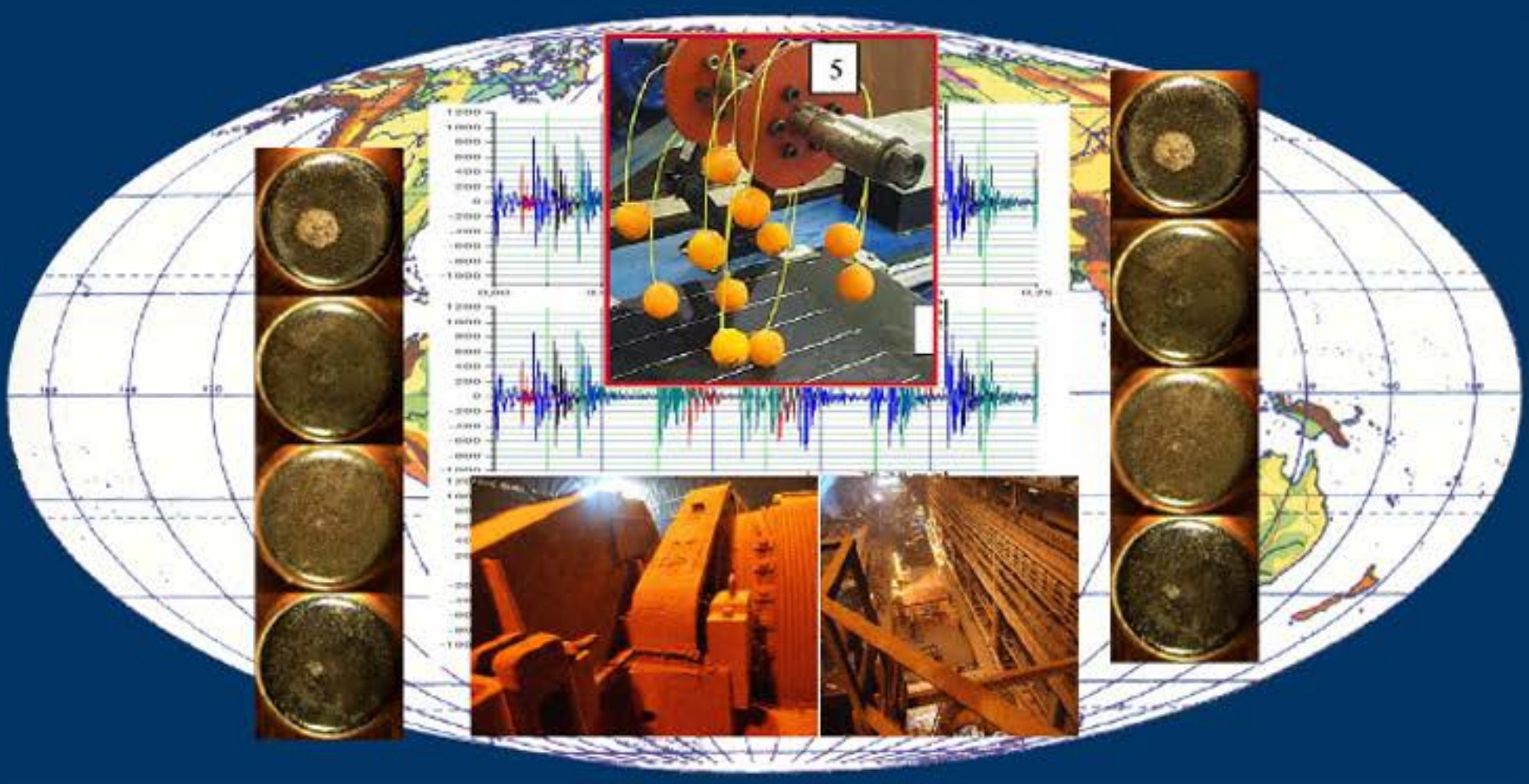
Vol. 21. No 2, 2019

ISSN 1507-2711
Cena: 25 zł (w tym 5% VAT)

EKSPLOATACJA I NIEZAWODNOŚĆ
MAINTENANCE AND RELIABILITY



Polskie Naukowo Techniczne Towarzystwo Eksploatacyjne
Warszawa

Polish Maintenance Society
Warsaw

Abstracts ... III

Leszek GOMÓŁKA, Andrzej AUGUSTYNOWICZ

Evaluation of applicability of dielectric constant in monitoring aging processes in engine oils
Ocena przydatności stałej dielektrycznej do monitorowania procesu starzenia oleju silnikowego ... 177

Kangkang SUN, Guoqiang WANG, Yanpeng LU

Optimization method of bevel gear reliability based on genetic algorithm and discrete element
Metoda optymalizacji niezawodności przekładni stożkowej z zastosowaniem algorytmu genetycznego i elementów dyskretnych ... 186

Andrzej URBAŚ, Marek SZCZOTKA

The influence of the friction phenomenon on a forest crane operator's level of discomfort
Wpływ zjawiska tarcia na poziom dyskomfortu pracy operatora żurawia leśnego ... 197

Agnieszka CHUDZIK, Bogdan WARDA

Effect of radial internal clearance on the fatigue life of the radial cylindrical roller bearing
Wpływ wewnętrznego luzu promieniowego na trwałość zmęczeniową promieniowego łożyska walcowego ... 211

Andrzej MARCZUK, Wojciech MISZTAL, Petr SAVINYKH, Nikolay TURUBANOV, Alexey ISUPOV, Dmitry ZYRYANOV

Improving efficiency of horizontal ribbon mixer by optimizing its constructional and operational parameters
Poprawa efektywności pracy poziomego mieszalnika wstęgowego poprzez optymalizację parametrów konstrukcyjnych i eksploatacyjnych ... 220

Vicente MACIÁN, Bernardo TORMOS, Jesús HERRERO

Maintenance management balanced scorecard approach for urban transport fleets
Zarządzanie utrzymaniem ruchu w zakładach komunikacji miejskiej w oparciu o zrównoważoną kartę wyników ... 226

Andrzej KSIAZKIEWICZ, Jerzy JANISZEWSKI

Welding tendency for selected contact materials under different switching conditions
Tendencja sczepiania wybranych materiałów stykowych w różnych warunkach łączeniowych ... 237

Junxing LI, Zhihua WANG, Chengrui LIU, Ming QIU

Accelerated degradation analysis based on a random-effect Wiener process with one-order autoregressive errors
Przyspieszona analiza degradacji w oparciu o proces Wienera z efektem losowym z błędami autoregresyjnymi pierwszego rzędu ... 246

Leszek KNOPIK, Klaudiusz MIGAWA

Semi-Markov system model for minimal repair maintenance
Semi-markowski model systemu obsługi z minimalną naprawą ... 256

Tomasz RYMARCZYK, Grzegorz KŁOSOWSKI

Innovative methods of neural reconstruction for tomographic images in maintenance of tank industrial reactors
Nowatorskie metody neuronowej rekonstrukcji obrazów tomograficznych w eksploatacji zbiornikowych reaktorów przemysłowych ... 261

Tahir KHANIYEV, M.Bahar BASKIR, Fikri GOKPINAR, Farhad MIRZAYEV

Statistical distributions and reliability functions with type-2 fuzzy parameters
Rozkłady statystyczne i funkcje niezawodności o parametrach rozmytych typu-2 ... 268

Kristina KILIKEVIČIENĖ, Jonas MATIJOŠIUS, Antanas FURSENKO, Artūras KILIKEVIČIUS

Tests of hail simulation and research of the resulting impact ,on the structural reliability of solar cells
Badania symulacyjne wpływu opadów gradu na niezawodność konstrukcji ogniw słonecznych ... 275

Andrzej GOŁAŚ, Wojciech CIESIELKA, Krystian SZOPA, Paweł ZYDROŃ, Wojciech BĄCHOREK, Mariusz BENESZ, Aleksander KOT, Szczepan MOSKWA

Analysis of the possibilities to improve the reliability of a 15 kV overhead line exposed to catastrophic icing in Poland
Analiza możliwości poprawy niezawodności napowietrznej linii 15 kV narażonej na katastrofalne oblodzenie w warunkach polskich ... 282

Yu WANG, Linhan GUO, Meilin WEN, Yi YANG

Availability analysis for a multi-component system with different k-out-of-n:G warm standby subsystems subject to suspended animation
Analiza gotowości systemu wieloelementowego składającego się z różnych wygaszalnych podsystemów typu k-z-n:G stanowiących rezerwę ciepłą ... 289

Grzegorz MIECZKOWSKI

Criterion for crack initiation from notch located at the interface of bi-material structure
Kryterium pękania struktury bi-materiałowej z ostrym karbem usytuowanym na interfejsie 301

Ahmed AL-GARNI, Wael ABDELRAHMAN, Ayman ABDALLAH

ANN-based failure modeling of classes of aircraft engine components using radial basis functions
Modelowanie uszkodzeń elementów silnika samolotowego w oparciu o sztuczne sieci neuronowe o radialnych funkcjach bazowych 311

Cher Ming TAN, Udit NARULA , Lu An LAI, Sumit PANDEY, Jung Hua TUNG, Chung Yi LI

Optimal maintenance strategy on medical instruments used for haemodialysis process
Optymalna strategia konserwacji urządzeń medycznych wykorzystywanych w procesie hemodializy 318

Jacek WAWRZOSEK, Szymon IGNACIUK, Andrzej BOCHNIAK

Selected problems of ambiguity of the dual price of water in the post-optimization analysis of the water supply system
Wybrane problemy niejednoznaczności ceny dualnej wody w postoptymalizacyjnej analizie systemu wodociągów 329

Marek SIKORA, Krzysztof SZCZYRBA, Łukasz WRÓBEL, Marcin MICHALAK

Monitoring and maintenance of a gantry based on a wireless system for measurement and analysis of the vibration level
Monitorowanie i utrzymanie suwnicy bramowej na podstawie bezprzewodowego systemu pomiaru i analizy poziomu drgań 341

Bolesław KARWAT, Ryszard MACHNIK, Jerzy NIEDŹWIEDZKI, Magdalena NOGAJ, Piotr RUBACHA, Emil STAŃCZYK

Calibration of bulk material model in Discrete Element Method on example of perlite D18-DN
Kalibracja modelu materiału sypkiego w Metodzie Elementów Dyskretnych na przykładzie perlitu D18-DN 351

GOMÓŁKA L, AUGUSTYNOWICZ A. **Evaluation of applicability of dielectric constant in monitoring aging processes in engine oils**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 177–185, http://dx.doi.org/10.17531/ein.2019.2.1.

This paper reports on the study involving the development of a dependence between the selected physicochemical properties of engine oil and the variation of its dielectric constant, considered as an evaluation parameter for the degree of its exploitation. In order to get to know the selected phenomena that accompany oil aging, a decision was made to perform tests of the condition of engine oils at various stages of their exploitation. In addition to the analysis of the dielectric constant, measurements of total base number and (TBN) and infrared spectroscopy (IR) were also performed. The interdependence between the examined values is developed using the Pearson's linear correlation coefficient. As a result, the variability of TBN and IR absorbance in the context of the change in the dielectric constant were evaluated. Following the realized experimental tests, it was found that the results demonstrate a satisfactory correlation between the selected physicochemical properties of the oil, which resulted in the adoption of an assumption of a high feasibility of application of dielectric constant as a diagnostic parameter in monitoring the state of exploitation of engine oil.

GOMÓŁKA L, AUGUSTYNOWICZ A. **Ocena przydatności stałej dielektrycznej do monitorowania procesu starzenia oleju silnikowego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 177–185, http://dx.doi.org/10.17531/ein.2019.2.1.

Celem pracy było określenie związków pomiędzy wytypowanymi właściwościami fizykochemicznymi oleju silnikowego a zmianą jego stałej dielektrycznej, rozważanej jako parametr oceniający stan jego zestarzenia. W celu szczegółowego poznania wybranych zjawisk towarzyszących starzeniu się oleju zdecydowano się na przebadanie oleju w różnych stadiach jego użytkowania. Poza badaniem stałej dielektrycznej wykonano również badania całkowitej liczby zasadowej TBN oraz badania spektrometryczne w podczerwieni IR. Stopień współzależności badanych wielkości przedstawiono wykorzystując do tego współczynnik korelacji liniowej Pearsona. W efekcie oceniono zmienność całkowitej liczby zasadowej oraz absorbancji IR w kontekście zmiany stałej dielektrycznej. Po przeprowadzeniu badań eksperymentalnych, stwierdzono, że uzyskane wyniki wskazują na zadawalającą korelację pomiędzy wytypowanymi własnościami fizykochemicznymi oleju, co pozwoliło z dużą dozą pewności przyjąć, iż stała dielektryczna może być użyta w charakterze parametru diagnostycznego do monitorowania stanu oleju silnikowego.

SUN K, WANG G, LU Y. **Optimization method of bevel gear reliability based on genetic algorithm and discrete element**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 186–196, http://dx.doi.org/10.17531/ein.2019.2.2.

Gear transmission is the most basic transmission component in mechanical transmission system. Many scholars have done a lot of research on gear reliability. When the variation coefficient is used to calculate and optimize the reliability of bevel gear, in order to calculate the reliability of bevel gear, it is often assumed that the gear works under constant torque, that is, the coefficient of variation (COV) is zero, but this is not the case in practice. In this paper, a gear reliability method based on discrete element simulation is proposed. The purpose of this method is to simulate the actual working conditions of gears, calculate more accurate coefficient of variation in the real world, and improve the accuracy of gear reliability design. Firstly, the real working conditions of the bevel gear transmission are simulated by discrete element method (DEM), and in the transmission system, the tangential force COV of the bevel gear is proved to be equal to the torque COV of the crusher central shaft. Secondly, the multi-objective function model of the gear transmission system is established based on the double tooth roll crusher (DTRC). The optimal volume and reliability of the bevel gear transmission are taken as the objective function, and the teeth number, module and face width factor of basic parameters of gear are optimized by genetic algorithm (GA). Finally, the accuracy of the optimization results is verified by Monte Carlo method. The main purpose of the manuscript is to analyse the effect of actual conditions (DEM simulation) on the optimization results. The results show that the COV of nominal tangential load of bevel gear is about 0.65 under actual working conditions, so in order to guarantee the same reliability, total volume need to be increased by 34.4%. This method is similar to the selection of gear safety factor. In practical production, the selection of safety factor is often based on experience. This paper provides a new method to optimize the reliability of bevel gear, combining with DEM simulation, which provides theoretical guidance for optimal design of bevel gear.

SUN K, WANG G, LU Y. **Metoda optymalizacji niezawodności przekładni stożkowej z zastosowaniem algorytmu genetycznego i elementów dyskretnych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 186–196, http://dx.doi.org/10.17531/ein.2019.2.2.

Przekładnia zębata to podstawowy element mechanicznego układu napędowego. Niezawodność przekładni jest przedmiotem wielu badań. Przy obliczeniach i optymalizacji niezawodności przekładni stożkowej z wykorzystaniem współczynnika zmienności, często przyjmuje się, że przekładnia pracuje w warunkach stałego momentu obrotowego, t.j. że współczynnik zmienności wynosi 1. Sytuacja taka jednak nie występuje w praktyce. W niniejszej pracy zaproponowano metodę optymalizacji niezawodności przekładni opartą na symulacji metodą elementów dyskretnych. Celem tej metody jest zasymulowanie rzeczywistych warunków pracy przekładni, dokładniejsze obliczenie rzeczywistego współczynnika zmienności oraz poprawa dokładności projektowania niezawodności przekładni. W pierwszej kolejności, na przykładzie kruszarki podwójnej, wyznaczono model działania układu przekładni stożkowej wykorzystujący wielokryterialną funkcję celu. Optymalną objętość i niezawodność przekładni stożkowej przyjęto jako funkcje celu. Następnie, za pomocą metody elementów dyskretnych, symulowano rzeczywiste warunki pracy przekładni. Wyznaczono moment obrotowy przekładni stożkowej i współczynnik zmienności siły wypadkowej, a podstawowe parametry koła zębatego: liczbę zębów, moduł zęba i współczynnik szerokości zębów, zoptymalizowano za pomocą algorytmu genetycznego. Trafność wyników optymalizacji weryfikowano metodą Monte Carlo. Wyniki pokazują, że badana metoda może skutecznie poprawiać niezawodność przekładni stożkowej.

URBAŚ A, SZCZOTKA M. **The influence of the friction phenomenon on a forest crane operator's level of discomfort**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 197–210, http://dx.doi.org/10.17531/ein.2019.2.3.

A mathematical model of a forest crane that is suitable for dynamics analysis of its operation cycle is presented in this paper. The flexibility of the operator's seat, drives and supports is taken into account. Joint coordinates are applied to describe the motion of the links together with the homogeneous transformations technique. Lagrange equations of the second order are used when deriving the equations of motions. Joint forces and torques are determined based on recursive Newton-Euler algorithms. These joint forces are then used in the LuGre friction model, which allows to calculate the friction coefficients and friction forces. Numerical analyses performed here show the influence of various friction forces on the vibration level as perceived by the operator of the crane. The level of discomfort is discussed based on standards commonly used in the vehicle and transportation industry for evaluations of vibration comfort.

URBAŚ A, SZCZOTKA M. **Wpływ zjawiska tarcia na poziom dyskomfortu pracy operatora żurawia leśnego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 197–210, http://dx.doi.org/10.17531/ein.2019.2.3.

W niniejszym artykule przedstawiono model matematyczny żurawia leśnego, który jest stosowany do analizy dynamiki cyklu jego pracy. Uwzględniono podatność podparcia fotela operatora, napędów oraz podpór. Do opisu ruchu członów stosuje się współrzędne złączowe i macierze przekształceń jednorodnych. Do wyprowadzenia równań ruchu modelu żurawia zastosowano podejście bazujące na formalizmie równań Lagrange'a drugiego rodzaju. Siły i momenty węzłowe są określane na podstawie rekurencyjnego algorytmu Newtona-Eulera. Siły te są następnie wykorzystywane w modelu tarcia LuGre, który pozwala obliczyć współczynniki i siły tarcia. Przeprowadzone analizy numeryczne pokazują wpływ różnych sił tarcia na poziom drgań odczuwany przez operatora żurawia. Poziom dyskomfortu operatora wywołany przez drgania maszyny został oszacowany w oparciu o często stosowane w przemyśle samochodowym i transportowym odpowiednie standardy.

CHUDZIK A, WARDA B. **Effect of radial internal clearance on the fatigue life of the radial cylindrical roller bearing**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 211–219, http://dx.doi.org/10.17531/ein.2019.2.4.

The paper presents result of the research on influence of internal radial clearance in radial cylindrical roller bearing on its fatigue durability. By solving the Boussinesq problem for the elastic half-space and finite elements method, stress distributions were determined, necessary to estimate predicted fatigue life of the bearing. The calculations took into account geometrical parameters of the bearing: its radial clearance and shape of rolling parts. Predicted radical clearance was computed by using the Lundberg and Palmgren model. ANSYS program allowed to introduce the

CHUDZIK A, WARDA B. **Wpływ wewnętrznego luzu promieniowego na trwałość zmęczeniową promieniowego łożyska walcowego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 211–219, http://dx.doi.org/10.17531/ein.2019.2.4.

W pracy przedstawiono wyniki badań wpływu wewnętrznego luzu promieniowego w promieniowym łożysku walcowym na jego prognozowaną trwałość zmęczeniową. Wykorzystując zagadnienie Boussinesqa dla półprzestrzeni sprężystej i metodę elementów skończonych, określono rozkłady naprężeń podpowierzchniowych niezbędne do oszacowania prognozowanej trwałości zmęczeniowej łożyska. W obliczeniach uwzględniono geometryczne parametry łożyska: jego luz promieniowy i kształt części tocznych. Do określenia wartości luzu promieniowego wykorzystano model Lundberga i Palmgrena.

analysis of von Mises stress distribution in any plane of cooperating components. The outcome revealed, radial cylindrical roller bearing will have highest endurance with slight radial clearance.

MARCZUK A, MISZTAL W, SAVINYKH P, TURUBANOV N, ISUPOV A, ZYRYANOV D. **Improving efficiency of horizontal ribbon mixer by optimizing its constructional and operational parameters**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 220–225, http://dx.doi.org/10.17531/ein.2019.2.5.

The homogeneity of mixing various feed mixture ingredients and the reliability of mixers directly depend on the constructional features of mixing devices and operational parameters of their work. This article presents a theoretical underpinning for the structural (design) specifications of a horizontal ribbon auger mixer (blender) with regard to the fulfilment of the condition when the material is transported by auger tapes in different directions, ensuring the efficient mixing of the feed mixture ingredients and timely material unloading from the mixer tank (hopper). The performed optimization of the structural and operational specifications as a result of experimental studies has allowed to increase the reliability of the mixer and to determine the homogeneity coefficient of the finished product, which reaches its the highest value of 94.13% with the forward direction of rotation of the mixer movable operating elements, when mixing time is 13 minutes and when the tank load is 52.4% of its volume.

MACIÁN V, TORMOS B, HERRERO J. **Maintenance management balanced scorecard approach for urban transport fleets**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 226–236, http://dx.doi.org/10.17531/ein.2019.2.6.

Attending the important role of maintenance function in any production or service provider company, the measurement and assessment of maintenance performance is crucial for competitiveness and future survival. That situation is even more critical in urban transport fleets where some specific boundary conditions and special characteristics will affect maintenance policy and implementation. This paper presents a deep review of different studies worldwide to define the most proper and effective maintenance performance indicators, selecting and refining the most important ones to obtain a reduced maintenance management balanced scorecard. That balanced scorecard is proposed as a main tool for urban transport fleet maintenance managers to assess efficiency and effectiveness of maintenance processes and will be used as a basis for a future benchmarking process for this type of companies.

KSIAZKIEWICZ A, JANISZEWSKI J. **Welding tendency for selected contact materials under different switching conditions**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 237–245, http://dx.doi.org/10.17531/ein.2019.2.7.

The flow of significant current through electric contacts may lead to contact welding. In a.c. circuits this phenomena is not only dependent on properties of contact material (i.e. resistance to welding) but on the phase in which current is switched on. Welding tendency for contact materials made from AgNi, AgCdO and AgSnO2 was evaluated based on selected phase at which make operation took place. The test circuit was protected by overcurrent apparatus to simulate real working environment. It is observed that welding tendency for the selected contact materials is contingent to current phase at which make operation is done.

LI J, WANG Z, LIU C, QIU M. **Accelerated degradation analysis based on a random-effect Wiener process with one-order autoregressive errors**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 246–255, http://dx.doi.org/10.17531/ein.2019.2.8.

For highly reliable and long-life products, accelerated degradation test (ADT) is often an effective and attractive way to assess the reliability. To analyze the accelerated degradation data, it has been well recognized that it is necessary to incorporate three sources of variability including the temporal variability, the unit-to-unit variability and measurement errors into the ADT model. The temporal variability can be properly described by the Wiener process. However, the randomness of the initial degradation level, which is an important part of the unit-to-unit variability, has been often neglected. In addition, regarding the measurement errors, current ADT models often assumed them to follow a mutually independent normal distribution and ignored the autocorrelation that may probably exist in them. These problems lead to a poor accuracy for reliability evaluation in some situation. Thus, a random-effect Wiener process-based ADT model considering one-order autoregressive (AR(1)) errors is proposed. Then closed-form expressions for the failure time distribution (FTD) is derived based on the concept of first hitting time (FHT). A statistical inference method is adopted to estimate unknown parameters. Finally, a comprehensive simulation study and a practical application are given to demonstrate the rationality and effectiveness of the proposed model.

Zastosowanie programu ANSYS pozwoliło na przeprowadzenie analizy rozkładu naprężeń von Misesa w dowolnej płaszczyźnie współpracujących ze sobą elementów tocznych łożyska. Analiza obliczeń wykazała, że promieniowe łożysko walcowe będzie miało najwyższą wytrzymałość z niewielkim luzem promieniowym.

MARCZUK A, MISZTAL W, SAVINYKH P, TURUBANOV N, ISUPOV A, ZYRYANOV D. **Poprawa efektywności pracy poziomego mieszalnika wstęgowego poprzez optymalizację parametrów konstrukcyjnych i eksploatacyjnych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 220–225, http://dx.doi.org/10.17531/ein.2019.2.5.

Jednorodność mieszania różnych składników mieszanki paszowej i niezawodność mieszalników zależy bezpośrednio od cech konstrukcyjnych urządzeń mieszających i parametrów roboczych ich pracy. W artykule przedstawiono teoretyczne podstawy dotyczące parametrów konstrukcyjnych (projektowych) poziomego mieszalnika ślimakowego (mieszarki) wywierających wpływ na warunki przemieszczania (podczas transportu w różnych kierunkach) zapewniających efektywne mieszanie składników mieszanki paszowej i terminowy rozładunek materiału ze zbiornika mieszalnika (zasobnika). Przeprowadzona na bazie badań eksperymentalnych optymalizacja parametrów konstrukcyjnych i eksploatacyjnych pozwoliła na zwiększenie niezawodności mieszalnika i wyznaczenie współczynnika homogeniczności gotowego produktu, który osiąga najwyższą wartość 94,13% podczas przemieszczania przy normalnym kierunku obrotów ruchomych elementów mieszalnika, z czasem mieszania wynoszącym 13 minut, gdy ładunek stanowi 52,4% objętości zbiornika.

MACIÁN V, TORMOS B, HERRERO J. **Zarządzanie utrzymaniem ruchu w zakładach komunikacji miejskiej w oparciu o zrównoważoną kartę wyników**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 226–236, http://dx.doi.org/10.17531/ein.2019.2.6.

Biorąc pod uwagę ważną rolę jaką pełni utrzymanie ruchu w firmach produkcyjnych i usługowych, pomiar i ocena wydajności eksploatacji ma kluczowe znaczenie dla konkurencyjności tych firm i ich przetrwania na rynku. Sytuacja ta jest szczególnie ważna w zakładach komunikacji miejskiej, w których pewne szczególne warunki brzegowe i szczególne cechy floty transportowej mają wpływ na politykę utrzymania ruchu i jej realizację. W niniejszym artykule przedstawiono dokładny przegląd różnych badań prowadzonych na całym świecie w celu określenia najbardziej odpowiednich i skutecznych wskaźników efektywności utrzymania ruchu, wybierając najważniejsze z nich i i udoskonalając je tak aby uzyskać zrównoważoną kartę wyników zarządzania utrzymaniem ruchu z ograniczoną liczbą parametrów. Zrównoważona karta wyników może być stosowana przez specjalistów utrzymania ruchu zakładów komunikacji miejskiej do oceny wydajności i skuteczności procesów konserwacji i stanowić będzie podstawę przyszłych analiz porównawczych dla tego typu przedsiębiorstw.

KSIAZKIEWICZ A, JANISZEWSKI J. **Tendencja sczepiania wybranych materiałów stykowych w różnych warunkach łączeniowych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 237–245, http://dx.doi.org/10.17531/ein.2019.2.7.

Przepływ znacznego prądu przez styki elektryczne może prowadzić do ich sczepienia. W obwodach prądu przemiennego to zjawisko jest nie tylko zależne od właściwości materiału stykowego (tj. odporności na sczepianie), ale od fazy, w której prąd jest załączany. Zbadana została tendencja do sczepiania się styków, w zależności od fazy załączania prądu, wykonanych z następujących materiałów stykowych: AgNi, AgCdO i AgSnO2. Obwód probierczy zabezpieczony był przez zabezpieczenia nadprądowe, w celu symulacji normalnych warunków pracy. Zauważono, że tendencja do sczepiania styków, dla wybranych materiałów stykowych, jest zależna od fazy, w której wykonuje się załączenie obwodu.

LI J, WANG Z, LIU C, QIU M. **Przyspieszona analiza degradacji w oparciu o proces Wienera z efektem losowym z błędami autoregresyjnymi pierwszego rzędu**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 246–255, http://dx.doi.org/10.17531/ein.2019.2.8.

W przypadku wysoce niezawodnych produktów o długim cyklu życia, przyspieszone badanie degradacji (ADT) często stanowi skuteczny i atrakcyjny sposób oceny niezawodności. Jak wiadomo, analiza danych z przyspieszonej degradacji wymaga włączenia do modelu ADT trzech źródeł zmienności, w tym zmienności czasowej, zmienności między jednostkami i błędów pomiarowych. Zmienność czasową można odpowiednio opisać za pomocą procesu Wienera. Jednak losowość początkowego poziomu degradacji, który stanowi ważną część zmienności między jednostkami, jest często w badaniach pomijana. Ponadto, w odniesieniu do błędów pomiaru, obecne modele ADT często zakładają, że mają one wzajemnie niezależne rozkłady normalne, ignorując możliwą autokorelację. Problemy te prowadzą w niektórych sytuacjach do niskiej trafności oceny niezawodności. W związku z powyższym, zaproponowano model ADT oparty na procesie Wienera z efektem losowym, w którym uwzględniono błędy autoregresyjne pierwszego rzędu (AR (1)). Następnie, w oparciu o pojęcie pierwszego czasu przejścia, wyprowadzono wyrażenia w postaci zamkniętej dla rozkładu czasu uszkodzenia (FTD). Do oszacowania nieznanych parametrów przyjęto metodę wnioskowania statystycznego. Na koniec przedstawiono kompleksowe studium symulacyjne i wskazano praktyczne zastosowanie modelu w celu wykazania jego racjonalności i skuteczności.

KNOPIK L, MIGAWA K. **Semi-Markov system model for minimal repair maintenance**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 256–260, http://dx.doi.org/10.17531/ein.2019.2.9.

This paper analyzes the semi-Markov model of technical objects age-replacements. The model includes two types of repairs: perfect repairs and minimal repairs. Minimal repairs in semi-Markov models have been studied in literature only to an extent. In this paper, the asymptotic availability coefficient and profit per time unit are considered as criteria for the quality of the system operation. The paper formulates various conditions for the occurrence of the maximum of criteria functions. The two numerical examples given at the end of the paper illustrate the results obtained in the paper.

RYMARCZYK T, KŁOSOWSKI G. **Innovative methods of neural reconstruction for tomographic images in maintenance of tank industrial reactors**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 261–267, http://dx.doi.org/10.17531/ein.2019.2.10.

The article presents an innovative concept of improving the monitoring and optimization of industrial processes. The developed method is based on a system of many separately trained neural networks, in which each network generates a single point of the output image. Thanks to the elastic net method, the implemented algorithm reduces the correlated and irrelevant variables from the input measurement vector, making it more resistant to the phenomenon of data noises. The advantage of the described solution over known non-invasive methods is to obtain a higher resolution of images dynamically appearing inside the reactor of artifacts (crystals or gas bubbles), which essentially contributes to the early detection of hazards and problems associated with the operation of industrial systems, and thus increases the efficiency of chemical process control.

KHANIYEV T, BASKIR MB, GOKPINAR F, MIRZAYEV F. **Statistical distributions and reliability functions with type-2 fuzzy parameters**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 268–274, http://dx.doi.org/10.17531/ein.2019.2.11.

Type-2 fuzzy sets were initially given by Zadeh as an extension of type-1 fuzzy sets. There is a growing interest in type-2 fuzzy set and its memberships (named secondary memberships) to handle the uncertainty in type-1 fuzzy set and its primary membership values. However, arithmetical operators on type-2 fuzzy sets have computational complexity due to third dimension of these sets. In this study, we present some mathematical operators which can be easily applied to type-2 fuzzy sets and numbers. Also, mathematical functions of type-2 fuzzy numbers are given according to their monotonicity. These functions are adapted to reliability and distribution functions of the random variables with the type-2 fuzzy parameters. These functions are applied to Exponential, Chi-square, Weibull distributions with respect to monotonicity of the parameters of these distributions.

KILIKEVIČIENĖ K, MATIJOŠIUS J, FURSENKO A, KILIKEVIČIUS A. **Tests of hail simulation and research of the resulting impact on the structural reliability of solar cells**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 275–281, http://dx.doi.org/10.17531/ein.2019.2.12.

The mandatory tests of resistance to hail is carried out in order to qualify solar cell modules according to standards (IEC 61215 and IEC 61646). The efficiency of modern photovoltaic systems decreases significantly, when the crystalline structure of solar cells is damaged as a result of climatic factors, such as wind, hail, etc., which are similar to mechanical-dynamic effects. This work presents a conducted research of efficiency and reliability of solar cells, simulating hail effects. A testbed was created specifically for carrying out experimental research. During the research, solar elements were exposed to impact, cyclic dynamic loads, with the frequency of revolutions of the balls simulating hail ranging from 5 to 20 Hz, the amplitude of the impact excitation acceleration of the solar cell - up to 986 m / s2 and the force amplitude - up to 1129 N. Experimental research results revealed the reliability of photovoltaic modules of different sizes during the simulation of hail. The proposed assessment methodology of hail effects can be successfully applied in studies of the influence of mechanical-dynamic effects of solar cells of different structures.

GOŁAŚ A, CIESIELKA W, SZOPA K, ZYDROŃ P, BĄCHOREK M, BENESZ M, KOT A, MOSKWA S. **Analysis of the possibilities to improve the reliability of a 15 kV overhead line exposed to catastrophic icing in Poland**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 282–288, http://dx.doi.org/10.17531/ein.2019.2.13.

The paper is a result of a synergic cooperation of two academic teams, i.e. power engineering and mechanical teams, and a distribution system operator. A real 15 kV overhead line exposed to a catastrophic load of ice and rime was analyzed and three solutions to improve the reliability of the tested object in such conditions were examined. Authors considered: shortening the length of the line spans, heating the main line with increased current and rebuilding the overhead line to a cable line. The researches worked out a FEM model taking into account the newest normatives,

KNOPIK L, MIGAWA K. **Semi-markowski model systemu obsługi z minimalną naprawą**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 256–260, http://dx.doi.org/10.17531/ein.2019.2.9.

W pracy bada się semimarkowski model wymian według wieku obiektów technicznych. W modelu uwzględnia się dwa rodzaje napraw: naprawy dokładne i naprawy minimalne. Naprawy minimalne w modelach semimarkowskich były badane w literaturze w niewielkim stopniu. Jako kryteria jakości pracy systemu rozważa się asymptotyczny współczynnik gotowości i zysk przypadający na jednostkę czasu. W pracy sformułowano różne warunki istnienia maksimum funkcji kryterialnych. Podane na końcu pracy dwa przykłady numeryczne ilustrują wyniki uzyskane w pracy.

RYMARCZYK T, KŁOSOWSKI G. **Nowatorskie metody neuronowej rekonstrukcji obrazów tomograficznych w eksploatacji zbiornikowych reaktorów przemysłowych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 261–267, http://dx.doi.org/10.17531/ein.2019.2.10.

W artykule przedstawiono nowatorską koncepcję usprawnienia monitoringu i optymalizacji procesów przemysłowych. Opracowana metoda bazuje na systemie osobno wytrenowanych wielu sieci neuronowych, w którym każda sieć generuje pojedynczy punkt obrazu wyjściowego. Dzięki zastosowaniu metody elastic net zaimplementowany algorytm redukuje z wejściowego wektora pomiarowego zmienne skorelowane i nieistotne, czyniąc go bardziej odpornym na zjawisko zaszumienia danych. Przewagą opisywanego rozwiązania nad znanymi metodami nieinwazyjnymi jest uzyskanie wyższej rozdzielczości obrazów dynamicznie pojawiających się wewnątrz reaktora artefaktów (kryształów lub pęcherzy gazowych), co zasadniczo przyczynia się do wczesnego wykrycia zagrożeń i problemów związanych z eksploatacją systemów przemysłowych, a tym samym zwiększa efektywność sterowania procesami chemicznymi.

KHANIYEV T, BASKIR MB, GOKPINAR F, MIRZAYEV F. **Rozkłady statystyczne i funkcje niezawodności o parametrach rozmytych typu-2**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 268–274, http://dx.doi.org/10.17531/ein.2019.2.11.

Zbiory rozmyte typu 2 po raz pierwszy zaproponował Zadeh jako rozszerzenie zbiorów rozmytych typu 1. Zbiory rozmyte typu 2 oraz ich funkcje przynależności (zwane wtórnymi funkcjami przynależności) cieszą się rosnącym zainteresowaniem, ponieważ pozwalają na modelowanie niepewności w zbiorze rozmytym typu 1 oraz wartości pierwotnych funkcji przynależności do takiego zbioru. Ich wadą jest złożoność obliczeniowa operatorów arytmetycznych wynikająca z trójwymiarowości tych zbiorów. W artykule przedstawiono operatory matematyczne, które można z powodzeniem stosować w odniesieniu do zbiorów i liczb rozmytych typu 2. Podano również funkcje matematyczne liczb rozmytych typu 2 zgodnie z ich monotonicznością. Funkcje te są dostosowane do funkcji niezawodności i rozkładu zmiennych losowych z parametrami rozmytymi typu 2. Można je stosować do opisu rozkładów wykładniczych, chi-kwadrat, oraz Weibulla w odniesieniu do monotoniczności parametrów tych rozkładów.

KILIKEVIČIENĖ K, MATIJOŠIUS J, FURSENKO A, KILIKEVIČIUS A. **Badania symulacyjne wpływu opadów gradu na niezawodność konstrukcji ogniw słonecznych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 275–281, http://dx.doi.org/10.17531/ein.2019.2.12.

Przy kwalifikacji modułów ogniw słonecznych do użytkowania przeprowadza się obowiązkowe badania odporności na gradobicie zgodnie z normami IEC 61215 i IEC 61646. Wydajność nowoczesnych systemów fotowoltaicznych znacznie spada, gdy struktura krystaliczna ogniw słonecznych ulega uszkodzeniu w wyniku czynników klimatycznych, takich jak wiatr, gradobicie itp., które przypominają w swoim działaniu obciążenia mechaniczno-dynamiczne. W pracy przedstawiono wyniki badań symulacyjnych wpływu gradobicia na wydajność i niezawodność ogniw słonecznych. Badania prowadzono na specjalnie do tego celu skonstruowanym stanowisku testowym. Podczas badań, elementy słoneczne były wystawiane na cykliczne obciążenia dynamiczne wywoływane uderzeniami kulek symulujących grad o częstotliwości obrotów od 5 do 20 Hz przy amplitudzie przyspieszenia wzbudzenia uderzeniowego ogniwa słonecznego wynoszącej do 986 m/s2 oraz amplitudzie siły do 1129 N. Wyniki symulacji pozwoliły ocenić niezawodność modułów fotowoltaicznych o różnych rozmiarach. Proponowaną metodologię oceny wpływu opadów gradu można z powodzeniem stosować w badaniach oddziaływania obciążeń mechaniczno-dynamicznych na ogniwa słoneczne o różnych strukturach.

GOŁAŚ A, CIESIELKA W, SZOPA K, ZYDROŃ P, BĄCHOREK M, BENESZ M, KOT A, MOSKWA S. **Analiza możliwości poprawy niezawodności napowietrznej linii 15 kV narażonej na katastrofalne oblodzenie w warunkach polskich**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 282–288, http://dx.doi.org/10.17531/ein.2019.2.13.

Praca jest efektem synergicznej współpracy dwóch zespołów akademickich: elektroenergetycznego i mechanicznego oraz operatora systemu dystrybucyjnego. Analizie poddano rzeczywistą, napowietrzną linię średniego napięcia 15 kV narażoną na katastrofalne obciążenia lodem i szadzią. Zbadano możliwość zastosowania trzech rozwiązań mogących poprawić niezawodność badanego obiektu w takich warunkach. Rozważono: skrócenie długości przęseł linii, podgrzewanie magistrali zwiększonym prądem roboczym oraz przebudowę linii do linii kablowej. W celu realizacji pracy wykonano badania mode-

simulated the model, experimentally increased the load on the real line with measured wire temperature, and performed multi-variant calculations to determine indicators of reliability, i.e. SAIDI and SAIFI. The analyses were followed by conclusions thanks to which the reliability of power lines exposed to catastrophic icing could be increased. These inferences should be considered and applied by all distribution system operators in Poland.

lowo-symulacyjne MES z uwzględnieniem najnowszych wytycznych normatywnych, zrealizowano eksperyment dociążenia linii wraz z pomiarem temperatury przewodów oraz przeprowadzono wielowariantowe obliczenia niezawodnościowe prowadzące do wyznaczenia wskaźników SAIDI i SAIFI. W wyniku szczegółowych analiz sprecyzowano wnioski końcowe pozwalające na zwiększenie niezawodności linii elektroenergetycznych narażonych na katastrofalne oblodzenie, które powinny być rozważone i stosowane przez wszystkich operatorów systemów dystrybucyjnych w Polsce.

WANG Y, GUO L, WEN M, YANG Y. **Availability analysis for a multi-component system with different *k*-out-of-*n*:G warm standby subsystems subject to suspended animation**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 289–300, http://dx.doi.org/10.17531/ein.2019.2.14.

Industrial equipment or systems are usually constructed as a multi-component series system with *k*-out-of-*n*:G subsystems to fulfill a specified function. As a common type of standby, warm standby is considered in the multi-component series system with *k*-out-of-*n*:G standby subsystems. When a subsystem fails, the non-failed subsystems are shut off and cannot fail, which is defined as suspended animation (SA). If the SA is ignored the non-failed subsystems are assumed to keep working in the SA time, which will cause inaccuracy in the availability analysis for the system. In this paper, we focus on the SA to construct an availability model for a multi-component series system with *k*-out-of-*n*:G warm standby subsystems. Multiple continuous time Markov chains are constructed to model the system availability. A Monte Carlo simulation has been carried out to verify our method. Several interesting findings are obtained. 1) The failure rates of subsystems with SA and their limits are derived. 2) The closed-form expressions for the stationary availability of the system and subsystems, mean time to failure, mean time to repair and stationary failure frequency are obtained considering SA. 3) The system stationary availability is a monotone function for its parameters. 4) The SA effect on the stationary availability should be emphasized in two cases, one is both the value of n/k and the failure rate of active components in a *k*-out-of-*n* subsystem are relatively large or small, the other is both the value of n/k and the repair rate are relatively small.

WANG Y, GUO L, WEN M, YANG Y. **Analiza gotowości systemu wieloelementowego składającego się z różnych wygaszalnych podsystemów typu *k*-z-*n*:G stanowiących rezerwę ciepłą**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 289–300, http://dx.doi.org/10.17531/ein.2019.2.14.

Aby urządzenia i systemy przemysłowe mogły pełnić swoje określone funkcje, zwykle buduje się je w postaci wieloelementowych systemów szeregowych składających się z podsystemów typu *k*-z-*n*: G. W pracy rozważano zagadnienie rezerwy ciepłej w wieloelementowym systemie szeregowym składającym się z podsystemów rezerwowych typu *k*-z-*n*: G. W przypadku awarii jednego z takich podsystemów, pozostałe, działające podsystemy wyłącza się, dzięki czemu nie mogą one ulec uszkodzeniu. Procedurę taką określa się, przez analogię z organizmami żywymi mianem anabiozy (suspended animation) lub wygaszania. Pominięcie zjawiska wygaszania, prowadzi do założenia, że podsystemy, które nie uległy uszkodzeniu pracują w czasie wygaszenia, co skutkuje nieprawidłowościami w analizie gotowości systemu. W artykule koncepcję wygaszania podsystemów wykorzystano do budowy modelu gotowości wieloelementowego systemu szeregowego składającego się z podsystemów typu *k*-z-*n*: G stanowiących rezerwę ciepłą. W celu zamodelowania gotowości systemu, skonstruowano łańcuchy Markowa z czasem ciągłym. Przedstawioną metodę zweryfikowano za pomocą symulacji Monte Carlo. Uzyskano szereg interesujących wyników. 1) Obliczono intensywność uszkodzeń podsystemów wygaszonych i ich wartości graniczne. 2) Wyprowadzono, z uwzględnieniem procedury wygaszania, wyrażenia w postaci zamkniętej dla stacjonarnej gotowości rozważanego systemu i podsystemów oraz określono średni czas do uszkodzenia, średni czas do naprawy oraz stacjonarną częstotliwość uszkodzeń. 3) Gotowość stacjonarna systemu jest funkcją monotoniczną dla parametrów wejściowych systemu. 4) Należy podkreślić wpływ wygaszania na stacjonarną gotowość systemu w dwóch przypadkach: gdy stosunek n/k i intensywność uszkodzeń aktywnych elementów *k*-tego spośród *n* podsystemów są względnie niskie lub względnie wysokie, oraz gdy stosunek n/k i intensywność napraw są względnie niskie.

MIECZKOWSKI G. **Criterion for crack initiation from notch located at the interface of bi-material structure**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 301–310, http://dx.doi.org/10.17531/ein.2019.2.15.

The fracture process of bi-material structure with the notch was analysed in this work. For fracture prediction, a criterion based on the Theory of Critical Distances was used. Under analysis were elements made of aluminium alloy and polymer combination (with a various structural notch-tip angle), which then were subjected to the three-point bending test. Values of critical loads resulting from the used hypothesis were compared with values obtained from the experiment. Validation of the selected criterion required defining a qualitative and quantitative description of singular stress fields present around the structural notch-tip area. Therefore, such solutions were obtained and methodology of their determining was discussed.

MIECZKOWSKI G. **Kryterium pękania struktury bi-materiałowej z ostrym karbem usytuowanym na interfejsie**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 301–310, http://dx.doi.org/10.17531/ein.2019.2.15.

W pracy przedstawiono wyniki badań dotyczących pękania struktury bi-materiałowej z karbem usytuowanym na interfejsie. Do prognozowania inicjacji procesu pękania zastosowano kryterium oparte na punktowej teorii krytycznych dystansów. Analizowano elementy wykonane ze stopu aluminium i polimerów (PC, PMMA), które poddane były trójpunktowemu zginaniu. Wartości obciążeń krytycznych wynikających z wykorzystanej hipotezy porównano z wartościami uzyskanymi z eksperymentu. Walidacja wybranego kryterium wymagała określania jakościowego i ilościowego opisu osobliwych pól naprężeń, występujących w okolicy wierzchołkowej karbu strukturalnego. W związku z tym, uzyskano takie rozwiązania i omówiono metodykę ich otrzymywania.

AL-GARNI A, ABDELRAHMAN W, ABDALLAH A. **ANN-based failure modeling of classes of aircraft engine components using radial basis functions**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 311–317, http://dx.doi.org/10.17531/ein.2019.2.16.

The objective of this research is to present a model to predict failure of two categories of critical aircraft engine components; non-rotating components such as valves and gearboxes, and rotating components such as engine turbines. The work utilizes Weibull regression and artificial neural networks employing Back Propagation (BP) as well as Radial Basis Functions (RBF). The model utilizes training failure data collected from operators of turboprop aircraft working in harsh desert conditions, where sand erosion is a detrimental factor in reducing turbine life. Accordingly, the model is more suited for accurate prediction of life of critical components of such engines. The algorithm, which uses Radial Basis Function (RBF) NN, uses a closest point specifier. The activation is based on the deviation of the earlier prototype from the input vector. Two earlier models are used for comparison purposes; namely Weibull regression modeling and Feed-Forward BP network. Comparison results show that the failure times represented by RBF are in better compromise with actual failure data than both earlier modeling methods. Moreover, the technique has comparatively higher efficiency as the neuron's number in each layer of ANN is reduced, to decrease computation time, with minimum effect on the accuracy of results.

AL-GARNI A, ABDELRAHMAN W, ABDALLAH A. **Modelowanie uszkodzeń elementów silnika samolotoweg ow oparciu o sztuczne sieci neuronowe o radialnych funkcjach bazowych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 311–317, http://dx.doi.org/10.17531/ein.2019.2.16.

Celem pracy jest przedstawienie modelu służącego do predykcji uszkodzeń dwóch kategorii krytycznych elementów silnika samolotowego: elementów nieobrotowych, takich jak zawory i skrzynie biegów oraz elementów obrotowych, takich jak turbiny silnika. W pracy wykorzystano regresję Weibulla i sztuczne sieci neuronowe oparte na propagacji wstecznej oraz radialnych funkcjach bazowych (RBF). Model wykorzystuje dane o błędach zebrane od operatorów samolotów turbośmigłowych pracujących w trudnych warunkach pustynnych, gdzie erozja powodowana przez piasek stanowi szkodliwy czynnik ograniczający żywotność turbin. Prezentowany model jest więc szczególnie przydatny do trafnego prognozowania żywotności krytycznych elementów takich silników. Algorytm, który wykorzystuje sieci neuronowe o radialnych funkcjach bazowych, używa specyfikatora najbliższego punktu. Aktywacja bazuje na odchyleniu wcześniejszego prototypu od wektora wejściowego. Dwa wcześniejsze modele oparte na regresji Weibulla (Weibull regression modeling) oraz sieciach typu Feed-Forward Backpropagation wykorzystano do badań porównawczych. Wyniki porównania pokazują, że czasy uszkodzeń odwzorowane przez RBF pozostają w większej zgodzie z rzeczywistymi danymi o uszkodzeniach niż w przypadku obu wcześniejszych metod modelowania. Co więcej, technika ta ma porównywalnie większą efektywność, ponieważ liczba neuronów w każdej warstwie sieci neuronowej została zredukowana tak aby zmniejszyć czas obliczeń, przy minimalnym wpływie na dokładność wyników.

TAN CM, NARULA U, LAI LA, PANDEY S, TUNG JH, LI CY. **Optimal maintenance strategy on medical instruments used for haemodialysis process**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 318–328, http://dx.doi.org/10.17531/ein.2019.2.17.

Haemodialysis machines are one of the important medical equipment which is used to treat renal failures and minimum downtimes are thus essential. Uninterrupted and constant use of these machines in hospitals worldwide makes them vulnerable to failures if not maintained properly. Consequently, the maintenance cost for dialysis machine is high. A method to implement a cost effective maintenance strategy is demonstrated in this work. Root Cause Based Maintenance (RCBM) strategy is employed at the component level to optimize the Reliability Based Maintenance schedules derived from the existing maintenance and failure data. In order to minimize the average cost of maintenance for Haemodialysis machines and ensure their high operational availability, a Cost-Model is derived, and Genetic Algorithm is employed for optimization in this work. The application of RCBM strategy results in cost saving of about 60% of the cost incurred using current maintenance scheme. Statistical and optimization calculations are performed using Reliasoft's Weibull++ and MATLAB tools respectively.

WAWRZOSEK J, IGNACIUK S, BOCHNIAK A. **Selected problems of ambiguity of the dual price of water in the post-optimization analysis of the water supply system**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 329–340, http://dx.doi.org/10.17531/ein.2019.2.18.

In literature it is believed that the dual price of water is an objective premise for shaping the market price of water. However, the authors note that a single vector of dual prices in the distribution of water, when ambiguous, should not become the basis for making decisions both regulating the price of water and affecting the procedures for modernizing the water supply network. This work cautions water management engineers not to duplicate common software errors and indicates how, despite the complete lack of literature tips, the technical problems encountered could be practically solved. The linear dependence of the row vectors of the left-hand parameters of binding constraints in the linear programming model for water consumption is identified here as the reason for the ambiguity of dual price vectors. This ambiguity in the issues of water distribution requires shaping alternative technical scenarios allowing for a variant selection of the method for modifying the water abstraction system. Therefore, the principles for determining the proportionality of simultaneous changes in certain parameters of the right-hand conditions of constraint conditions are described. These principles for the optimal selection of the most productive vectors for the parametric linear programming method were formulated and indicated on a simplified model of water distribution. The methodology developed in the work enables, among others, generating alternative technical scenarios for saving varying amounts of water, resulting in various financial savings.

SIKORA M, SZCZYRBA K, WRÓBEL Ł, MICHALAK M. **Monitoring and maintenance of a gantry based on a wireless system for measurement and analysis of the vibration level**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 341–350, http://dx.doi.org/10.17531/ein.2019.2.19.

The paper describes a system for monitoring and diagnosing a gantry. The main goal of the system is to acquire, visualize and monitor vibration levels of the gantry crucial elements. The system is also equipped with a computing and analytical part which enables predictive maintenance related to the vibration level assessment. The system architecture can be used in other applications too, i.e. those which require a wireless network of vibration sensors to carry out diagnostic tasks.

KARWAT B, MACHNIK R, NIEDŹWIEDZKI J, NOGAJ M, RUBACHA P, Stańczyk E. **Calibration of bulk material model in Discrete Element Method on example of perlite D18-DN**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 351–357, http://dx.doi.org/10.17531/ein.2019.2.20.

Analytical methods for calculations of the transport machinery are often insufficient especially when untypical granular materials are considered. Discrete Element Method (DEM) is a very useful numerical tool supporting designing and optimization of the transport equipment. However, to obtain reliable DEM simulation results an accurate set of input parameters values is needed. The most common calibration approach is to make use of a procedure where laboratory tests are performed and then the same experiments are numerically replicated in DEM. The article presents calibration of the DEM input parameters on the example of perlite D18-DN Based on the performed calibration, the model of perlite transport in a screw conveyor has been shown.

TAN CM, NARULA U, LAI LA, PANDEY S, TUNG JH, LI CY. **Optymalna strategia konserwacji urządzeń medycznych wykorzystywanych w procesie hemodializy**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 318–328, http://dx.doi.org/10.17531/ein.2019.2.17.

Aparaty do hemodializy to ważne urządzenia medyczne wykorzystywane w leczeniu niewydolności nerek, dlatego ich przestoje muszą być jak najkrótsze. Ciągłe, nieprzerwane korzystanie z tych urządzeń w szpitalach na całym świecie sprawia, że, w przypadku braku właściwej konserwacji, są one podatne na awarie. W związku z tym koszty konserwacji aparatów do dializy są wysokie. W prezentowanej pracy przedstawiono metodę wdrażania ekonomicznej strategii konserwacji. Wykorzystano strategię konserwacji opartą na analizie przyczyn źródłowych uszkodzenia (RCBM). Zastosowano ją na poziomie części składowych w celu optymalizacji harmonogramów konserwacji opartej na niezawodności (RBM) tworzonych na podstawie istniejących danych dotyczących konserwacji i uszkodzeń. Aby móc zminimalizować średni koszt konserwacji aparatów do hemodializy i zapewnić ich wysoką gotowość operacyjną, opracowano model kosztowy, a optymalizację przeprowadzono za pomocą algorytmu genetycznego. Zastosowanie strategii RCBM daje około 60-procentową oszczędność kosztów, jakie ponosi się przy użyciu obecnie wykorzystywanego programu konserwacji. Obliczenia statystyczne i optymalizacyjne wykonano, odpowiednio, przy użyciu oprogramowania Weibull ++ i MATLAB firmy Reliasoft.

WAWRZOSEK J, IGNACIUK S, BOCHNIAK A. **Wybrane problemy niejednoznaczności ceny dualnej wody w postoptymalizacyjnej analizie systemu wodociągów**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 329–340, http://dx.doi.org/10.17531/ein.2019.2.18.

W literaturze uważa się, że cena dualna wody jest obiektywną przesłanką do kształtowania rynkowej ceny wody. Jednak autorzy zauważają, że pojedynczy wektor cen dualnych w dystrybucji wody, gdy jest niejednoznaczny, nie powinien stać się podstawą do podejmowania decyzji zarówno normującej cenę wody jak i wpływającej na procedury modernizujące sieć wodociągową. Praca uczula inżynierów gospodarki wodnej by nie powielali powszechnych błędów oprogramowania oraz wskazuje jak, pomimo kompletnego braku literaturowych wskazówek, praktycznie rozwiązywać napotykane problemy techniczne. Liniowa zależność wektorów wierszowych parametrów lewych stron wiążących warunków ograniczających w modelu programowania liniowego dla zużycia wody identyfikowana jest tu jako przyczyna niejednoznaczności wektorów cen dualnych. Ta niejednoznaczność w zagadnieniach dystrybucji wody wymaga kształtowania alternatywnych scenariuszy technicznych pozwalających na wariantowy wybór sposobu modyfikacji systemu poboru wody. Dlatego opisano zasady wyznaczania proporcjonalności jednoczesnych zmian niektórych parametrów prawych stron warunków ograniczających. Na uproszczonym modelu dystrybucji wody sformułowano i wskazano te zasady optymalnego doboru najbardziej produktywnych wektorów dla metody parametrycznego programowania liniowego. Opracowana w pracy metodyka umożliwia m.in. wygenerowanie alternatywnych scenariuszy technicznych oszczędzania różnej ilości wody, skutkującej różnymi oszczędnościami finansowymi.

SIKORA M, SZCZYRBA K, WRÓBEL Ł, MICHALAK M. **Monitorowanie i utrzymanie suwnicy bramowej na podstawie bezprzewodowego systemu pomiaru i analizy poziomu drgań** Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 341–350, http://dx.doi.org/10.17531/ein.2019.2.19.

W artykule przedstawiono system monitorowania i diagnostyki suwnicy bramowej. Głównym zadaniem systemu jest akwizycja, wizualizacja i monitorowanie poziomu drgań newralgicznych elementów suwnicy. System wyposażony jest również w część obliczeniowo-analityczną, umożliwiającą realizację zadań predykcyjnego utrzymania ruchu (ang. predictive maintenance) związanych z oceną poziomu drgań. Architektura systemu umożliwia wykorzystanie go również do innych zastosowań, w których dla realizacji zadania diagnostyki wymagana jest bezprzewodowa sieć czujników drgań.

KARWAT B, MACHNIK R, NIEDŹWIEDZKI J, NOGAJ M, RUBACHA P, Stańczyk E. **Kalibracja modelu materiału sypkiego w Metodzie Elementów Dyskretnych na przykładzie perlitu D18-DN**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 351–357, http://dx.doi.org/10.17531/ein.2019.2.20.

Analityczne metody obliczeniowe parametrów konstrukcyjnych maszyn i urządzeń transportowych są często niewystarczające, zawłaszcza w przypadku transportu nietypowych materiałów sypkich. Pomocnym narzędziem numerycznym wspierającym proces projektowania i optymalizacji urządzeń do transportu materiałów sypkich jest Metoda Elementów Dyskretnych (DEM). Uzyskanie wiarygodnych wyników symulacji wymaga kalibracji parametrów wejściowych modelowanego materiału wykorzystując wyniki badań laboratoryjnych właściwości fizykochemicznych rzeczywistych materiałów. W artykule przedstawiono metodologię kalibracji modelu DEM na przykładzie perlitu D18-DN. W oparciu o przeprowadzoną kalibrację zaprezentowano możliwości zastosowania metody DEM do symulowania transportu materiału przenośnikiem ślimakowym.

Article citation info:
GOMÓŁKA L, AUGUSTYNOWICZ A. Evaluation of applicability of dielectric constant in monitoring aging processes in engine oils. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 177–185, http://dx.doi.org/10.17531/ein.2019.2.1.

Leszek GOMÓŁKA
Andrzej AUGUSTYNOWICZ

EVALUATION OF APPLICABILITY OF DIELECTRIC CONSTANT IN MONITORING AGING PROCESSES IN ENGINE OILS

OCENA PRZYDATNOŚCI STAŁEJ DIELEKTRYCZNEJ DO MONITOROWANIA PROCESU STARZENIA OLEJU SILNIKOWEGO*

This paper reports on the study involving the development of a dependence between the selected physicochemical properties of engine oil and the variation of its dielectric constant, considered as an evaluation parameter for the degree of its exploitation. In order to get to know the selected phenomena that accompany oil aging, a decision was made to perform tests of the condition of engine oils at various stages of their exploitation. In addition to the analysis of the dielectric constant, measurements of total base number and (TBN) and infrared spectroscopy (IR) were also performed. The interdependence between the examined values is developed using the Pearson's linear correlation coefficient. As a result, the variability of TBN and IR absorbance in the context of the change in the dielectric constant were evaluated. Following the realized experimental tests, it was found that the results demonstrate a satisfactory correlation between the selected physicochemical properties of the oil, which resulted in the adoption of an assumption of a high feasibility of application of dielectric constant as a diagnostic parameter in monitoring the state of exploitation of engine oil.

***Keywords**: oil exploitation, diagnostics of engine oil, dielectric constant.*

Celem pracy było określenie związków pomiędzy wytypowanymi właściwościami fizykochemicznymi oleju silnikowego a zmianą jego stałej dielektrycznej, rozważanej jako parametr oceniający stan jego zestarzenia. W celu szczegółowego poznania wybranych zjawisk towarzyszących starzeniu się oleju zdecydowano się na przebadanie oleju w różnych stadiach jego użytkowania. Poza badaniem stałej dielektrycznej wykonano również badania całkowitej liczby zasadowej TBN oraz badania spektrometryczne w podczerwieni IR. Stopień współzależności badanych wielkości przedstawiono wykorzystując do tego współczynnik korelacji liniowej Pearsona. W efekcie oceniono zmienność całkowitej liczby zasadowej oraz absorbancji IR w kontekście zmiany stałej dielektrycznej. Po przeprowadzeniu badań eksperymentalnych, stwierdzono, że uzyskane wyniki wskazują na zadawalającą korelację pomiędzy wytypowanymi własnościami fizykochemicznymi oleju, co pozwoliło z dużą dozą pewności przyjąć, iż stała dielektryczna może być użyta w charakterze parametru diagnostycznego do monitorowania stanu oleju silnikowego.

***Słowa kluczowe**: eksploatacyjne zużycie oleju, diagnostyka oleju silnikowego, stała dielektryczna.*

1. Introduction

In the era of developing technology, ICE forms the used sources applied to drive vehicles, machines and a variety of equipment. We can assume that despite its numerous drawbacks, ICE will remain to play the role of the dominant source of power for many years [9,12]. The lubrication system forms one of the most important systems in an engine, and a lubricating oil forms its integral part. In order to adequately verify the properties of oils during operation, there is a need to easily and accurately diagnose its quality. This fact leads to the need to adequately select a representative parameter by means of which it is possible to monitor changes occurring during its exploitation. Modern engine oils consist of various components and additives, by which a number of stringent operational requirements can be achieved and realized. Therefore, it is important to adopt and select such research methods and quantities that meet the requirements of a diagnostic parameter, can be applied to analyze the quality of the oil. The assessment of oil wear in an engine currently forms one of the most dynamically developing trends in research and is of interest to many scientific institutions [7]. This subject also forms the focus of interest of private companies operating vehicles due to the practical and economic outcomes that can be valuable to them [5,6]. All mechanisms or components of equipment are subjected to the processes of aging and wear so along with aging, engine oil changes its lubricating properties, which adversely affects the characteristics of the lubrication system. Heterogeneous conditions of vehicle and machine opera-

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

tion directly affect the intensity of changes occurring in the engine oil and; hence, individual approaches to oil quality assessment should be applied, taking into account the projected period of engine exploitation. For this reason, research was carried out with the purpose of selecting quantities that can effectively and accurately demonstrate the nature of exploitation changes and which can be easily measured and recorded.

On the basis of obtained characteristics of selected physicochemical properties of the oil and considerable level of a correlation between changes in dielectric constant and oil properties, a new methodology was proposed for the diagnostics of oil quality. To this aim, an approach to the analysis of the state of the lubricating oil was considered to be necessary.

This work focuses on demonstrating the relation between the physicochemical properties of the oil and the dielectric constant, which is an easily measurable quantity. The analysis involves the relations between the basic parameters applied to characterize the aging process of oils based on the measurements of the change in the dielectric constant of used oils in relation to the fresh ones. Due to the widespread use of electronics in vehicles, the procedures for diagnosing oil quality and methods of setting intervals of oil change were described in detail using guidelines developed by vehicle manufacturers' as well as on the basis of the logical algorithms applied for such purposes. On this basis, the development of a new method that can guide in the decisions regarding engine oil intervals was proposed as a secondary objective of the present study.

2. Initial assumptions

The review of selected service manuals shows that an oil change intervals form an ambiguous recommendation for vehicle manufacturer due to the lack of technical conditions applicable for assessing the suitability of its use in a vehicle. The advanced level of electrical and electronic measurements provides a variety of quantities and forms of control that can be applied in mechatronic systems [2,4]. However, the analysis of the state of knowledge in the area shows that the field concerned with aging of lubricating oil in internal combustion engines forms one of the challenges due to complex physicochemical processes occurring in oils. Aging processes can be delayed by adding enriching additives to oils during exploitation and such substances are known to affect anti-wear properties and increase the level of protection from engine dry running. The decrease in the value of total base number (TBN) is reported to be feasible due to the use of additives in a number of studies [10,11]. The measurements involving the total basic number belongs to routine tests as it forms a test used to determine the level of loss of alkaline reserve during operation. It can also be carried out to monitor the condition of oil additives.

The oil change intervals given by the manufacturers are often designed to occur well before engine oil starts to deteriorate its lubricating properties significantly. However, checking oil properties using analytical methods in a laboratory is expensive. In addition, the use of analytical techniques during normal exploitation can be extremely difficult. Therefore, oil is rarely diagnosed before scheduled replacement, which results in wasting huge amounts of oils.

During the exploitation of engine oil, many parameters are variable, including its kinematic viscosity, fractional composition, additive content, as well as impurity content. The kinematic viscosity test performed at 40 and 100°C demonstrates that during exploitation this value varies in a narrow range. The diagnostic parameter based on the viscosity parameter provides limited information, in particular in the conditions of varied temperatures.

Engine oil is a typical dielectric material. The study in [7] reports that dielectric spectroscopy forms one of the most powerful analytical technique applicable for measurements in dielectric materials. It can be easily used to classify oils according to their viscosity parameters. We can assume that dielectric spectroscopy can provide practical information regarding the composition and structure of the engine oil. Similarly as in [14], it was shown that the temperature coefficient of the dielectric constant of engine oils depends on whether oil is fresh or used. Therefore, an assumption was made that the measurement of the dielectric constant offers the potential to determine the condition of the engine oil.

A similar conclusion was made in [16] in which a principle of monitoring oil contamination over the Internet was proposed on the basis of the measurement of its dielectric constant. A measurement system was also developed including a capacitance sensor, a small capacitance detection circuit, as well as software applied for monitoring and data processing. The results show that the relative dielectric constant of oil can be effectively tracked and properly classified using a dedicated measuring system that can be used to determine oil change intervals.

The measurement of the dielectric constant of oils can provide valuable information regarding oil quality. A simple test dedicated to this purpose can be applied immediately to determine if the correct oil was used during a service procedure involving oil change. An increase of the dielectric constant of the oil indicates the presence of impurities or change in its chemical composition [3,15]. It was found that the rate of variation in the dielectric constant depends on the physical properties of the oil, which in turn are affected by deposition of impurities, such as density and viscosity, as well as the use of refining agents.

Synthetic and mineral oils used in engines are not perfect insulators as they are characterized by a minimal electrical conductivity. Therefore, the knowledge of dielectric properties, such as the relative electrical permittivity is important for the adequate performance of ICE engines. The monitoring of oil condition and its degradation in the online system and forecasting its replacement periods forms ones of the goals of the current automotive research [18]. Monitoring the oil condition and its degradation using an online system involving the analysis of its dielectric constant is associated with the need to get to know about its changes in the function of temperature. The reported tests [8,13] demonstrated a significant increase in the dielectric constant in the ranges of higher temperatures, which offers an easier measurement and classification of oil exploitation. The paper also analyzes the feasibility of using a capacitor sensor and its correlation with the dielectric constant. The present study found that along with the changes of the dielectric constant of the oil, its capacity changes as well.

Infrared spectrometry (IR) offers a tool that is practical in assessing oil quality and the changes occurring in it by application of spectral analysis. This is an instrumental technique based on the study of spectral analysis of electrons in chemical compounds. It is also a method used to determine the content of some chemical compounds and the measurements for such purposes are performed by application of a selected wavelength. As a result of the research, on the basis of obtained spectra and spectra of standards, the content of the compound in the tested substance can be derived. Infrared spectroscopy techniques are often applied to determine the content of undesirable components and the content of some additives [1].

Bearing the above in mind, a decision was made to design a scope of a study to include parameters such as: dielectric constant, total base number and absorbance levels.

3. Research program

For the purposes of the present study, five major steps were identified:

- identification of objects applied for extracting oil samples,
- development of sampling methodology,
- analysis of the influence of temperature on the variations in dielectric constant of oil,

– testing of samples in a laboratory,
– analysis of the results.

An important aspect was associated with the determination of an adequate representative amount of oil needed for laboratory analysis. For the purposes of comparison, tests to investigate the degradation of oil over a standard vehicle distance – 15,000 km have been determined in the first place. After each sampling (from 100 to 150 ml), the oil level was checked in the engine and it was supplemented to the level recommended by the manufacturer. A test drive was also carried out without adding oil after sampling to verify this type of operation, expecting faster oil aging. Due to the long-term duration of this type of study, a decision was made to perform only one such series on a single vehicle.

4. Objects of testing

The object applied for the testing involved six intermediate class passenger cars, normally driven in urban and non-urban environment, in particular used in everyday commute. The minimum daily distance that was covered by the cars was equal to around 40 km. The cars used petrol or operated in cycles where LPG was used alternatively with petrol and one car used diesel fuel. Each of the vehicles was used by a single driver. The vehicle engines applied semi-synthetic and synthetic oils.

The selection of objects applied for testing was also influenced by other factors, such as:
– common use of intermediate class cars in the domestic automotive market,
– varied age and mileage of vehicles,
– similar characteristics of exploitation (use in urban traffic, daily mileage, etc.),
– accessibility of vehicles for cyclic oil sampling.

Table 1 contains a summary of the basic data regarding the vehicles.

5. Laboratory testing

In addition to the assumptions adopted during the selection of test objects, the technical conditions for extraction and storing of engine oil samples for laboratory tests were defined. The program of the activities involved the following steps:
– sampling from the engine was always carried out when engine was fully heated; in cases when vehicle was parked for a longer period, sampling was preceded by a sufficient distance traveled to ensure thermal equilibrium state of the engine. This provided conditions for adequate mixing of the oil in the engine and caused water drainage from the oil,
– due to the high oil temperature at the time of the extraction, special care was taken in terms of health and safety,
– a syringe with a flexible hose was used for oil sampling,
– oil level indicator tube formed the place from which a constant, specific amount of oil was extracted for the study,
– a specific volume of oil was extracted in each case – ranging from 100 to 150 ml,

While maintaining the above guidelines, efforts were made to eliminate any errors during the sampling, in order to provide the same conditions for testing all vehicles. In order to store samples at various stages of oil exploitation, dedicated containers made of laboratory glass were purchased. They offer a very useful solution for oil storage until testing. The use of laboratory glass protects the oil from additional contamination and uncontrolled aging during storage.

Due to the selection of objects and research methodology, the analysis involved the changes of oil properties in the self-ignition (SI) and compression-ignition (CI) engines. Such an approach to the analysis had to be followed due to the actual loss of physicochemical properties, depending on the nature of use and technical condition of the vehicle, as well as progressive changes in oils following a specific mileage traveled. By conducting laboratory tests, the justification of adopting oil change intervals and the possibility of a quick and effective assessment of its state were verified on the basis of the selected diagnostic parameter.

The experimental setup comprised with the necessary equipment. Before each measurement, residual oil was cleaned from the sensor using extraction gas and it was cleaned with a paper towel and finally air-dried under pressure. This ensured reliable measurement as efforts were made to ensure that the Analexrs measuring electrodes did not contain residues of oil used in the previous tests. Another very important aspect of the research was associated with ensuring the homogeneity of the samples oil. Due to the need to store collected samples before testing, undesirable sedimentation occurred in the oils, which led to settling of residues and its delamination. Since such processes are imperceptible to the naked eye, so to ensure the homogeneity of the suspension, the oil samples were shaken gently to prevent the effect of aging before each of the test series.

The scope of research included:
– measurement of the dielectric constant using Lubri Sensor
– measurement of the dielectric constant using Analexrs sensor
– variability of the total base number (TBN),
– infrared spectrometry, variability in the level of absorbance – IR.

Table 1. Summary of data on vehicles used in the study

No.	Brand	Fuel	Engine capacity [dm^3]	Type of oil	Mileage [km]
1	Opel Astra 1.7D	Diesel	1.7	Lotos synt. 5W/40; API SN/CF, ACEA A3/B4	No available data
2	Daewoo Nexia 1.5	Petrol + LPG	1.5	Lotos synt. 5W/40;API SN/CF, ACEA A3/B4	98 400
3	Fiat Punto 1.1	Petrol	1.1	Elf Semi-Syntetic 10W/40;API SL/CF, ACEA A3/B4	81 700
4	Opel Astra 1.6 *Testing performed in two series: S1 and S2*	Petrol + LPG	1.6	Genuine GM 10W/40;API SL/CF, ACEA A3/B3	125 000
5	Volkswagen Passat 1.8	Petrol + LPG	1.8	Castrol Magnatec 10W/40;API SL /CF, ACEA A3/B3	242 038
6	Fiat Seicento 1.1 *Testing performed in two series S1 and S2*	Petrol	1.1	Mobil 1 Formula S 10W/40;API SL/SJ/CF, ACEA A3/B3	99 382

The equipment applied in the measurements oil quality by determining the oil parameters, also known as the dielectric constant, included: Analexrs mechatronic sensor and the Lubri-Sensor portable meter shown in Figs. 1 and 2. The variations in IR absorbance as a function of the course were examined by application of Specord M80 mesh spectrophotometer (Fig. 3).

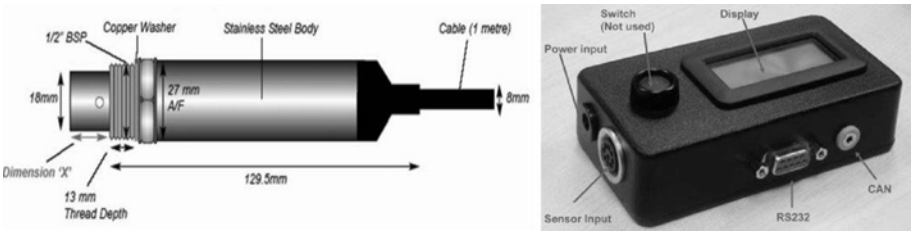


Fig.1. Analexrs sensor suite for measuring dielectric constant [19]

Fig. 2. Lubri Sensor applied for measurements of dielectric constant [20]

Fig. 3. General view of Specord M80 spectrophotometer [21]

6. Multi-aspect analysis

The essence of the concept reported in this paper involves the analysis of oil degradation based on the variations in the dielectric constant in the conditions of selected exploitation cycles. Such an assumption was connected with the necessity of a thorough study of the influence of various physicochemical properties of the aging processes in oils and determining changes in dependencies between them. Despite many studies, comprehensive insight in this area has not been gained and exiting knowledge needs to be extended, in particular due to the use of new oil components, which in principle should ensure the highest level of technical and ecological parameters while maintaining minimum production costs.

6.1. Dielectric constant

The measurement of dielectric constant was made on the basis of two sensors: Lubri and Analexrs. The characteristics of the tested parameter are outlined below.

The summary of data in charts in Fig. 4 contains the characteristic variations in dielectric constant in the particular objects used in the study. For the analysis in question, the reference was expressed as the mileage of the vehicle after which the testing of oil was carried out in individual objects.

When we analyze the diagrams representing the changes in the dielectric constant measured by two independent sensors, we found that a significant increase in the dielectric constant during use was very evident. Small differences result from the applied technique of measurement and the nominal scale of the sensors, as well as from the time intervals of the performed tests. A similar trend was observed in the characteristics of the two series of tests performed on the same Fiat Punto power unit, measured by Analex and Lubri sensors. The same applies to the function (qualitative similarity) in the two test series using Opel Astra 1.6. For the S1 series in the last car, faster oil degradation was recorded in cycles without oil refills, which resulted in earlier intervals after which values associated with the need to change oil were obtained.

This is proof that it is necessary to check the oil level in the engine during use in order to maintain the required quality of lubrication. This ensures fault-free operation of the lubrication system, and refills refresh the oil and slow down the process of the gradual oil concentration – after reaching the mileage of 25,000 km, the dielectric constant has reached the level of 65 units (Analexrs) for the Opel Astra 1.6 following oil refills. The characteristic variations of the adopted value of the diagnostic parameter are characterized by considerable levels of correlation coefficients, ranging from R^2=0.87 to 0.98.

6.2. Total base number

The measurement of the total base (TBN) number was performed by potentiometric titration according to PN–88/C-04049 standard. It is clear that the use of oil additives determines the appropriate level of the total base number during exploitation. For this reason, oils of different types cannot be mixed. This action leads to the fact that the oils in the engine can react with each other, causing a reduction in the total base number and formation of inert and even harmful chemicals for the engine. The characteristic variations in this quantity in the tested vehicles are illustrated below in Figure 5.

The results derived on the basis of the analysis of TBN in the oil during its exploitation found that the initial value of the alkaline number representing the alkaline reserve for the further period of operation forms the most important aspect. In the presented results, the initial value of TBN is in the range from 8 to 13 [mgKOH/g] and depends on the type of engine oil and its manufacturer – the higher the initial value, the longer the service life of an oil. For Fiat Punto, after comparing both series of data, a very similar loss of alkaline reserve (measured by a decrease in the base number) was observed, which most probably indicates a similar type of vehicle exploitation? For Opel Astra 1.6 (in the first test series), the effect of accelerated aging of the oil was recorded as no refills were made in this case after each oil sampling. The second test series using the Opel Astra 1.6 car served as the source of reference. During this series, the volume that was extracted was subsequently supplemented with fresh oil after each extraction.

An interesting example is offered by the course of variation in the total base number for Opel Astra 1.7 diesel. For this vehicle the lowest decrease in the alkaline reserve was recorded. Due to the fact that

this was the only vehicle with a SI engine applied in the tests, it is not possible to clarify this fact more extensively.

Another interesting combination is offered by the pair of vehicles, i.e. VW Passat 1.8 running on petrol and LPG and Fiat Seicento 1.1 driving only on petrol. By comparing the courses in Fig. 5 for these two cars, wed can see a sharp decrease in the total base number for the VW vehicle in comparison to Fiat, whose total base number was lower by 0.52 mgKOH/g at the beginning of the test. The phenomenon can be explained by the fact that the vehicles differed in the engine capacity and moreover, by the fact that the VW Passat had a significantly greater mileage from Fiat Seicento at the beginning of the study (Table 1).

The variations in the value of the total base number during exploitation, i.e. the reduction of the alkaline reserve is a commonly recognized process, and this progress depends mainly on the type of exploitation and condition of the powertrain, i.e. the basic factors affecting the degradation of the oil. An initial analysis of the vehicles could as well be performed in terms of the type of applied engine oil, which is characterized by different values of the initial TBN. Bearing in mind the above, a parameter was developed for the purpose of comprising the degree of the variation in TBN between specific vehicles. Only the boundary parameters of the changes in TBN were taken into account for his purpose, i.e. TBN_{max} and TBN_{min}, after which the result was compared to the mileage traveled on a particular oil. For a broader analysis, we can consider a change in the particular periods of use.

The calculation of ΔTBN applied the formula:

$$\Delta TBN = TBN_{max} - TBN_{min} \tag{1}$$

The change in per cent was derived from the formula:

$$zTBN = 100\% - \frac{TBN_{min} \cdot 100\%}{TBN_{max}} \tag{2}$$

On the basis of a comparing of the percentage change in the total base number (representing the loss of the alkaline reserve in per cent) in relation to the vehicle mileage, we are able to compare all examined vehicles to each other (Table 2). Knowing the guidelines of the manufacturers of the particular oils as to the moment of oil change in the context of a decrease in the alkaline reserve, we are able to determine this instant in time by correlation with the course of the vehicle or time period (motohours). At this point, it is also possible to determine whether the period specified by a given manufacturer resulting from the mileage or operating hours forms a correct determinant of service operations.

The analysis demonstrates that in some cases the adopted distance is too short on the one hand for Fiat Punto and on the other hand too short when looking at the VW Passat, where the base number after the distance of 10,000 km reached the value of 3.34 [mgKOH/r], which means the loss of 60% of alkaline reserve. The result is a result for Opel Astra 1.6 in the second test series, where despite a low initial value of the total base number, a significant decrease of around 78% in the alkaline reserve was recorded for the series during which oil was supplemented with fresh oil after sampling.

The comparisons also show that an important parameter of the oil is associated with its initial TBN, which means that engines should use selected quality oils with accurately controlled initial values of TBN. The permissible TBN limits in oil are regulated by the ACEA (Association des Constructeurs Européens d'Automobiles) and depend on the type of fuel. Vehicles driving on petrol, LPG and CNG as well as ultra-low-sulfur diesel oil do not require high TBN. On the other hand, when the engine is driving on a biofuel, there is a need to select of oil with a higher TBN [17].

After a detailed analysis of the change in the total base number, the research demonstrates that it is therefore advisable to compile the measurement results with the diagnostic parameter determined using the Lubri Sensor and Analexrs index.

6.3. Distribution of spectrum radiation using infrared spectroscopy (IR)

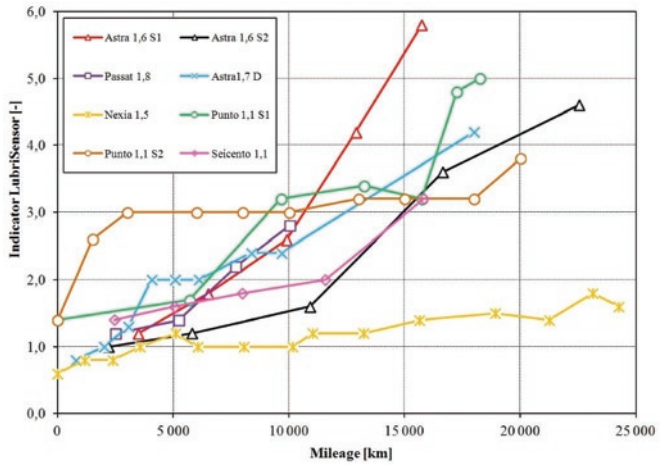
The analysis Spectral distribution (absorbance) of infrared radiation for motor oils proves most effective in the oxygen band, i.e. in the wavenumber range around 1824-1520 [cm^{-1}]. In this range, significant changes occur during exploitation, and they can be relatively easily examined and subjected to a further analysis. The basic conditions for recording IR spectrum are: SPECORD M80 spectrophotometer, measuring range: 1860 – 1480$cm^{-1,}$ layer thickness: 0.105mm, resolution: 4cm^{-1}, fresh oil used as reference selected was selected on the basis of manufacturers' recommendations as most suitable for each vehicle. The recorded IR spectra were calculated according to the baseline based on points 1764cm^{-1} and 1516cm^{-1}, and then numerically integrated within the boundaries determined by the base points. Fig. 6 contains a summary with the variations in absorbance levels for individual vehicles and oils. Measurement and analysis of infrared spectra (IR) offers the presentation of the dynamic characteristics of thermooxidative changes in oil during exploitation, but does not meet the criteria of a diagnostic parameter. It is either very difficult or impossible to determine the boundary values to be applied for qualifying the oil for replacement (especially for the Opel Astra 1.7D), because the changes in the absorbance value do not have characteristic points in the curves that demonstrate the change in its condition.

Fig. 6 also contains information regarding the changes in the absorbance value for the Fiat Punto (in two series) that significantly deviate from the curves for other cars. The existence of the fact can be explained by the fact that calculations made for the Fiat Punto were carried out in a wider bandwidth than for other test objects.

Table 2. Variation in total base number

Vehicle	Total base number TBN_{max} [mgKOH/g]	Total base number TBN_{min} [mgKOH/g]	ΔTBN [mgKOH/g]	Percenetage change zTBN [%]	Distance [km]
Punto 1.1 series 1	12.59	7.69	4.90	38,92	18250
Punto 1.1 series 2	12.78	8.54	4.24	33,18	20000
OpelAstra 1.7D	11.58	10.45	1.13	9,76	17983
Daewoo Nexia 1.5	10.2	7.04	3.16	30,99	24259
Astra 1.6 series 1	8.22	1.94	6.28	76,40	15730
Astra 1.6 series 2	8.22	1.79	6.43	78,22	22529
VW Passat 1.8	8.36	3.34	5.02	60,05	10038
Fiat Seicento 1.1	7.84	4.70	3.14	40,05	15828

a)



b)

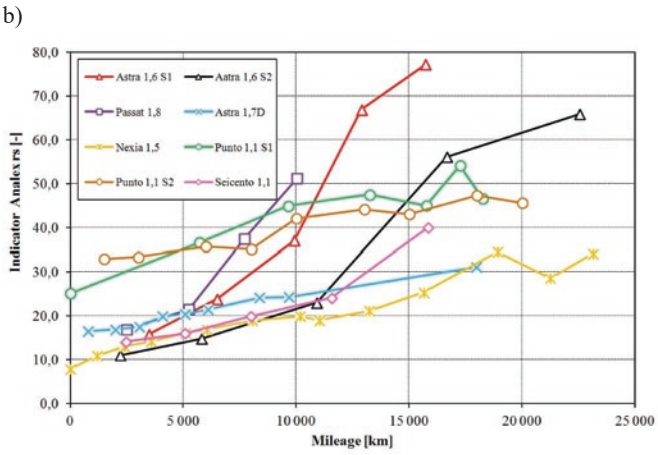


Fig. 4. Variation in dielectric constant measured by application of: a) Lubri, b) Analexrs sensors

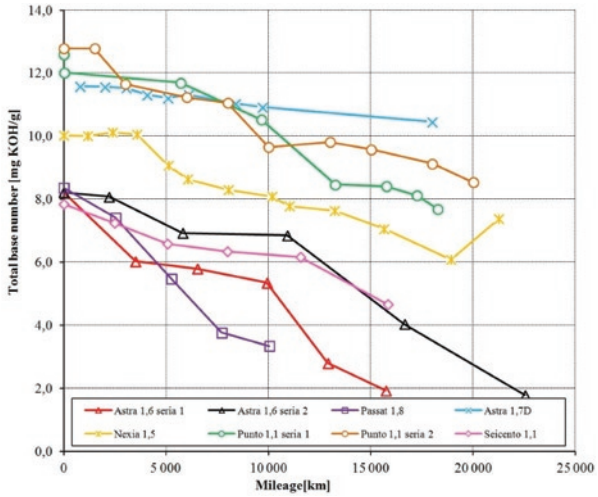


Fig. 5. Variation in total base number

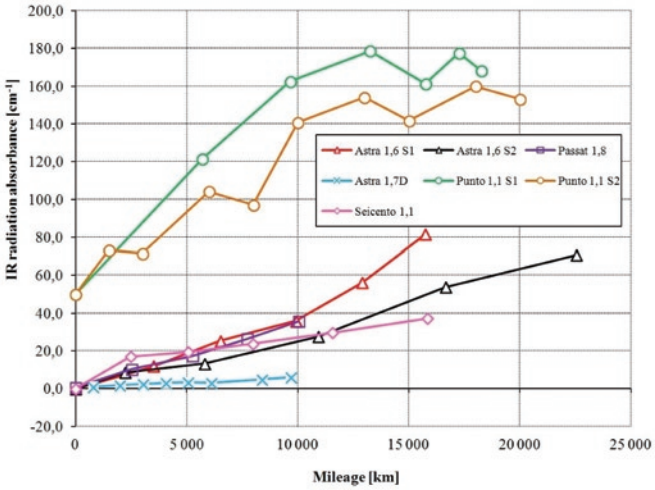
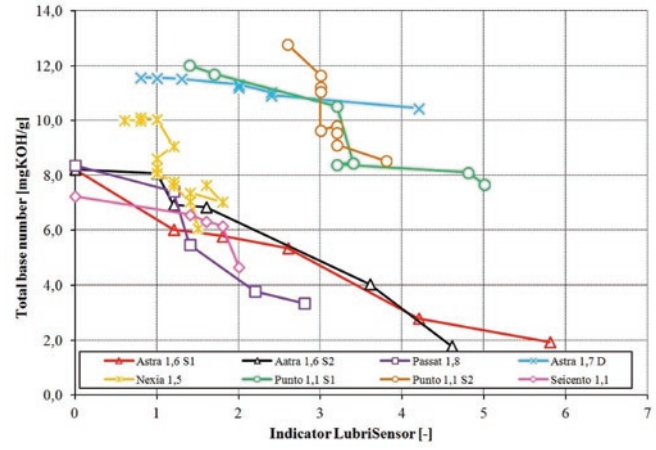


Fig. 6. Distribution of the spectrum of infrared radiation

a)



b)

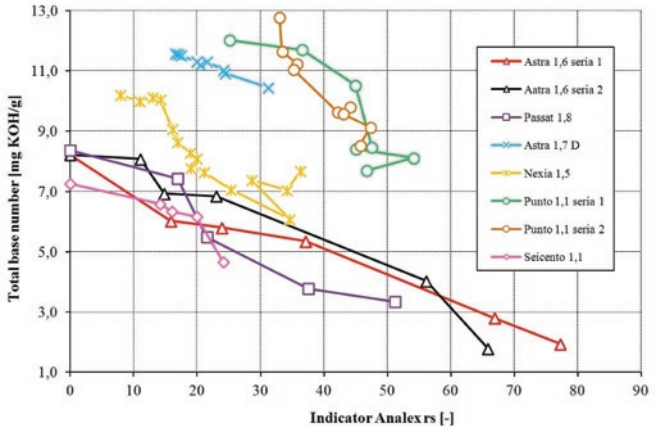
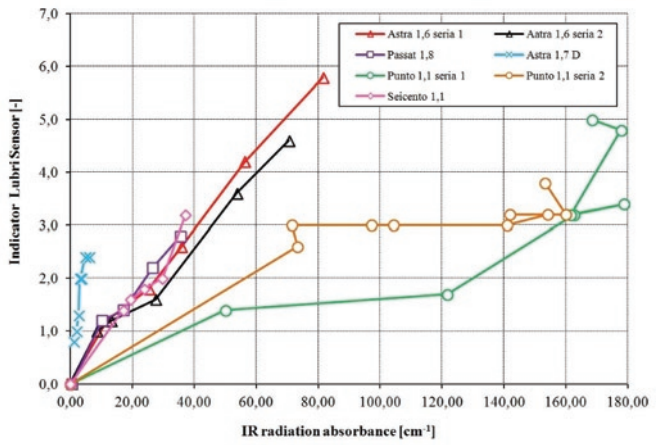


Fig. 7. Variation in dielectric constant compared to the total base number measured by application of:a) Lubri Sensor b) Analexrs

6.4. Comparative analysis

All measurements of the analyzed properties of engine oil were performed with a level of error resulting from the inaccuracy of the measurements of the applied equipment. For these reasons, it is impossible to determine the accurate value of the measured quantity and therefore it is important to assess the reliability of the results of the measurements. Therefore, in order to better visualize the variability and determine the correlation between the properties of engine oil during its exploitation, it was necessary to perform a comparative analysis. The degree of interdependence is presented using the Pear-

a)



b)

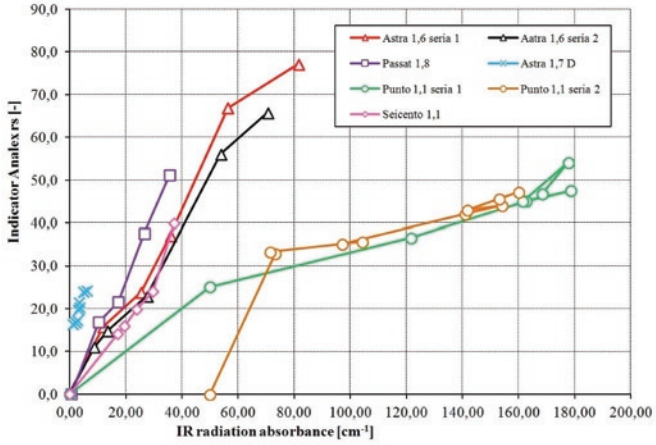


Fig . 8. Variation in infrared absorbance IR compared to the total base number measured by application of: a) Lubri Sensor b) Analexrs

son's linear correlation coefficient. For this purpose, the variations of the total base number (Fig. 7) and IR absorbance (Fig. 8) in the context of the dielectric constant were compared with the aim of its application as a diagnostic parameter.

Pearson's linear correlation coefficient forms a coefficient that defines the degree of linear dependence between random variables. A correlation was used to investigate whether there is a relationship between two variables (properties, quantities). The correlation coefficient tells us about the strength of the relation. It is defined as a value in the range from 0 to 1. The value of the coefficient tends to 1; the strength of the relationship is greater. The strength of a correlation in the range:

below 0,2– denotes a weak correlation (virtually no dependence),
0.2 ÷ 0.4 – low correlation (moderate level of dependence),
0.4 ÷ 0.6 – moderate correlation (significant level of dependence),
0.6 ÷ 0.8 – high correlation (very significant dependence),
0.8 ÷ 0.9 – very high correlation (very large level of dependence),
0.9 ÷ 1.0 – the dependence is almost full.

Pearson correlation coefficient r_{xy} was applied to the study of linear dependence of the examined variables in which an increase in the value of one of the variables results in a proportional change in the mean values of the other quantity (i.e. either increase or decrease).

This coefficient was calculated based on the formula:

$$r_{xy} = \frac{cov(x,y)}{Sd_x \cdot Sd_y} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \overline{x_x})^2 \cdot \Sigma(y_i - \overline{x_y})^2}}. \quad (3)$$

Tables 3 to 8 present a comparison of Pearson's linear correlation coefficients for individual vehicles.

The numbers presented in the Tables indicate that a high and very high level of dependence was determined between the analyzed variables. A detailed analysis of the results obtained leads to the following conclusions:

Table 3. Summary of correlation coefficient for Fiat Punto 1.1 – series No. 1

Fiat Punto 1,1 (1)	Lubri-Sensor factor [-]	Analex rs factor [-]
Analex rs factor [-]	0.814	1.000
Total base number [mgKOH/g]	0.823	0.715
IR absorbance [cm^{-1}]	0.804	0.949

Table 4. Summary of correlation coefficient for Fiat Punto 1.1 – series No. 2

Fiat Punto 1.1 (series no. 2)	Lubri-Sensor factor [-]	Analex rs factor [-]
Analex rs factor [-]	0.888	1.000
Total base number [mgKOH/g]	0.72	0.897
IR absorbance [cm^{-1}]	0.731	0.692

Table 5. Summary of correlation coefficient for Opel Astra 1.6 – series No. 1

Opel Astra 1.6 (series no. 1)	Lubri-Sensor factor [-]	Analex rs factor [-]
Analex rs factor [-]	0.982	1.000
Total base number [mgKOH/g]	0.966	0.974
IR absorbance [cm^{-1}]	0.995	0.971

Table 6. Summary of correlation coefficient for Opel Astra 1.6 – series No. 2

Opel Astra 1.6 (series no. 2)	Lubri-Sensor factor [-]	Analex rs factor [-]
Analex rs factor [-]	0.992	1.000
Total base number [mgKOH/g]	0.960	0.953
IR absorbance [cm^{-1}]	0.984	0.988

Table 7. Summary of correlation coefficient for Volkswagen Passat 1.8

Volkswagen Passat 1.8	Lubri-Sensor factor [-]	Analex rs factor [-]
Analex rs factor [-]	0,926	1,000
Total base number [mgKOH/g]	0,903	0,916
IR absorbance [cm^{-1}]	0,974	0,990

Table 8. Summary of correlation coefficient for Fiat Seicento 1.1

Fiat Seicento 1.1	Lubri-Sensor factor [-]	Analex rs factor [-]
Analex rs factor [-]	0.982	1.000
Total base number [mgKOH/g]	0.927	0.963
IR absorbance [cm^{-1}]	0.962	0.946

– The measurement of the dielectric constant was performed by application of two sensors that differ in terms of the design. One that is known as Lubri Sensor and can be considered as a portable meter whereas Analexrs form a modern sensor that can be combined with a data acquisition system to monitor the relative permittivity in real time. Despite the different designs of the applied equipment and their measurements, a very high correlation coefficient in the range of 0.81÷0.99 was obtained.
– The methodology applying infrared spectroscopy offers the analysis of its composition and observation of the variations in quantities that occur throughout oil aging. However, selection of an unambiguous parameter characterizing the quality of oil in a given case is very difficult. Given the above and the fact that testing involves the need to apply expensive equipment, the use of this method in a vehicle is unfeasible. The calculated correlation between the values of the absorbance of radiation and the dielectric constant is in the range of 0.73÷0.99 for Lubri Sensor and 0.69÷0.99 for Analexrs, respectively.
– As we already mentioned earlier, the decrease in the alkaline reserve accompanying the exploitation of engine oil affects the degradation of oil. At the moment referred to as the point of equilibrium, the alkaline reserve is exhausted to the degree that prevents the current neutralization of acidic products of combustion and the replacement of oil with fresh oil is most desirable. The determination of the instant when oil needs to be changed with on-board diagnostic methods is very difficult. The calculated correlation between the values of the total base number and the dielectric constant is in the range of 0.72÷0.97 for Lubri sensor and 0.71÷0.97 for Analexrs sensor, respectively.
– The measurement of the dielectric constant with the Analexrs sensor or a similar sensor installed in the lubrication oil system of an ICE could provide a source of diagnostic information – to play the roles of a parameter of the condition of engine oil.

7. Concept of original method of predicting instant of oil exchange

The final stage of the study took the form of an attempt to develop a systematic methodology for forecasting oil change intervals. In this approach, it was considered to take into account all the necessary information that is needed to determine an optimal time to change the engine oil, in particular:

– parameters of fresh oil, applied or selected (quality class, viscosity class, value of TBN, etc.) that determine the starting point of the analysis,
– conditions of vehicle use and driver profile (engine speed range, engine load, coolant temperature, oil temperature and level, ambient temperature, oil pressure in various parts of the lubrication system, number of cold starts, distance traveled, etc.),
– measurements from sensors dedicated to determine oil condition (dielectric constant, TBN, viscosity or other that are applied in the market).

A control algorithm was responsible for identification of the instant corresponding to the optimal instant corresponding to the oil change. Such control algorithm would apply the available information from the data communication network with continuous data input from the vehicle and on-board sensors. This data should contain information on the level of oil degradation (and form degradation characteristic), such as:

- value of alkaline reserve loss in per cent,
- the value or level of the TBN,
- dielectric constant as a value or a percentage change,
- period of service work (kilometers, hours, etc.).

The system would take into account corrections from on-line measurements as well as given as reference parameters (commonly referred to as rough use or operation in difficult environment), by means of which oil change interval would be determined, such as:

- driver or operator profile,
- type of use,
- trend of changes in measured values,
- ambient conditions, etc.

All forecasting would be based on a real-time system that can be easily implemented in the calculation of neural networks so as to develop an artificial intelligence system. However, we can emphasize that the decisive role is taken on by the measurement of dielectric constant which, as a universal quantity, takes into account most factors affecting oil degradation (fuel, soot, sludge, lacs, gels, ferromagnetic abrasion, water, etc.) and strongly correlates with the quantities that are used for accurate laboratory measurements (spectrometry, total base number, volume of pollution produced, etc.). Any characteristic representing the changes of individual physicochemical quantities in oils derived in a correlation using a selected measured parameter can be combined in the logic of such a system.

In the age of up-to-date trends and technologies implemented in the design of lubrication systems for vehicles and machines, such as:

– oil nanotechnology – modern oil additives applied to improve not only lubricating but also capable of maintaining specific operating parameters at the projected level and rebuilding the structure of friction pairs,
– downsizing – a small volume of oil in the oil sump, high operating temperatures, high power output from small engine capacity,
– modular design of the oil system, where the used oil and filter are exchanged with an completely new module,
– on-board oil cleaning systems.

We have to realize that accurate diagnostics plays an increasing role and its role continuously gains in importance. Oil degradation depends on many factors and the proposed conceptual method is based on the analysis of several variables, which is why it is included in the proposed intelligent system. A huge amount of input data is needed for the algorithm to operate effectively. It is not enough to monitor only one of the properties of the oil and to select the oil change interval on its basis.

This concept forms a new category of vehicle use that can be only implemented in modern hybrid and autonomous vehicles. This will ensure ease of operation, effective recycling and minimization of waste as well as meeting ecological standards.

8. Conclusions

The mileage as well as the duration of the vehicle's exploitation do not offer feasible parameters for determining oil change intervals. It is necessary to establish an easy and cheap diagnostic method that can be applied with regard to engine oils. The objective during the development of such a method would involve the application of the potential information of physicochemical properties of oils to eliminate ones that have lost their lubricating properties and no longer provide engine protection. With the aim of the insight into aging of oil, an analysis concerned with selected phenomena accompanying the various stages of oil use was performed. The detailed conclusions are as follows:

– The study of IR spectra confirmed the characteristics of thermo-oxidative degradation of oil during exploitation.
– The alkaline reserve test demonstrated the tendency to decrease reserve with exploitation, and indicates the increase of the acidic environment of an aging oil.

– A change in the relative permittivity accompanies the aging of oils. The dielectric constant forms an easily measurable parameter. The measurement of the dielectric constant with a sensor installed in the lubrication oil system of ICE could provide adequate diagnostic information.
– The study of the correlation of variability in the dielectric constant with respect to variables such as the TBN and IR absorbance demonstrates a high level of dependence, which means that dielectric constant can be successfully applied as a diagnostic parameter for monitoring the condition of engine oil.
– Determination of the optimal moment of oil change requires the development of a methodology for forecasting oil change applying systematic measures.

References

1. Al-Ghouti M A, Al-Toum L. Virgin and recycled engine oil differentiation: A spectroscopic study. Journal of Environmental Management 2009; 90, https://doi.org/10.1016/j.jenvman.2007.08.018.
2. Bolognesi P, Bruno O, Landi A, Sani L, Taponecco L. Electric Machines and Drives for X-by-Wire System in Ground Vehicles 2003. 10th European Conference on Power Electronics and Application: 1-16.
3. Carey A A, Hayzen A J. The dielectric constant and oil analysis Practicing Oil Analysis Magazine 2001; 9.
4. Cheng H. Autonomous Inteligent Vehicles: Theory, Algorithms and Implementation. London Limited: Springer-Verlag, 2011, https://doi.org/10.1007/978-1-4471-2280-7.
5. Gomółka L. Ocena oleju silnikowego w eksploatacji. Praca doktorska, Wydział Mechaniczny, Politechnika Opolska, 2017.
6. Gomółka L, Augustynowicz A, Maciąg A. Analysis of the rank spreading the decline of oil in internal combustion engines. Comustion Engines 2011; 1 (144): 1-4.
7. Guan L, Feng X L, Xiong G. Engine lubricating oil classification by SAE grade and source based on dielectric spectroscopy data. Analytica Chimica Acta 2008; 628: 117-120, https://doi.org/10.1016/j.aca.2008.09.004.
8. Han Z, Wang YQing X. Characteristics Study of In-Situ Capacitive Sensor for Monitoring Lubrication Oil Debris. Sensors 2017; 17 (12) 2851: 1-13.
9. Idzior M. The development of combustion engines in the aspect of their manufacturing. Comustion Engines 2006; 1 (124): 60-70.
10. Laber A, Adamczuk K. Improving the working conditions of friction by using operation additives to lubricating oil. Tribologia 2014; 1: 27-39.
11. Laber S, Laber A. Property assessment lubricants and tribological lotos dynamic engine oil. Tribologia 2015; 3: 89-97.
12. Serrano J R. Imagining the Future of the Internal Combustion Engine for Ground Transport in the Current Context. Applied Sciences 2017; 7 (10) 1001: 1-5.
13. Stevan Junior S L, Paiter L, Galvao J R, Roque D V, Chaves E S. Sensor and Methodology for Dielectric Analysis of Vegetal Oils Submitted to Thermal Stress. Sensors 2015; 15 (10): 26457-26477, https://doi.org/10.3390/s151026457.
14. Torrents J M, Pallas-Areny R. Sensing oil condition through temperature coefficient of dielectric constant. 2003 XVII IMEKO World Congress - Metrology in the 3rd millennium, Dubrovnik, Croatia: 917-919.
15. Patch R, Kamenka D, Menzel J. Return voltage measurements Diagnostic interpretations based on the dielectric time constants. Materials Science-Poland 2009; 27 (4): 1157-1169.
16. Yang Y, Yang D, Hu Z, Zhang X. Oil Contamination Monitoring Based on Dielectric Constant Measurement. 2009 International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Zhangjiajie, Hunan, China: 245-252, https://doi.org/10.1109/ICMTMA.2009.588.
17. Zhmud B. TBN, What`s in It for me?, Lube Magazine online 2018; 4.
18. Zhu J, Yoon J M, He D W, Qu Y, Bechhoefer E. Lubrication oil condition monitoring and remaining useful life prediction with particle filtering. International Journal of Prognostics and Health Management 2013; 4 Special issue 2: 1-15.
19. kittiwake.com/online-sensors
20. pmlubricants.com.au.
21. specord-service.de

Leszek GOMÓŁKA
JAWO TECH S.R.O. Sp. z o.o., Branch in Poland
Niepodległości 67, 44-370 Pszów, Poland

Andrzej AUGUSTYNOWICZ
Faculty of Mechanical Engineering
Opole University of Technology
Mikołajczyka 5, 45-271 Opole, Poland

E-mails: leszek.gomolkaa@gmail.com,
a.augustynowicz@po.opole.pl

Article citation info:
SUN K, WANG G, LU Y. Optimization method of bevel gear reliability based on genetic algorithm and discrete element. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 186–196, http://dx.doi.org/10.17531/ein.2019.2.2.

Kangkang SUN
Guoqiang WANG
Yanpeng LU

OPTIMIZATION METHOD OF BEVEL GEAR RELIABILITY BASED ON GENETIC ALGORITHM AND DISCRETE ELEMENT

METODA OPTYMALIZACJI NIEZAWODNOŚCI PRZEKŁADNI STOŻKOWEJ Z ZASTOSOWANIEM ALGORYTMU GENETYCZNEGO I ELEMENTÓW DYSKRETNYCH

Gear transmission is the most basic transmission component in mechanical transmission system. Many scholars have done a lot of research on gear reliability. When the variation coefficient is used to calculate and optimize the reliability of bevel gear, in order to calculate the reliability of bevel gear, it is often assumed that the gear works under constant torque, that is, the coefficient of variation (COV) is zero, but this is not the case in practice. In this paper, a gear reliability method based on discrete element simulation is proposed. The purpose of this method is to simulate the actual working conditions of gears, calculate more accurate coefficient of variation in the real world, and improve the accuracy of gear reliability design. Firstly, the real working conditions of the bevel gear transmission are simulated by discrete element method (DEM), and in the transmission system, the tangential force COV of the bevel gear is proved to be equal to the torque COV of the crusher central shaft. Secondly, the multi-objective function model of the gear transmission system is established based on the double tooth roll crusher (DTRC). The optimal volume and reliability of the bevel gear transmission are taken as the objective function, and the teeth number, module and face width factor of basic parameters of gear are optimized by genetic algorithm (GA). Finally, the accuracy of the optimization results is verified by Monte Carlo method. The main purpose of the manuscript is to analyse the effect of actual conditions (DEM simulation) on the optimization results. The results show that the COV of nominal tangential load of bevel gear is about 0.65 under actual working conditions, so in order to guarantee the same reliability, total volume need to be increased by 34.4%. This method is similar to the selection of gear safety factor. In practical production, the selection of safety factor is often based on experience. This paper provides a new method to optimize the reliability of bevel gear, combining with DEM simulation, which provides theoretical guidance for optimal design of bevel gear.

Keywords: *bevel gear, reliability, discrete element method, monte carlo simulation, double tooth roll crusher.*

Przekładnia zębata to podstawowy element mechanicznego układu napędowego. Niezawodność przekładni jest przedmiotem wielu badań. Przy obliczeniach i optymalizacji niezawodności przekładni stożkowej z wykorzystaniem współczynnika zmienności, często przyjmuje się, że przekładnia pracuje w warunkach stałego momentu obrotowego, t.j. że współczynnik zmienności wynosi 1. Sytuacja taka jednak nie występuje w praktyce. W niniejszej pracy zaproponowano metodę optymalizacji niezawodności przekładni opartą na symulacji metodą elementów dyskretnych. Celem tej metody jest zasymulowanie rzeczywistych warunków pracy przekładni, dokładniejsze obliczenie rzeczywistego współczynnika zmienności oraz poprawa dokładności projektowania niezawodności przekładni. W pierwszej kolejności, na przykładzie kruszarki podwójnej, wyznaczono model działania układu przekładni stożkowej wykorzystujący wielokryterialną funkcję celu. Optymalną objętość i niezawodność przekładni stożkowej przyjęto jako funkcje celu. Następnie, za pomocą metody elementów dyskretnych, symulowano rzeczywiste warunki pracy przekładni. Wyznaczono moment obrotowy przekładni stożkowej i współczynnik zmienności siły wypadkowej, a podstawowe parametry koła zębatego: liczbę zębów, moduł zęba i współczynnik szerokości zębów, zoptymalizowano za pomocą algorytmu genetycznego. Trafność wyników optymalizacji weryfikowano metodą Monte Carlo. Wyniki pokazują, że badana metoda może skutecznie poprawiać niezawodność przekładni stożkowej.

Słowa kluczowe: *przekładnia stożkowa, niezawodność, metoda elementów dyskretnych, symulacja Monte Carlo, kruszarka podwójna.*

1. Introduction

Gear transmission is one of the most common transmission in mechanical systems, and is widely used in various precision mechanical transmission components, such as machine tools, vehicles, etc. In mechanical systems, the basic requirement of gear mechanism is accurate and smooth transmission of motion, considering the requirements of manufacturing costs and convenience, in the premise of ensuring the requirements of gear transmission. The optimal design of the volume or weight of the wheel drive system has become an important subject for many scholars [2, 3, 4, 7, 42, 43].

With the rapid development of computer technology and various optimization algorithms, there have been a lot of research results on the optimal design of gear system. The methods and objects of optimal design have been continuously enriched [22, 38, 49]. Mendi [25] used GA to optimize the module of spur gear and then the optimal size of gear box shaft and rolling bearing is obtained. Savsani [34] used the particle swarm optimization and simulated annealing algorithm to optimize the design of multistage spur gear. Sa'id Golabi [13] established the objective function and constraint conditions of the optimum design of the volume / weight of the gearbox, compiled the optimization program using MATLAB, and verified the practicability of the design result by comparison with the known gearbox.

Nomenclature

Z_M	midpoint factor
Z_H	node region factor
Z_E	elasticity factor $\left(\sqrt{\mathrm{N/mm^2}}\right)$
Z_{LS}	stress correction factor
Z_β	helix angle factor
Z_K	bevel gear factor
K	calculating factor
K_A	application factor
K_V	dynamic factor
$K_{H\beta}$	face load factor
$K_{H\alpha}$	transverse load factor
F_t	nominal tangential load (N)
d_m	reference diameter (mm)
l_{bm}	length of contact line in the middle of gear (mm)
μ	gear ratio
σ_{Flim}	nominal stress number from reference test gears ($\mathrm{N/mm^2}$)
σ_{FP}	allowable bending stress ($\mathrm{N/mm^2}$)
Y_{NT}	life factor
$Y_{\delta relT}$	relative notch sensitivity factor
Y_{RrelT}	relative surface factor
Y_X	size factor
σ_{Hlim}	experimental surface contact fatigue strength ($\mathrm{N/mm^2}$)
σ_{HP}	allowable contact stress ($\mathrm{N/mm^2}$)
Z_{NT}	life factor
Z_L	lubricant factor
Z_V	velocity factor
Z_R	surface factor
Z_W	work harden factor
Z_X	size factor
m_{nm}	normal module (mm)
$K_{F\beta}$	face load factor
$K_{F\alpha}$	transverse load factor
Y_{FS}	form factor
Y_ε	rim thickness factor
Y_K	bevel gear factor
Y_{LS}	stress correction factor
b	face width (mm)

With the rapid development of science and technology, people have higher requirements for various machines, products and parts. For gear transmission, it is necessary not only to have enough precision and strength, but also to ensure that it can complete its specified function, namely reliability design under the specified working conditions and time [18, 45, 46]. In recent years, scholars have conducted in-depth research on reliability design, and achieved considerable results. Savage [35] established the reliability model of planetary gear system based on Weibull distribution of each unit. Thompson [41] presented a generalized optimal design formula with multi-objective, which can calculate the fatigue strength of multi-stage gear. Li. [21] established the relationship between the wind turbine system and the factors and mechanisms affecting the failure based on the global tree, the success tree and the main logic diagram. Zhang [47] established the optimization design of reliability of large ball mill gear transmission based on the Bayesian analysis algorithm of Kriging model and verified the reliability calculation results by Monte Carlo method. Huang [15] proposed a method to determine the accuracy and reliability of gear motion based on truncated random variables, and discussed the practical application value of the model in detail. Gallego-Calderon and Nejad [11, 30] studied the reliability of gear transmission system in wind turbine. Zhou [48] analysed the dynamic reliability of planetary gear drive system of shearer and verified the results by Monte Carlo simulation method.

In recent decades, a large number of studies have been carried out on the failure probability of structural systems [8, 10, 23, 31, 36, 39]. Many methods have been developed to identify major failure modes, such as (A) "probabilistic" methods, including branch and bound methods [16, 19, 29, 40] and simulation-based techniques [9, 12, 24, 27, 32]; (B) "deterministic" methods, such as incremental loading method [20, 26, 28] b-unzipper method [31], mathematical programming based method [5] or heuristic technique [37, 44]. Kim [17] proposed an effective method to identify dominant failure modes in random variable space, and then analysed the system reliability and calculated the failure probability of the system, and identified the dominant failure mode in the decreasing order of its contribution to the system failure probability. Savage [33] presented a reliability model for single input pinion or equal size double input pinion reducer. Hao [14] mainly studied the influence of different failure modes of bevel gear transmission under incomplete probabilistic information on the modeling of interdependent structures.

In this paper, the probability method is used to calculate the gear reliability model based on the COV, and a new method is proposed to optimize the bevel gear parameters in combination with DEM and GA to improve the gear reliability effectively. This paper is mainly divided into three parts. In the Section 1, the bevel gear working condition is introduced, and the reliability model of single-stage transmission bevel gear is built based on reliability theory. In the Section 2, the model of bevel gear reliability is optimized by GA, and the optimization results is verified by the Monte Carlo method. The Section 3 is a summary of the full text.

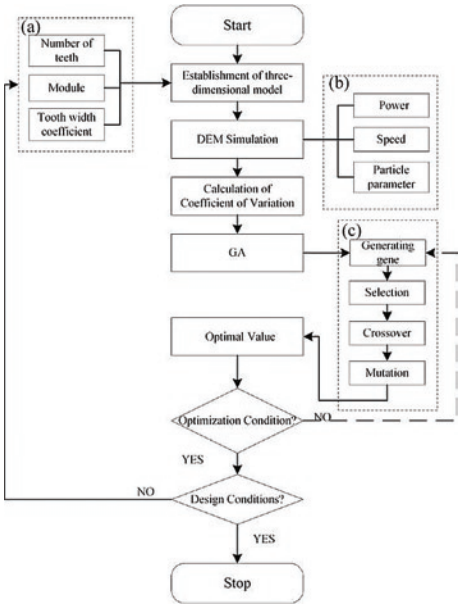


Fig. 1. Analysis flow chart

The optimal design of the bevel gear transmission system could be carried out according to the established optimization model based on DEM. The whole process of the method is shown in Figure 1. Firstly the 3D model of the DTRC is established according to the actual size, as shown in Figure 1 (a); Then the whole model is imported into EDM simulation in Figure 1 (b) according to the actual conditions of crusher (Table 2), and the torque COV of the crusher shaft torque is calculated in order to deduce the torque COV of bevel gear; Finally, based on GA, the optimization design of bevel gear is carried out in Figure 1 (c).

2. Modeling of bevel gear reliability system

2.1. Gear working condition analysis

In this paper, the double tooth roll crusher (2PGC1040×3610) is taken as an example to simulate and calculate the torque change of the central shaft of the DTRC under a certain working condition with the discrete element method (DEM). Through the mean and variance of the central shaft torque, the torque COV of the bevel gear in the transmission process is deduced.

In the reducer of the DTRC, torque change of the DTRC central shaft under a certain working condition directly affect the force situation of the bevel gear in the transmission system. According to the load of transmission system, as shown in Figure 2, a series transmission in this paper is adopted, so the load of the crusher central shaft is transferred to the bevel gear in a certain proportion. Through calculation of COV, the torque COV of the bevel gear is equal to the torque COV of the crusher central shaft and meanwhile is equal to the nominal tangential load (F_t) COV of the bevel gear.

The basic theory of DEM is that the discontinuities of the system are discretized into elements, and each element can interact with each other and satisfy Newton's law of motion when the equations of motion of the system are satisfied. The static or dynamic relaxation iteration method is used to solve the force and motion of each unit in each time step, and then the macroscopic motion law of the whole system is obtained. Based on discrete element, the flow model of material particles in machine can be established, and the information of particle velocity, flow rate, force and wear with machine can be obtained [1,6].

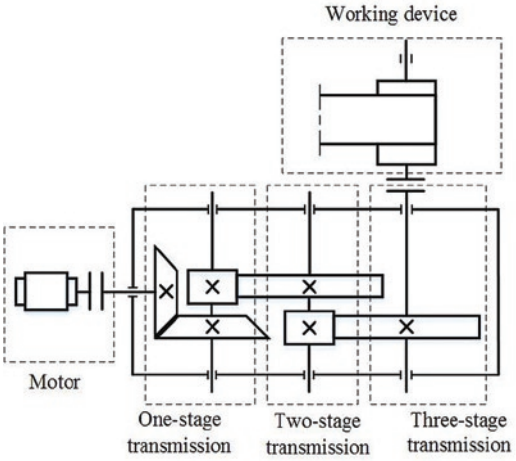


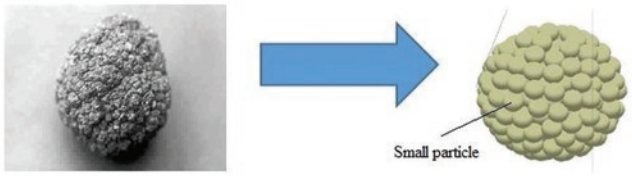
Fig. 2. Reducer of DTRC structure

Table 1. Material parameters in DEM

Material	Steel	Grit stone
Density (kg / m^3)	7850	2500
Poisson ratio	0.3	0.28
Modulus of shearing (Pa)	7.9×10^{10}	2.7×10^{8}
Recovery coefficient	0.5	0.2
Static friction coefficient	0.3	0.4
Kinetic friction coefficient	0.01	0.01

The discrete element model of DTRC mainly includes: material condensation model, prototype operation model and particle factory, material parameter setting Table 1.

The equivalent form of simulated particles in DEM is usually irregular particles, as shown in Figure 3 (a). The theoretical model of particles is assumed to be spherical in order to facilitate the establishment of equivalent model. The radius of equivalent model is 300mm, which is formed by the condensation of small particles (as shown in the Figure 3 (b)) with radius 40mm. The condensation radius is set to 40.5 mm.



a) Real particle *b) Equivalent model in DEM*

Fig. 3. Particle model transformation

Table 2. Working parameters of bevel gears

Name	Value	Name	Value
Power of motor(P)	710kW	Life expectancy(hours)	8.76×10^{4}
Motor speed(n)	175rad/s	Lubricant	Mine Gear Oil 100
Crossed axis angle	90°	Mode of production	Medium-duty
Accuracy of manufacture	6	Bevel gear ratio	3
Main gear	42CrMo(41~45HRC) Tempered	Pinion	20CrMnMo(56~62HRC) Tempered

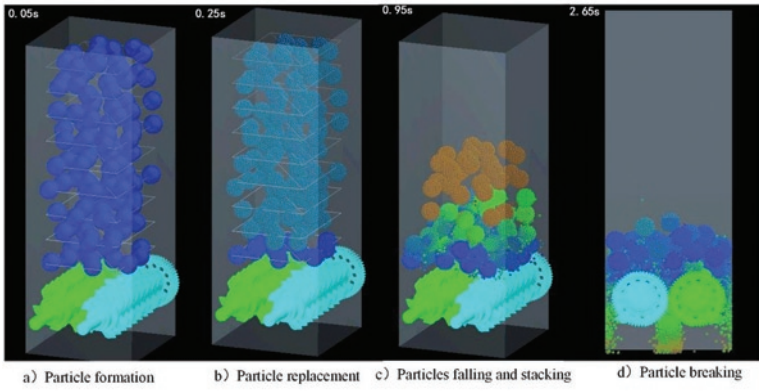


Fig. 4. Simulation process in DEM

Table 2 shows the working parameters of the bevel gear deceleration part of the DTRC. The rotational speed of the motor is 175 rad/s, after deceleration of the reducer (Figure 2), the rotational speed is

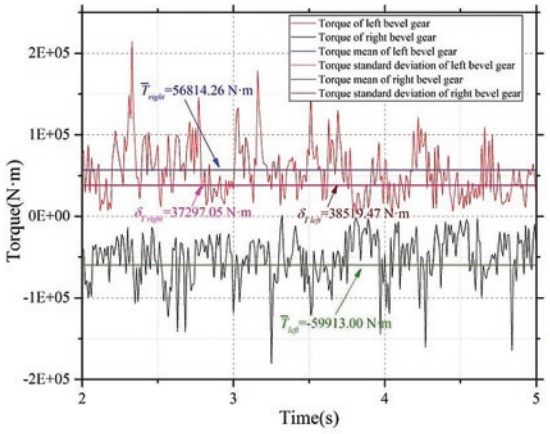


Fig. 5. Torque output curve of DTRC in DEM

Table 3. Coefficient of variation of parameters in bevel gear transmission system [14]

Variables	Mean	Std	COV	Unit	Variables	Mean	Std	COV	Unit
F_t	5.85×10^4	3.8×10^4	**0.6495726**	N·m	Z_X	1	0.033	0.033	-
K_A	1	0.033	0.033	-	Z_L	0.9658	0.0318714	0.033	-
K_V	1.034147	0.034127	0.033	-	Z_V	0.968213	0.031951	0.033	-
$K_{H\beta}$	1.65	0.05445	0.033	-	Z_R	0.936408	0.030901	0.033	-
$K_{F\beta}$	1.65	0.05445	0.033	-	Z_W	1	0.033	0.033	-
$K_{H\alpha}$	1	0.033	0.033	-	Y_{Fa1}	2.243285	0.074028	0.033	-
$K_{F\alpha}$	1	0.033	0.033	-	Y_{Fa2}	2.249664	0.074238	0.033	-
l_{bm}	27.138	0.13569	0.005	mm	Y_{sa1}	1.880157	0.06204	0.033	-
b	28	0.14	0.005	mm	Y_{sa2}	1.889086	0.06234	0.033	-
d_{V1}	50.886	0.25433	0.005	mm	Y_e	0.717334	0.0035867	0.005	-
m_{mn}	2.54827	0.01274	0.005	mm	Y_K	1.000244	0.005	0.005	-
Z_{M-B}	0.9874	0.00494	0.005	-	σ_{Flim}	380	76	0.2	N/mm^2
Z_H	2.49457	0.01247	0.005	-	Y_{ST}	2.0	0.066	0.033	-
Z_E	189.8117	9.4905	0.05	$\left(N/mm^2\right)^{1/2}$	Y_{NT1}	0.912	0.030096	0.033	-
F_t	5.85×10^4	3.8×10^4	**0.6495726**	N·m	Y_{NT2}	0.933	0.0300789	0.033	-
Z_{LS}	1	0.033	0.033	-	Y_{relT1}	1.004782	0.033158	0.033	-
Z_β	1	0.033	0.033	-	Y_{relT2}	1.003269	0.033108	0.033	-
Z_K	0.8	0.004	0.005	-	Y_{RelT1}	1.024202	0.033809	0.033	-
σ_{Hlim}	1370	164.4	0.12	N/mm^2	Y_{RelT2}	1.024202	0.033809	0.033	-
Z_{NT}	1.07	0.03531	0.033	-	Y_{X1}	1	0.033	0.033	-

reduced to 2.77 rad/s, the total gear ratio is 63.18, and the input power of motor is 710 kW. The bevel gear works at the one-stage transmission with a gear ratio of 3, a rotational speed of 175 rad/s for the input shaft, and a rotational speed of 58.3 rad/s for the output shaft. The material and processing parameters are shown in the table.

Figure 4 shows the change of particle position at different time in DEM simulation. The working speed of DTRC shaft is 2.77 rad/s. It can be seen that the particle does not contact the crusher when the simulation time reaches 1 s, so the torque of the crusher central shaft is zero. After 1 s, the crusher starts to work. The torque change curve of the central shaft under the steady working condition of 2-5s crusher is exported, as shown in Figure 5.

Figure 5 is torque output curve of DTRC in DEM, the output torque of the left and right tooth rollers can be seen to be basically equal. The load distribution information is extracted and the mean value and standard deviation are calculated as follows:

$$\begin{cases} \overline{T}_e = \dfrac{\overline{T}_{right} + \overline{T}_{left}}{2} = 58363.63N \cdot m \\ \delta_{T_e} = \dfrac{\delta_{T_{right}} + \delta_{T_{left}}}{2} = 37908.26N \cdot \boldsymbol{m} \end{cases} \tag{1}$$

For accuracy of calculation, the mean and standard deviation (Std) are taken as the mean values of the left and right axes.

$$COV(F_t) = COV(\overline{T}) = \frac{\delta_{T_e}}{\overline{T}_e} \tag{2}$$

In Table 3, except that the force F_t is simulated by DEM, the other gear parameters are assumed to follow the normal distribution [14], so the COV of the parameters is calculated, and the reliability calculation process is shown in the appendix of this paper.

2.2. Gear modeling analysis

The theory of system reliability bound, such as Cornell bounds [4] and Ditlevsen bounds [7], is often used to estimate the reliability of the system. It is also called the narrow reliability bound. By evaluating the joint failure probability of each pair of failure modes, a narrower estimate of the system failure probability is obtained. For a series system with m failure modes, a narrow bound of system reliability estimation is given:

$$P_{f_1} + \sum_{i=2}^{m} \max(P_{f_i} - \sum_{j=1}^{i-1} P_{f_{ij}}, 0) \le P_{f_1} \le \sum_{i=1}^{m} P_{f_1} - \sum_{i=2}^{m} \max(P_{f_{ij}}) \tag{3}$$

Where, P_{fi} is the failure probability of the first failure mode and P_{fij} is the joint failure probability of each pair of failure modes. Considering the convenience of the analysis, we only consider the primary bevel gear drive, which is a series system.

2.3. Determination of Objective Functions

According to the definition of robust optimal design involving reliability sensitivity, the robust design optimization model of bevel gear transmission system based on reliability is established:

$$\min f(\boldsymbol{X}) = \sum_{k=1}^{M} w_k f_k(\boldsymbol{X}) \tag{4}$$

$$s.t.\ R[g_i(X) > 0] \le R_i \quad ,i = 1,...,N \tag{5}$$

$$\boldsymbol{X}^L \le \boldsymbol{X} \le \boldsymbol{X}^U \ , \ \boldsymbol{X} \in \boldsymbol{X}^{ndv} \tag{6}$$

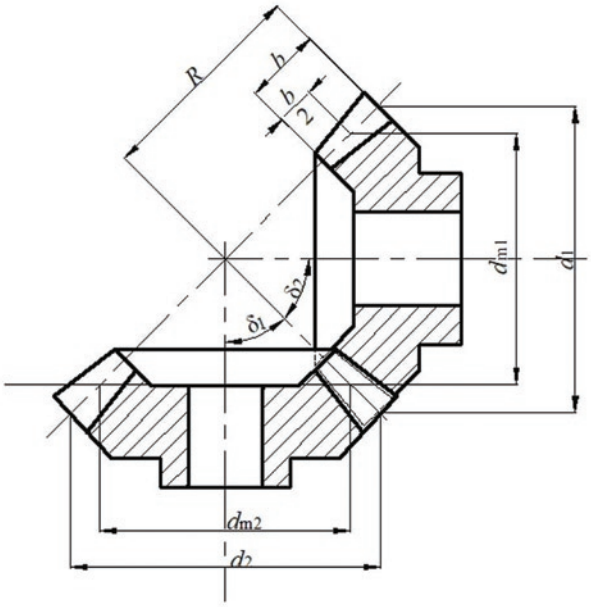


Fig. 6. Bevel gear transmission geometric parameters

Where, $\boldsymbol{X}$ is the design variables, $\boldsymbol{X} = [m_{nm}, z_1, \varphi_R]^T$, $\boldsymbol{X}^L$ and $\boldsymbol{X}^U$ is the lower and upper bounds of the design variables, respectively. *M*, *N* and *ndv* represent the number of sub-objective functions, reliability functions and design variables, respectively. ω_k is the weight of each sub-function which shows the importance of each sub-function in calculating reliability. The formulas are as follows:

$$\omega_k = [f_1(\boldsymbol{X}^{*k} - f_1(\boldsymbol{X}^{*1}))] / \{[f_1(\boldsymbol{X}^{*k}) - f_1(\boldsymbol{X}^{*1})] + [f_2(\boldsymbol{X}^{*(k-1)}) - f_2(\boldsymbol{X}^{*2})] + ... + [f_k(\boldsymbol{X}^{*1}) - f_k(\boldsymbol{X}^{*k})]\} \tag{7}$$

where, the weighting factor should satisfy the following conditions:

$$\sum_{k=1}^{n} \omega_k = 1 \quad (\omega_k \ge 1) \tag{8}$$

The reliability function is the objective function f_1, and the bevel gear transmission is series system. $R_1(\boldsymbol{X})$, $R_2(\boldsymbol{X})$, $R_3(\boldsymbol{X})$, $R_4(\boldsymbol{X})$ represent the reliability of contact strength of the gear tooth surface, bending strength of the gear tooth root, contact strength of pinion tooth surface and bending strength of pinion tooth root, respectively. In order to ensure the reliability of gear transmission, the structure of transmission will be destroyed if any gear is damaged. Therefore, the reliability objective function of bevel gear transmission can be designed as follows:

$$f_1(\boldsymbol{X}) = f(\boldsymbol{X}) = 1 - R_1(\boldsymbol{X})R_2(\boldsymbol{X})R_3(\boldsymbol{X})R_4(\boldsymbol{X}) \tag{9}$$

where, calculation formula of $R_1(\boldsymbol{X})$, $R_2(\boldsymbol{X})$, $R_4(\boldsymbol{X})$ is in the appendix.

Meanwhile, the volume of bevel gear group is also the objective function f_2, the formula is:

$$f_2(\boldsymbol{X}) = \left(f_p(\boldsymbol{X}) + f_g(\boldsymbol{X})\right) / V_{\max} \tag{10}$$

where, $f_p(\boldsymbol{X})$, $f_g(\boldsymbol{X})$ represent the volume of the pinion and the main gear, respectively.

$$f_p(\boldsymbol{X}) = \frac{\pi u d_1^2}{12\times 10^9}\left(1-\left(\frac{d_1 - b/u}{d_1}\right)\right) \tag{11}$$

$$f_g(\boldsymbol{X}) = \frac{\pi d_2^3}{12\times u\times 10^9}\left(1-\left(\frac{d_2 - bu}{d_2}\right)\right) \tag{12}$$

$$b = \varphi_R m_n \tag{13}$$

$$d_1 = z_1 m_n \tag{14}$$

$$d_2 = d_1 u \tag{15}$$

where, u is bevel gear ratio.

Therefore, the formula (1)-(13) is simultaneous, then the objective function is as follows:

$$\min f(\boldsymbol{X}) = \omega_1 f_1(\boldsymbol{X}) + \omega_2 f_2(\boldsymbol{X}) \tag{16}$$

$$f_1(\boldsymbol{X}) = 1 - R_1(\boldsymbol{X})R_2(\boldsymbol{X})R_3(\boldsymbol{X})R_4(\boldsymbol{X}) \tag{17}$$

$$f_2(\boldsymbol{X}) = \frac{\pi u(z_1 m_n)^2}{12\times 10^9}\left(1-\left(\frac{z_1 - \varphi_R/u}{z_1}\right) + u - u\left(\frac{z_1 - \varphi_R}{z_1}\right)^2\right) \Big/ V_{\max} \tag{18}$$

$$\textbf{\textit{s.t.}}\ R[g_i(\boldsymbol{X}) > 0] \le R_i \tag{19}$$

$$\left(R_1 \cap R_2 \cap R_3 \cap R_4\right) \ge 0.9 \tag{20}$$

$$g_i(\boldsymbol{X}) \ge 0 \tag{21}$$

$$\begin{cases} 1 \le m_e \le 50 \\ z_{\min} \le z_1 \le 44 \\ 1/4 \le \varphi_R \le 1/3 \end{cases} \tag{22}$$

where, $i = 1,2$, z_{min} is the minimum number of pinion teeth, determined by the type of bevel gear in Table 4. $z_{min} = 13 \sim 15$.

Table 4. Selection of gear types and minimum module

Type of Gear	Coniflex gear	Skew bevel gear	Zerol bevel gear	Gleason spiral bevel gear
Power range (kW)	<50	<500	<3700	>3700
Minimum teeth number	13	13	15	12

3. Case Analysis

3.1. Analysis of results

Figure 7-11 shows the optimization results of DEM and GA. From those figures, it can be seen that the optimization result is good. As shown in Figure 7, the optimization result of GA tends to be stable after 70 iterations by using the objective function of weight $\omega_1 = 0.3$ and $\omega_2 = 0.7$, that is, the optimal value may be reached. It can be seen that the total volume and failure rate of gears will affect the objective function, and the objective function decreases first and then rise. That indicates that the reliability of gear system has a great influence on the objective function before 25 generations, while after 25 generations, the total volume of gear system has great influence on the objective function.

Figure 8 shows number of teeth and module curve (DEM and GA) and Figure 9 shows face width factor curve (DEM and GA). Combined with Figure 7-9, GA is used to optimize bevel gear parameters (teeth number, module and face width factor), which change trend is approximate, finally after 80 iterations tend to balance. Variable approximation is $\mathbf{X}_{op} = \left(m_n, z_1, \varphi_R\right)^T \approx \left(11, 32.23, 0.264\right)^T$, optimal value $f\left(\boldsymbol{X}\right)_{min} = 0.5661$.

Figure 10 shows the failure rate and reliability curve (DEM and GA), after 30 iterations, the reliability gradually tends to 1, while Gear failure rate gradually approach 0. Figure 11 shows the change of contact stress on tooth surface and bending stress on tooth root during iteration. According to experience, the pinion is most easily worn out during transmission. The allowable contact stress of pinion is $783.57\,\text{MPa}$ and the bending stress of pinion tooth root is $466.95\,\text{MPa}$. As the Figure 11 shows, after 30 generations, the contact strength and bending strength of gear are lower than the allowable value, and the curve tends to be stable.

Table 5 shows the differences of optimization results using DEM and without DEM. In mode 1, the result is not using DEM and the COV of nominal tangential load (COV_{Ft}) is assumed to be 0; in mode 2, the COV of nominal tangential load (COV_{Ft}) is obtained from DEM. The results of mode 1 and mode 2 is all optimized by GA.

Comparing mode 1 with mode 2, it can be seen that when mode 1 reaches the best point $\left(m_n, z_1, \varphi_R\right)^T \approx \left(11, 28.57, 0.298\right)^T$ and mode 2 reaches the best point $\left(m_n, z_1, \varphi_R\right)^T \approx \left(11, 32.23, 0.264\right)^T$, the interesting phenomenon is occur that there is a small change in module relative to mode 1, which is always 11mm. The teeth number increased from 28.57 to 32.23, while the face width factor decreased from 0.298 to 0.264. The reliability changes little and the gear volume increases by 34.4%, which shows that the mode 2 design increases the cost, but the mode 2 design is more suitable for the actual situation. The method is same as the safety factor in gear design. Because of the different working conditions of gear, the selection of safety factor is often based on experience. This method provides theoretical guidance for the selection of gear safety factor.

In the mode 2, the optimization result based on the DEM increase the number of gear tooth than the mode 1, resulting in the increase of the volume, but the reliability change is relatively small. The result shows that when the variation coefficient method is used to optimize the gear reliability, the gear working condition also has a great influence on the reliability optimization.

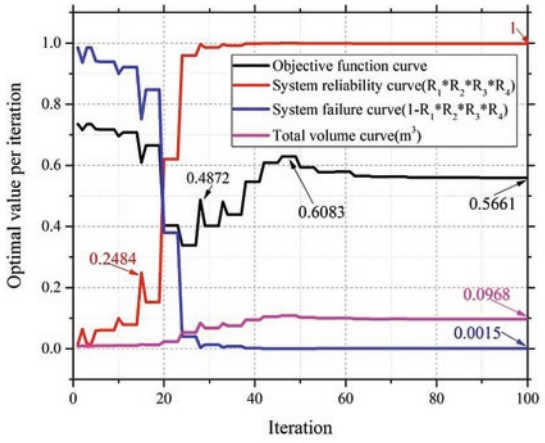


Fig. 7. Gear parameters curve

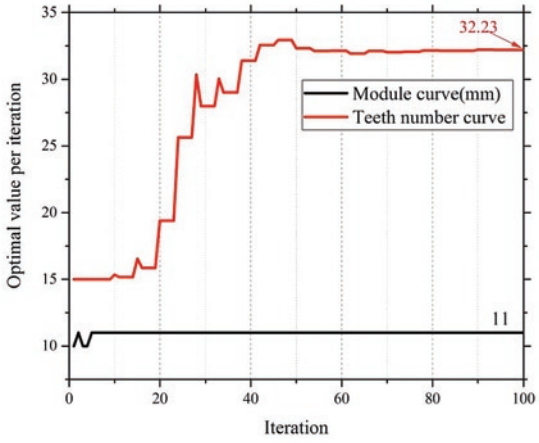


Fig. 8. Number of teeth and module curve

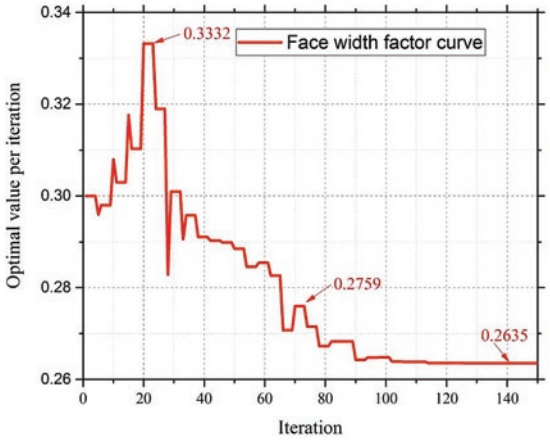


Fig.9. Face width factor curve

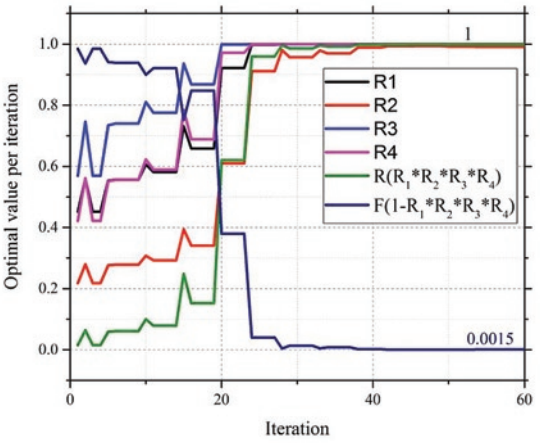


Fig. 10. Reliability and failure rate curve

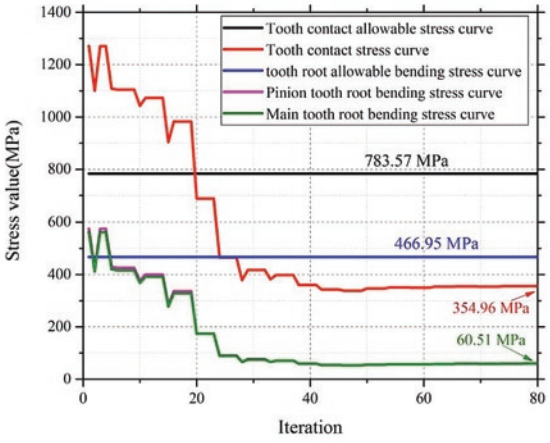


Fig. 11. Stress iterative curve

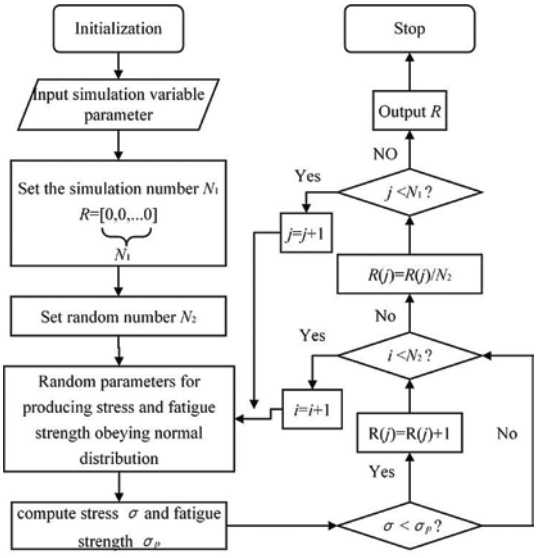


Fig. 12. Reliability calculation using Monte Carlo method

Table 5. Differences of optimization results using DEM and without DEM

	Design mode	m_e	z_1	φ	R_1	R_2	R_3	R_4	System Reliability	Volume (m^3)	F_{ob}
1	Reliability optimization design ($COV_{Ft}=0$)	11	28.57	0.298	0.9999	0.9903	1	1	**0.9903**	**0.0721**	0.4227
	Teeth number upward rounding ($COV_{Ft}=0$)	11	29	0.298	0.9999	0.9925	1	1	0.9925	0.0754	0.4403
2	Reliability optimization design (DEM) ($COV_{Ft}=0.6495726$)	11	**32.23**	0.264	0.9998	0.9900	1	1	**0.9900**	0.0969	**0.5661**
	Teeth number upward rounding (DEM) ($COV_{Ft}=0.6495726$)	11	33	0.264	0.9999	0.9931	1	1	0.9931	0.104	0.6050

Note: $F_{ob}=\omega_1\times Reliability+\omega_2\times\frac{Volume}{Vmax}, V_{max}=0.052\mathrm{m}^3, \omega_1=0.3, \omega_2=0.7$

Table 6. Results Comparison of Monte Carlo (MC) and theoretical value (TV)

Reliability	Teeth number upward rounding ($COV_T=0$)	Teeth number upward rounding (DEM) ($COV_T=0.6495726$)	
		MC	TV	MC
R_1	0.9999	0.998201	0.9999	0.999896
R_2	0.9925	0.992051	0.9931	0.994681
R_3	1	1	1	1
R_4	1	0.992494	1	0.999439

3.2. Verification of reliability with Monte Carlo method

Monte Carlo method is also called statistical test method, stochastic simulation method, or random sampling technique or statistical test method.

When the Monte Carlo method is used to calculate the reliability of gear transmission, the parameters related to the geometric dimensions of the gears are regarded as constants, while the other parameters are considered to obey normal distribution.

As shown in Figure 12, this paper uses the Monte Carlo method to calculate the gear transmission reliability flow chart, the basic process is:

Step 1. Input the parameters of simulation variables that need to verify the results;

Step 2. Set the number of simulations N_1 and make the N_1 dimension line vector $R=[0,0,\ldots 0]$;

Step 3. Generating N_2 random variables of normal distribution;

Step 4. Calculate the stress σ and fatigue strength σ_p;

Step 5. Comparing the magnitude of stress and fatigue strength σ_p, if $\sigma<\sigma_p$, it accumulates to the reliable database, otherwise it goes directly to the next step;

Step 6. If the total number of random numbers reaches N_2, go to the next step, otherwise, return to Step4;

Step 7. Calculate the reliability of jth order $R(j)=R(j)/N_2$;

Step 8. If the number of simulations reaches N_1. , enter the reliability R of all the simulations and end the program. Otherwise, $j=j+1$, return to Step3.

Suppose the simulation number N_1 is 1 and the number of random numbers N_2 is one million. In Table 6, comparison of reliability between Monte Carlo and theoretical calculations (Section 3.1) can be obtained.

In Table 6, the simulation results of reliability obtained by Monte Carlo method and theoretical value are very similar, so it can be considered that the calculation of results by normal distribution is correct.

4. Summary

In this paper, the multi-objective function model of bevel gear is established, and the parameters of bevel gear are optimized by GA and DEM. The feasibility of the method is proved theoretically. In the design of gear reliability, the calculation of reliability by COV often assumes that the moment of transmission is subjected to constant load, but the actual situation is not always so. The innovation of this paper is to use DEM. Combined with DEM of real working conditions, the torque COV of gear transmission is calculated, which makes the designed gear more in line with the actual situation. The results of this paper verify the feasibility of the method by Monte Carlo method and provide reference for the design and optimization of gears.

The results show that it is closer to gear design when the optimum design of gear under actual working conditions are considered, and this method is similar to the selection of gear safety factor. In practical production, the selection of safety factor is often based on experience.

This paper provides a new method to optimize the reliability of bevel gear, combining with DEM simulation, which provides theoretical guidance for optimal design of bevel gear.

There are also many defects in this paper. After all, the DEM is only theoretical simulation, and there is still a lot of gap with the actual situation. At the same time, the GA is used in the process of global optimization, because of the defects of the algorithm. It is possible to fall into a local optimal situation in the process of optimization. Of course, with the continuous improvement of computer performance and research of algorithm, this method will have a lot of room for improvement.

Acknowledgement

The research work financed with the National Natural Science Foundation of China (Grant No. 51775225) and key scientific and technological coal based research projects in Shanxi (Grant No. MJ2014-02).

References

1. Boemer D, Ponthot J. A generic wear prediction procedure based on the discrete element method for ball mill liners in the cement industry. Minerals Engineering 2017; 109: 55-79, https://doi.org/10.1016/j.mineng.2017.02.014.
2. Chong T H, Lee J S. A Design Method of Gear Trains Using a Genetic Algorithm. International Journal of Precision Engineering & Manufacturing 2000; 1(1): 62-70, http://pdfs.semanticscholar.org/3460/f3437791bde00fa9f5e21c98e5836980a17a.pdf.
3. Chong T H, Bae I, Park G J. A new and generalized methodology to design multi-stage gear drives by integrating the dimensional and the configuration design process. Mechanism & Machine Theory 2002; 37(3): 295-310. https://doi.org/10.1016/S0094-114X(01)00078-7.
4. Chong T H, Bae I, Kubo A. Multiobjective Optimal Design of Cylindrical Gear Pairs for the Reduction of Gear Size and Meshing Vibration 2002; 44(1): 291-298, https://doi.org/10.1299/jsmec.44.291
5. Corotis RB, Nafday A M. Structural system reliability using linear programming and simulation. Journal of Structural Engineering ASCE 1989; 115(10): 2435–2447, https://doi.org/10.1061/(ASCE)0733-9445(1989)115:10(2435)
6. Delaney G W, Morrison R D, Sinnott M D, et al. DEM modelling of non-spherical particle breakage and flow in an industrial scale cone crusher. Minerals Engineering 2015; 74: 112-122, https://doi.org/10.1016/j.mineng.2015.01.013.
7. Deb K, Jain S. Multi-Speed Gearbox Design Using Multi-Objective Evolutionary Algorithms. Journal of Mechanical Design 2003; 125(3): 609-619, http://link.aip.org/link/?JMDEDB/125/609/1.
8. Ditlevsen O, Madsen HO. Structural reliability methods. Chichester, UK: JohnWiley & Sons; 1996.
9. Ditlevsen O, Bjerager P. Plastic reliability analysis by directional simulation. Journal of Engineering Mechanics ASCE 1989; 115(6): 1347–1362, https://doi.org/10.1061/(ASCE)0733-9399(1989)115:6(1347).
10. Freudenthal AM, Garrelts JM, Shinozuka M. The analysis of structural safety. Journal of Structures Division, ASCE 1966; 92: 267–325.
11. Gallego-Calderon J, Natarajan A, Dimitrov N K. Effects of Bearing Configuration in wind Turbine Gearbox Reliability. Energy Procedia 2015; 80 (Pt2): 392-400, https://doi.org/10.1016/j.egypro.2015.11.443.
12. Grimmelt M, Schueller GI. Benchmark study on methods to determine collapse failure probabilities of redundant structures. Structural Safety 1982;1:93–106.
13. Golabi S, Fesharaki J J, Yazdipoor M. Gear train optimization based on minimum volume/weight design. Mechanism & Machine Theory 2014; 73(2): 197-217, https://doi.org/10.1016/j.mechmachtheory.2013.11.002.
14. Hao L. The probabilistic analysis and optimal design of a bevel gear transmission system with failure interaction. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19(2): 220-228, http://dx.doi.org/10.17531/ein.2017.2.9.
15. Huang X, Hu S, Zhang Y, et al. A method to determine kinematic accuracy reliability of gear mechanisms with truncated random variables. Automotive Engineering 2015; 92: 200-212, https://doi.org/10.1016/j.mechmachtheory.2015.04.017.
16. Karamchandani A. Structural system reliability analysis methods. Report no. 83. Department of Civil Engineering, Stanford University; 1987.
17. Kim, D., et al., System reliability analysis using dominant failure modes identified by selective searching technique. Reliability Engineering & System Safety 2013; 119: 316-331, https://doi.org/10.1016/j.ress.2013.02.007
18. Li M, Xie L, Ding L. Load sharing analysis and reliability prediction for planetary gear train of helicopter[J]. Mechanism & Machine Theory, 2017,115:97-113.https://doi.org/10.1016/j.mechmachtheory.2017.05.001
19. Lee YJ, Song J. Finite-element-based system reliability analysis of fatigueinduced sequential failures. Reliability Engineering & System Safety 2012; 108: 131–41, https://doi.org/10.1016/0167-4730(82)90018-2.
20. Lee JS. Basic study on the reliability analysis of structural systems. Journal of Ocean Engineering and Technology 1989; 12: 145–57.
21. Li Y F, Valla S, Zio E. Reliability assessment of generic geared wind turbines by GTST-MLD model and Monte Carlo simulation. Renewable Energy 2015; 83: 222-233, https://doi.org/10.1016/j.renene.2015.04.035.
22. Marjanovic N, Isailovic B, Marjanovic V, et al. A practical approach to the optimization of gear trains with spur gears. Mechanism & Machine Theory 2012; 53(7): 1-16, https://doi.org/10.1016/j.mechmachtheory.2012.02.004.
23. Melchers RE. Structural reliability: analysis and prediction. 2nd ed New York,NY: John Wiley; 1999. ISBN:0471987719
24. Melchers RE. Structural system reliability assessment using directional simulation. Structural Safety 1994; 16: 23–37, https://doi.org/10.1016/0167-4730(94)00026-M.
25. Mendi F, Boran F E. Optimization of module, shaft diameter and rolling bearing for spur gear through genetic algorithm. Expert Systems with Applications An International Journal 2010; 37(12): 8058-8064, https://doi.org/10.1016/j.eswa.2010.05.082.
26. Moses F, Stahl B. Reliability analysis format for offshore structures. In: Proceedings of the 10th annual offshore technology conference, 1978; Paper 3046. DOI: 10.4043/3046-MS
27. Moses F, Fu G. Important sampling in structural system reliability. Fifth ASCE EMD/GTD/STD specialty conference on probabilistic mchanics; 1988, http://cedb.asce.org/CEDBsearch/record.jsp?dockey=0056556.
28. Moses F. System reliability developments in structural engineering. Structural Safety 1982; 1: 3–13, https://doi.org/10.1016/0167-4730(82)90011-X.
29. Murotsu Y, Okada H, Taguchi K, Grimmelt M, Yonezawa M. Automatic generation of stochastically dominant failure modes of frame structures.Structural Safety 1984; 2: 17–25, https://doi.org/10.1016/0167-4730(84)90004-3.

30. Nejad A R, Gao Z, Moan T. On long-term fatigue damage and reliability analysis of gears under wind loads in offshore wind turbine drivetrains. International Journal of Fatigue 2014, 61(2): 116-128, https://doi.org/10.1016/j.ijfatigue.2013.11.023.
31. Nikolaidis E, Ghiocel DM, Singhal S. Engineering design reliability handbook.Boca Raton, FL: CRC Press; 2005, https://doi.org/10.1201/9780203483930.
32. Rashedi MR. Studies on reliability of structural systems. Department of Civil Engineering, Case Western Reserve University; 1983
33. Savage M, Brikmanis C, Lewicki D G, Coy J J. Life and reliability modeling of bevel gear reductions. Journal of Mechanisms, Transmissions, and Automation in Design 1988; 110(2): 189-196, https://doi.org/10.1115/1.3258925.
34. Savsani V, Rao R V, Vakharia D P. Optimal weight design of a gear train using particle swarm optimization and simulated annealing algorithms. Mechanism & Machine Theory 2010; 45(3): 531-541, https://doi.org/10.1016/j.mechmachtheory.2009.10.010.
35. Savage M, Paridon C A, Coy J J. Reliability Model for Planetary Gear Trains. Journal of Mechanical Design 1983; 105(3): 291-297, https://doi.org/10.1115/1.3267359
36. Shao S, Murotsu Y. Approach to failure mode analysis of large structures.Probabilistic Engineering Mechanics 1999; 14: 169–177, https://doi.org/10.1016/S0266-8920(98)00028-9.
37. Shetty NK. Selective enumeration method for identification of dominant failure paths of large structures. In: Proceedings of OMAE conference, vol. II. ASME. Safety and Reliability 1994: 381–391, https://www.osti.gov/biblio/55808.
38. Swantner A, Campbell M I. Topological and parametric optimization of gear trains. Engineering Optimization 2012, 44(11): 1351-1368, https://doi.org/10.1080/0305215X.2011.646264.
39. Thoft-Christensen P, Baker MJ. Structural reliability theory and its applications. Springer-Verlag; 1982, https://doi.org/10.1007/978-3-642-68697-9_11.
40. Thoft-Christensen P, Murotsu Y. Application of structural systems reliability theory. Berlin: Springer-Verlag; 1986, https://link.springer.com/book/10.1007%2F978-3-642-82764-8.
41. Thompson D F, Gupta S, Shukla A. Tradeoff analysis in minimum volume design of multi-stage spur gear reduction units. Mechanism & Machine Theory 2000; 35(5): 609-627, https://doi.org/10.1016/S0094-114X(99)00036-1.
42. Wang H, Wang H P. Optimal engineering design of spur gear sets. Mechanism & Machine Theory 1994; 29(7): 1071-1080, https://doi.org/10.1016/0094114X(94)90074-4.
43. Wang Y. Optimized tooth profile based on identified gear dynamic model. Mechanism & Machine Theory 2007; 42(8): 1058-1068, https://doi.org/10.1016/j.mechmachtheory.2006.02.011
44. Xiao Q, Mahadevan S. Fast failure mode identification for ductile structural system reliability. Structural Safety 1994; 13(4): 207–26, https://doi.org/10.1016/0167-4730(94)90029-9.
45. Xie L, Wu N, Qian W. Time domain series system definition and gear set reliability modeling. Reliability Engineering & System Safety 2016; 155: 97-104, https://doi.org/10.1016/j.ress.2016.06.009.
46. Yang Q J. Fatigue test and reliability design of gears. International Journal of Fatigue 1996; 18(3): 171-177, https://doi.org/10.1016/0142-1123(95)00096-8.
47. Zhang G, Wang G, Li X, et al. Global optimization of reliability design for large ball mill gear transmission based on the Kriging model and genetic algorithm. Mechanism & Machine Theory 2013; 69(6): 321-336, https://doi.org/10.1016/j.mechmachtheory.2013.06.003.
48. Zhou D, Zhang X, Zhang Y. Dynamic reliability analysis for planetary gear system in shearer mechanisms. Mechanism & Machine Theory 2016; 105: 244-259, https://doi.org/10.1016/j.mechmachtheory.2016.07.007.
49. Zolfaghari A , Goharimanesh M , Akbari A A . Optimum design of straight bevel gears pair using evolutionary algorithms. Journal of the Brazilian Society of Mechanical Sciences & Engineering 2017; 39(6): 1-9, https://doi.org/10.1007/s40430-017-0733-9.

Appendix:

1. Calculation of Reliability of Bevel Gear Transmission system

According to the IOS standard, reliability design is to treat the design variables such as load, strength and their influencing factors as random variables, and apply the reliability theory and method to make the designed products meet the expected reliability requirements. In this paper, the reliability design of bevel gear transmission system mainly includes: the reliability of tooth surface contact strength of the pinion and the gear, the reliability of tooth surface contact strength of the pinion and the gear are expressed as R_1, R_2, R_3, R_4 , respectively.

1.1. Reliability calculation of tooth surface contact strength (R_1, R_2)

In the calculation of tooth surface contact fatigue strength, destructive pitting is the limit state of tooth surface work. The strength condition of the tooth surface without fatigue pitting is that the contact stress of the tooth surface joint is not greater than the contact allowable stress. Its reliability calculation formula is as follows:

$$R = P\{\sigma_H < \sigma_{HP}\} \tag{23}$$

1.1.1. Bevel gear tooth surface contact stress

Computational formula:

$$\sigma_{\mathrm{H}}=Z_M Z_H Z_E Z_{LS} Z_\beta Z_K \sqrt{\frac{KK_A K_V K_{H\beta} K_{H\alpha} F_t}{d_{m1} l_{bm}} \times \frac{\sqrt{u^2+1}}{u}} \tag{24}$$

Formula of COV:

$$\sigma_{\mathrm{H}}=Z_M Z_H Z_E Z_{LS} Z_\beta Z_K \sqrt{\frac{KK_A K_V K_{H\beta} K_{H\alpha} F_t}{d_{m1} l_{bm}} \times \frac{\sqrt{u^2+1}}{u}} \tag{25}$$

where, $\overline{Z}_E, \overline{F}_t, \overline{K}, \overline{K}_{H\beta}, \overline{K}_{H\alpha}$ are the mean values of the corresponding parameters.

$$V_{\sigma H}=[V_{Z_E}^2+\frac{1}{4}(V_K^2+V_{F_t}^2+V_{K_V}^2+V_{K_{H\beta}}^2+V_{K_{H\alpha}}^2+V_{K_V}V_{K_{H\alpha}}+V_{K_V}V_{K_{H\beta}}+V_{K_{H\alpha}}V_{K_{H\beta}})]^{0.5} \tag{26}$$

$$S_{\sigma_H} = \overline{\sigma_H} V_{\sigma_H} \tag{27}$$

The average tangential force on the bevel gear face is:

$$\bar{F}_t = \frac{2000\bar{T}}{d_{m1}} \tag{28}$$

1.1.2. Contact fatigue strength of bevel gears

Formula of tooth surface contact fatigue strength:

$$\sigma_{\mathrm{HP}} = \sigma_{Hlim} Z_{\mathrm{NT}} Z_{\mathrm{L}} Z_{\mathrm{V}} Z_{\mathrm{R}} Z_{\mathrm{W}} Z_{\mathrm{X}} \tag{29}$$

Formula of COV:

$$\bar{\sigma}_{\mathrm{HP}} = \bar{\sigma}_{\mathrm{Hlim}} \bar{Z}_{\mathrm{NT}} \bar{Z}_{\mathrm{L}} \bar{Z}_{\mathrm{V}} \bar{Z}_{\mathrm{R}} \bar{Z}_{\mathrm{W}} Z_{\mathrm{X}} \tag{30}$$

$$V_{\sigma_{HP}} = \left(V_{\sigma_{H\mathrm{lim}}}^2 + V_{Z_{NT}}^2 + V_{Z_L}^2 + V_{Z_V}^2 + V_{Z_R}^2 + V_{Z_W}^2 \right)^{0.5} \tag{31}$$

$$S_{\sigma_{HP}} = \bar{\sigma}_{\mathrm{HP}} V_{\sigma_{HP}} \tag{32}$$

Assuming that the contact stress limit function is:

$$g_1(\boldsymbol{X}_1) = \sigma_H - \sigma_{HP} \tag{33}$$

The formula of the contact fatigue strength of tooth surfaces is as follows:

$$Z_R = \frac{\bar{\sigma}_{HP} - \bar{\sigma}_H}{\left(S_{\sigma_{HP}}^2 + S_{\sigma_H}^2 \right)^{0.5}} \tag{34}$$

$$R = \Phi(Z_R) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_R} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x \tag{35}$$

1.2. Reliability calculation of tooth root bending strength $(\boldsymbol{R}_3, \boldsymbol{R}_4)$

In the calculation of tooth root bending strength, the fracture of tooth root is regarded as the limit state. The condition that the tooth root does not break is that the bending stress of the tooth root joint is not greater than the bending allowable stress. The reliability calculation method is as follows:

$$R = P\{\sigma_F < \sigma_{FP}\} \tag{36}$$

1.2.1. Root bending stress

Formula of bending stress of tooth root:

$$\sigma_F = K K_{\mathrm{A}} K_{\mathrm{V}} K_{\mathrm{F}\beta} K_{\mathrm{F}\alpha} F_{\mathrm{t}} \frac{Y_{\mathrm{FS}} Y_\varepsilon Y_{\mathrm{K}} Y_{\mathrm{LS}}}{b m_{\mathrm{nm}}} \tag{37}$$

Formula of COV:

$$\bar{\sigma}_F = \bar{K} K_{\mathrm{A}} \bar{K}_{\mathrm{V}} \bar{K}_{\mathrm{F}\beta} \bar{K}_{\mathrm{F}\alpha} \bar{F}_{\mathrm{t}} \frac{Y_{\mathrm{FS}} Y_\varepsilon Y_{\mathrm{K}} Y_{\mathrm{LS}}}{b m_{\mathrm{nm}}} \tag{38}$$

$$V_{\sigma_F} = \left(V_K^2 + V_{F_t}^2 + V_{K_V}^2 + V_{K_{F\beta}}^2 + V_{K_{F\alpha}}^2 + V_{K_{F\alpha}} V_{K_{F\beta}} + V_{K_{F\alpha}} V_{K_V} + V_{K_{F\beta}} V_{K_V} \right)^{0.5} \tag{39}$$

$$S_{\sigma_F} = \bar{\sigma}_F V_{\sigma_F} \tag{40}$$

1.2.2. Tooth root bending fatigue strength

Formula of tooth root bending fatigue strength:

$$\sigma_{FP} = \sigma_{F\,\mathrm{lim}} Y_{NT} Y_{\delta\mathrm{rel}T} Y_{\mathrm{Rrel}T} Y_X \tag{41}$$

Formula of COV:

$$\bar{\sigma}_{FP} = \bar{\sigma}_{F\,\mathrm{lim}} \bar{Y}_{NT} Y_{\delta\mathrm{rel}T} Y_{\mathrm{Rrel}T} Y_X \tag{42}$$

$$V_{\sigma_{FP}} = \left(V_{\sigma_{F\,\mathrm{lim}}}^2 + V_{Y_{NT}}^2 \right)^{0.5} \tag{43}$$

Assuming that the bending stress limit function is:

$$g_2(\boldsymbol{X}_2) = \sigma_F - \sigma_{FP} \tag{44}$$

The formula for the reliability of tooth root bending fatigue strength is:

$$Z_R = \frac{\bar{\sigma}_{FP} - \bar{\sigma}_F}{\left(S_{\sigma_{HF}}^2 + S_{\sigma_F}^2 \right)^{0.5}} \tag{45}$$

$$R = \Phi(Z_R) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_R} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x \tag{46}$$

Kangkang SUN
Guoqiang WANG
Yanpeng LU
College of Mechanical & Aeronautics
& Astronautics Engineering
Jilin University, Changchun, 130025, P. R. China

E-mails: szpsunkk@163.com, sunkk17@mails.jlu.edu.cn

Article citation info:
URBAŚ A, SZCZOTKA M. The influence of the friction phenomenon on a forest crane operator's level of discomfort. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 197–210, http://dx.doi.org/10.17531/ein.2019.2.3.

Andrzej Urbaś
Marek Szczotka

THE INFLUENCE OF THE FRICTION PHENOMENON ON A FOREST CRANE OPERATOR'S LEVEL OF DISCOMFORT

WPŁYW ZJAWISKA TARCIA NA POZIOM DYSKOMFORTU PRACY OPERATORA ŻURAWIA LEŚNEGO*

A mathematical model of a forest crane that is suitable for dynamics analysis of its operation cycle is presented in this paper. The flexibility of the operator's seat, drives and supports is taken into account. Joint coordinates are applied to describe the motion of the links together with the homogeneous transformations technique. Lagrange equations of the second order are used when deriving the equations of motions. Joint forces and torques are determined based on recursive Newton-Euler algorithms. These joint forces are then used in the LuGre friction model, which allows to calculate the friction coefficients and friction forces. Numerical analyses performed here show the influence of various friction forces on the vibration level as perceived by the operator of the crane. The level of discomfort is discussed based on standards commonly used in the vehicle and transportation industry for evaluations of vibration comfort.

***Keywords**: crane dynamics, friction, seat vibrations.*

W niniejszym artykule przedstawiono model matematyczny żurawia leśnego, który jest stosowany do analizy dynamiki cyklu jego pracy. Uwzględniono podatność podparcia fotela operatora, napędów oraz podpór. Do opisu ruchu członów stosuje się współrzędne złączowe i macierze przekształceń jednorodnych. Do wyprowadzenia równań ruchu modelu żurawia zastosowano podejście bazujące na formalizmie równań Lagrange'a drugiego rodzaju. Siły i momenty węzłowe są określane na podstawie rekurencyjnego algorytmu Newtona-Eulera. Siły te są następnie wykorzystywane w modelu tarcia LuGre, który pozwala obliczyć współczynniki i siły tarcia. Przeprowadzone analizy numeryczne pokazują wpływ różnych sił tarcia na poziom drgań odczuwany przez operatora żurawia. Poziom dyskomfortu operatora wywołany przez drgania maszyny został oszacowany w oparciu o często stosowane w przemyśle samochodowym i transportowym odpowiednie standardy.

***Słowa kluczowe**: dynamika żurawia, tarcie, drgania siedzenia.*

Nomenclature

b, c, s, d, j	–	flexible supported base, crane, seat, drive and joint indexes, respectively
$(c,l)\big\|_{l=1,\ldots,7}$	–	link index
g	–	acceleration of gravity
$l^{(c,l)}$	–	length of link
$m^{(b)}, m^{(c,l)}, m^{(s)}$	–	masses of bodies
n_b	–	number of bodies
n_{dof}	–	number of generalised coordinates describing the motion of system
$\tilde{n}_{dof}^{(c,l)}$	–	number of generalised coordinates describing the motion of link (c,l) with respect to link $(c,l-1)$
$n_{dof}^{(b)}, n_{dof}^{(c,l)}, n_{dof}^{(s)}$	–	number of generalised coordinates describing the motion of link with respect to reference system $n_{dof}^{(c,l)} = n_{dof}^{(b)} + n_{dof}^{(c,l-1)} + \tilde{n}_{dof}^{(c,l)}, n_{dof}^{(c,l)} = 0$
$s_\beta^{(sup_\alpha,i)}, d_\beta^{(sup_\alpha,i)}\Big\|_{\substack{\alpha\in\{b,s\}\\ \beta\in\{x,y,z\}}}$	–	stiffness and damping coefficients of support

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

$s^{(d,i)}, d^{(d,i)}$ – stiffness and damping coefficients of drive

$t_f^{(j,i)}, f_f^{(j,i)}$ – friction torque in revolute joint, friction force in prismatic joint, respectively

$sup_{\alpha}|_{\alpha\in\{b,s\}}$ – support index

$t^{(d,i)}, f^{(d,i)}$ – driving torques and force, respectively

$\mathbf{r}_A^{(a)} = \begin{bmatrix} x_A^{(a)} & y_A^{(a)} & z_A^{(a)} & 1 \end{bmatrix}^T$ – vector of position of point A defined in the local coordinate system of link a

$\mathbf{H}^{(b)}, \mathbf{H}^{(c,l)}, \mathbf{H}^{(s)}$ – pseudo-inertia matrices

$\tilde{\mathbf{T}}^{(c,l)}$ – homogeneous transformation matrix from the local coordinate system of link (c,l) to the system of link $(c,l-1)$

$\mathbf{T}^{(b)}, \mathbf{T}^{(c,l)}, \mathbf{T}^{(s)}$ – homogeneous transformation matrices from the local coordinate systems to reference system

$\mathbf{T}^{(c,l)} = \mathbf{T}^{(b)}\mathbf{T}^{(c,l-1)}\tilde{\mathbf{T}}^{(c,l)}$, $\mathbf{T}^{(c,0)} = \mathbf{I}$

$\mathbf{T}_i = \frac{\partial \mathbf{T}}{\partial q_i}, \mathbf{T}_{i,j} = \frac{\partial^2 \mathbf{T}}{\partial q_i \partial q_j}$

Friction parameters

$\boldsymbol{\sigma}_0^{(j)} = \left(\sigma_{0,i}^{(j)}\right)_{i=1,\ldots,7}$, $\boldsymbol{\sigma}_1^{(j)} = \left(\sigma_{1,i}^{(j)}\right)_{i=1,\ldots,7}$, $\boldsymbol{\sigma}_2^{(j)} = \left(\sigma_{2,i}^{(j)}\right)_{i=1,\ldots,7}$ – vectors of stiffness, damping and viscous friction coefficients of bristles, respectively

$\mathbf{z}^{(j)} = \left(z_i^{(j)}\right)_{i=1,\ldots,7}$ – vector of deflections of bristles

$\boldsymbol{\mu}^{(j)} = \left(\mu_i^{(j)}\right)_{i=1,\ldots,7}$ – vector of friction coefficients

$\boldsymbol{\mu}_s^{(j)} = \left(\mu_{s,i}^{(j)}\right)_{i=1,\ldots,7}$, $\boldsymbol{\mu}_k^{(j)} = \left(\mu_{k,i}^{(j)}\right)_{i=1,\ldots,7}$ – vectors of static and kinetic friction coefficients, respectively

$\dot{\mathbf{q}}_S^{(c)} = \left(\dot{q}_{S,i}^{(c)}\right)_{i=1,\ldots,7}$ – vector of transition velocities between friction regimes

Parameters used for comfort assessment (BS 6841, 1987, ISO 2631-1, 1997)

$y(t)$ Time history of a signal (discrete, calculated by the numerical simulation). The signal $y(t)$ should be filtered using the frequency filters.

$RMS = \sqrt{\frac{1}{T}\int_0^T y^2(t)\mathrm{d}t}$ Root-Mean-Square describes the energetic content of a vibrational signal, T is the time period of vibration.

$C_f = \frac{\max\left(|y(t)|\right)}{RMS}$ Crest factor, to be used in presence of shocks (short high magnitude transient events).

$RMQ = \sqrt[4]{\frac{1}{T}\int_0^T y^4(t)\mathrm{d}t}$ Quad-Mean-Square, similar measure to RMS, but better describes the effect of vibration discomfort when $C_f > 9$.

$$K = \frac{1}{n\sigma^4}\sum_{i=1}^{n}\left(y_i - \tilde{y}\right)^4$$ Kurtosis, used for highly impulsive time domain signals, where n is the number of discrete data, σ is the standard deviation, $\tilde{y}$ is the average value of the analyzed signal.

$$VDV = \sqrt[4]{\int_0^T y^4(t)\mathrm{d}t}$$ Vibration dose value, gives a measurement of a cumulative vibration level received over a time period (often $8\mathrm{hr}$ or $16\mathrm{hr}$). Location and direction dependent filters to be applied.

$$T_{15} = \left(\frac{15}{VDV}\right)^4 t$$ The duration in seconds, required to reach the value of $VDV = 15\mathrm{ms}^{-1.75}$ which is defined as a severe discomfort. Parameter t is the duration of measured or calculated signal.

1. Introduction

Crane control problems are becoming increasingly important for designers and operators. The efficiency of load handling and increased safety and level of comfort for all personnel involved and crane operators are the main driving factors for developments in this field. For this purpose, modern machines are equipped with quite advanced and expensive sensors and other control devices. Computer simulations, even in the very early design phase, are a very useful approach that aims to reduce the overall cost and to eliminate some errors that can be predicted without building real physical prototypes of the crane. Moreover, one can easily simulate complex machine behaviour without the risk of damage or injuries. In this context, many crane models have been developed with varying complexity and level of details.

In this work, a mathematical model of a grab crane is presented [12, 13, 18, 22-24, 19, 20, 26-30]. Its main purpose is to investigate dynamics during various operation modes and handling scenarios, including estimation of loads, load motion, drive system control and others. The flexibility of the support system [22-24, 26-30], which is modelled as one-dimensional spring-damping elements, is taken into account in the mathematical model. In the similar way a flexible system is built to represent the operator's seat connection to the crane column. In the crane model developed here, all drive units also have flexible features [26-30]. Homogeneous transformation and joint coordinates are applied to describe the geometry of the crane units [7, 14, 15]. Equations of motion can be derived based on methods and algorithms presented in [11, 32]. The Runge-Kutta method of the fourth order is applied for integration of the governing equations of motion, with constant time step. In order to determine the joint forces and torques, which are necessary to calculate the friction forces and torques, each integration time step involves recursive loops defined as the Newton-Euler recursive dynamics task [5]. The friction coefficients for each kinematic pair are calculated by applying the LuGre friction model [1, 2, 17, 21] which takes into account pre-sliding displacement [4], as well as the Stribeck effect [25], among other features.

The influence of friction on forest crane dynamics has been discussed in some previous papers. The Dahl friction model was investigated in [27], while works [29, 30] concentrated on the LuGre friction model.

The analyses performed here concentrated on the dynamic properties of the crane with particular interest in the operator's seat properties during selected modes of operation. The LuGre friction model, with two different friction levels characterised by joint conditions, was assumed. Prediction of the discomfort level, determined by vibrations transmitted from the column to the seat and the body, was examined by taking into account the standard approach [6, 10] applied in vehicle N.V.H. (Noise, Vibration and Harshness) analyses). An analysis of the comfort level in various systems, including all vehicle types, buildings and other structures, is very important and required by certification authorities [9, 16] Many tests have been reported on how the human body perceives the discomfort [8]. It is a common practice to simulate and test comfort parameters also in special or construction machinery, as, for example, in [3] but especially in many branches of the ground vehicle industry.

2. Mathematical model of the forest crane

The model of the forest crane which consists of eight rigid links is presented in Fig. 1. These links are driven by flexible drive models generating drive torques $\mathbf{t}^{(d,i)}\big|_{i=1,2,3,7}$ and drive force $\mathbf{f}^{(d,4)}$ The whole crane (its platform) is supported on eight flexible legs. Similarly, mass-less spring damping elements model the connection between the seat and the crane's column.

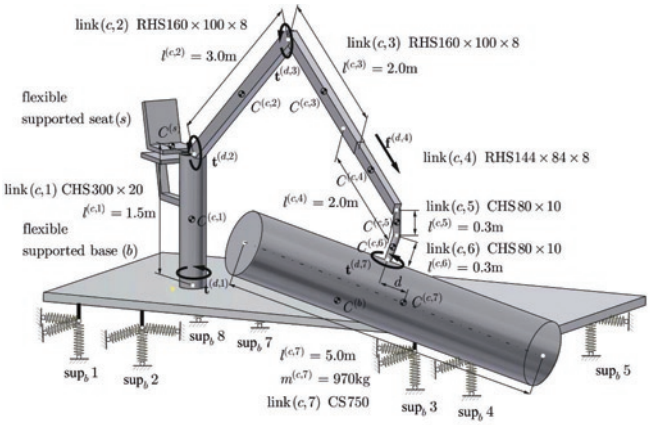


Fig. 1. Model of the forest crane

Joint coordinates and homogeneous transformation matrices are used to describe the geometry of the forest crane. The local coordinate systems and numeration of certain points are shown in Fig. 2.

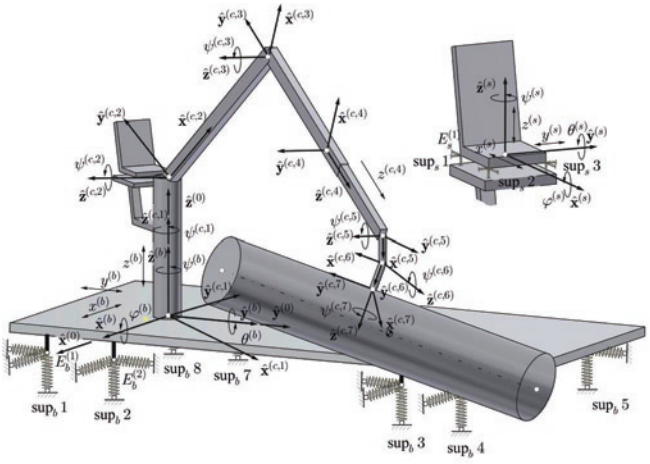


Fig. 2. Coordinate systems and notation applied to crane links

The vector of the model's generalised coordinates has the following form:

$$\mathbf{q} = \left(q_k\right)_{k=1,\ldots,n_{dof}} = \left[\left(\mathbf{q}^{(b)}\right)^T \quad \left(\mathbf{q}^{(c)}\right)^T \quad \left(\mathbf{q}^{(s)}\right)^T\right]^T \tag{1}$$

where:

$$\mathbf{q}^{(b)} = \left(q_j^{(b)}\right)_{j=1,\ldots,6} = \left[x^{(b)} \quad y^{(b)} \quad z^{(b)} \quad \psi^{(b)} \quad \theta^{(b)} \quad \varphi^{(b)}\right]^T$$

$$\mathbf{q}^{(c)} = \left(q_j^{(c)}\right)_{j=1,\ldots,7} = \left[\psi^{(c,1)} \quad \psi^{(c,2)} \quad \psi^{(c,3)} \quad z^{(c,4)} \quad \psi^{(c,5)} \quad \psi^{(c,6)} \quad \psi^{(c,7)}\right]^T$$

$$\mathbf{q}^{(s)} = \left(q_j^{(s)}\right)_{j=1,\ldots,6} = \left[x^{(s)} \quad y^{(s)} \quad z^{(s)} \quad \psi^{(s)} \quad \theta^{(s)} \quad \varphi^{(s)}\right]^T$$

The homogeneous transformation matrices have the following forms:

$$\mathbf{T}^{(\alpha)}\Big|_{\alpha\in\{b,s\}} \begin{bmatrix} c\psi^{(\alpha)}c\theta^{(\alpha)} & c\psi^{(\alpha)}s\theta^{(\alpha)}s\varphi^{(\alpha)} - s\psi^{(\alpha)}c\varphi^{(\alpha)} & c\psi^{(\alpha)}s\theta^{(\alpha)}c\varphi^{(\alpha)} + s\psi^{(\alpha)}s\varphi^{(\alpha)} & x^{(\alpha)} \\ s\psi^{(\alpha)}c\theta^{(\alpha)} & s\psi^{(\alpha)}s\theta^{(\alpha)}s\varphi^{(\alpha)} + c\psi^{(\alpha)}c\varphi^{(\alpha)} & s\psi^{(\alpha)}s\theta^{(\alpha)}c\varphi^{(\alpha)} - c\psi^{(\alpha)}s\varphi^{(\alpha)} & y^{(\alpha)} \\ -s\theta^{(\alpha)} & c\theta^{(\alpha)}s\varphi^{(\alpha)} & c\theta^{(\alpha)}c\varphi^{(\alpha)} & z^{(\alpha)} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\tilde{\mathbf{T}}^{(c,1)} = \begin{bmatrix} c\psi^{(c,1)} & -s\psi^{(c,1)} & 0 & 0 \\ s\psi^{(c,1)} & c\psi^{(c,1)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \tilde{\mathbf{T}}^{(c,2)} = \begin{bmatrix} c\psi^{(c,2)} & -s\psi^{(c,2)} & 0 & 0 \\ 0 & 0 & -1 & 0 \\ s\psi^{(c,2)} & c\psi^{(c,2)} & 0 & l^{(c,1)} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\tilde{\mathbf{T}}^{(c,3)} = \begin{bmatrix} c\psi^{(c,3)} & -s\psi^{(c,3)} & 0 & l^{(c,2)} \\ s\psi^{(c,3)} & c\psi^{(c,3)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \tilde{\mathbf{T}}^{(c,4)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -z^{(c,4)} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\tilde{\mathbf{T}}^{(c,5)} = \begin{bmatrix} c\psi^{(c,5)} & -s\psi^{(c,5)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -s\psi^{(c,5)} & -c\psi^{(c,5)} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \tilde{\mathbf{T}}^{(c,6)} = \begin{bmatrix} c\psi^{(c,6)} & -s\psi^{(c,6)} & 0 & l^{(c,5)} \\ 0 & 0 & 1 & 0 \\ -s\psi^{(c,6)} & -c\psi^{(c,6)} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\tilde{\mathbf{T}}^{(c,7)} = \begin{bmatrix} c\psi^{(c,7)} & -s\psi^{(c,7)} & 0 & 0 \\ 0 & 0 & 1 & l^{(c,6)} \\ -s\psi^{(c,7)} & -c\psi^{(c,7)} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad s\alpha^{(\beta)} = \sin\alpha^{(\beta)}, \quad c\alpha^{(\beta)} = \cos\alpha^{(\beta)}.$$

The equations of motion are derived using Lagrange equations of the second order. The following general form is commonly used:

$$\dot{\mathbf{z}}^{(j)} = \mathbf{LuGre}\left(t, \mathbf{q}^{(c)}, \mathbf{z}^{(j)}\right) \tag{2.1}$$

$$\mathbf{M}\ddot{\mathbf{q}} = \mathbf{e}\left(t, \mathbf{q}, \dot{\mathbf{q}}\right) + \mathbf{s}\left(\mathbf{q}, \dot{\mathbf{q}}\right) + \mathbf{d}\left(t, \mathbf{q}^{(c)}, \dot{\mathbf{q}}^{(c)}\right) - \mathbf{f}\left(t, \mathbf{q}^{(c)}, \dot{\mathbf{q}}^{(c)}\right) \tag{2.2}$$

where:

$$\left(LuGre_i\right)_{i=1,\ldots,7} = \dot{q}_i^{(c)}\left(1 - \frac{\sigma_{0,i}^{(j)} z_i^{(j)} \operatorname{sgn}\left(\dot{q}_i^{(c)}\right)}{\mu_{k,i}^{(j)} + \left(\mu_{s,i}^{(j)} - \mu_{k,i}^{(j)}\right)\exp\left(-\left(\frac{\dot{q}_i^{(c)}}{\dot{q}_{S,i}^{(c)}}\right)^2\right)}\right)$$

$$\boldsymbol{\mu}^{(j)} = \boldsymbol{\sigma}_0^{(j)}\mathbf{z}^{(j)} + \boldsymbol{\sigma}_1^{(j)}\dot{\mathbf{z}}^{(j)} + \boldsymbol{\sigma}_2^{(j)}\dot{\mathbf{q}}^{(c)}$$

$$\mathbf{M} = \sum_{\alpha\in\{b,c,s\}} \mathbf{M}^{(\alpha)}$$

$$\mathbf{M}^{(b)} = \begin{bmatrix} \mathbf{M}_{1,1}^{(b)} & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{M}^{(c)} = \begin{bmatrix} \mathbf{M}_{1,1}^{(c)} & \cdots & \mathbf{M}_{1,j}^{(c)} & \cdots & \mathbf{M}_{1,n_b-1}^{(c)} & 0 \\ \vdots & \ddots & \vdots & \iddots & \vdots & \vdots \\ \mathbf{M}_{i,1}^{(c)} & \cdots & \mathbf{M}_{i,j}^{(c)} & \cdots & \mathbf{M}_{i,n_b-1}^{(c)} & 0 \\ \vdots & \iddots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{M}_{n_{b-1},1}^{(c)} & \cdots & \mathbf{M}_{n_{b-1},j}^{(c)} & \cdots & \mathbf{M}_{n_{b-1},n_{b-1}}^{(c)} & 0 \\ 0 & \cdots & 0 & \cdots & 0 & 0 \end{bmatrix}$$

$$\mathbf{M}^{(s)} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{M}_{1,1}^{(s)} \end{bmatrix}$$

$$\mathbf{M}_{1,1}^{(\alpha)}\Big|_{\alpha\in\{b,s\}} = \left(m_{i,j}^{(\alpha)}\right)_{i,j=1,\ldots,n_{dof}^{(\alpha)}}, \quad m_{i,j}^{(\alpha)} = \operatorname{tr}\left\{\mathbf{T}_i^{(\alpha)}\mathbf{H}^{(\alpha)}\left(\mathbf{T}_j^{(\alpha)}\right)^T\right\},$$

$$\mathbf{M}_{i,j}^{(c)} = \sum_{l=\max\{i,j\}}^{n_b-1} \mathbf{M}_{i,j}^{(c,l)}, \quad m_{i,j}^{(c,l)} = \operatorname{tr}\left\{\mathbf{T}_i^{(c,l)}\mathbf{H}^{(c,l)}\left(\mathbf{T}_j^{(c,l)}\right)^T\right\},$$

$$\mathbf{e} = \sum_{\alpha\in\{b,c,s\}} \mathbf{e}^{(\alpha)},$$

$$\mathbf{e}^{(b)} = \begin{bmatrix} \mathbf{e}_1^{(b)} \\ 0 \end{bmatrix}, \quad \mathbf{e}^{(c)} = \begin{bmatrix} \mathbf{e}_1^{(c)} \\ \vdots \\ \mathbf{e}_i^{(c)} \\ \vdots \\ \mathbf{e}_{n_{b-1}}^{(c)} \\ 0 \end{bmatrix}, \quad \mathbf{e}^{(s)} = \begin{bmatrix} 0 \\ \mathbf{e}_1^{(s)} \end{bmatrix},$$

$$\mathbf{e}_1^{(\alpha)}\Big|_{\alpha\in\{b,s\}} = -\left(\mathbf{h}^{(\alpha)} + \mathbf{g}^{(\alpha)}\right),$$

$$\mathbf{h}^{(\alpha)} = \left(h_i^{(\alpha)}\right)_{i=1,\ldots,n_{dof}^{(\alpha)}},$$

$$h_i^{(\alpha)} = \sum_{m=1}^{n_{dof}^{(\alpha)}}\sum_{n=1}^{n_{dof}^{(\alpha)}} \operatorname{tr}\left\{\mathbf{T}_i^{(\alpha)}\mathbf{H}^{(\alpha)}\left(\mathbf{T}_{m,n}^{(\alpha)}\right)^T\right\}\dot{q}_m^{(\alpha)}\dot{q}_n^{(\alpha)},$$

$$\mathbf{g}^{(\alpha)} = \left(g_i^{(\alpha)}\right)_{i=1,\ldots,n_{dof}^{(\alpha)}}, \quad g_i^{(\alpha)} = m^{(\alpha)}\mathbf{g}\mathbf{j}_3\mathbf{T}_i^{(\alpha)}\mathbf{r}_{C^{(\alpha)}}^{(\alpha)},$$

$$\mathbf{e}_i^{(c)} = -\sum_{l=i}^{n_b-1}\left(\mathbf{h}_i^{(c,l)} + \mathbf{g}_i^{(c,l)}\right),$$

$$\mathbf{h}_i^{(c,l)} = \left(h_{n_{dof}^{(c,i-1)}+k}^{(c,l)}\right)_{k=1,\ldots,n_{dof}^{(c,i)}},$$

$$h_i^{(c,l)} = \sum_{m=1}^{n_{dof}^{(c,l)}} \sum_{n=1}^{n_{dof}^{(c,l)}} \mathrm{tr}\left\{ \mathbf{T}_i^{(c,l)} \mathbf{H}^{(c,l)} \left(\mathbf{T}_{m,n}^{(c,l)} \right)^T \right\} \dot{q}_m^{(c,l)} \dot{q}_n^{(c,l)} \,,$$

$$\mathbf{g}_i^{(c,l)} = \left(g^{(c,l)}_{n_{dof}^{(c,i-1)}+k} \right)_{k=1,\ldots,n_{dof}^{(c,i)}}, \quad g_i^{(c,l)} = m^{(c,l)} \mathbf{g} \mathbf{j}_3 \mathbf{T}_i^{(c,l)} \mathbf{r}_{C^{(c,l)}},$$

$$\mathbf{s} = \sum_{\alpha \in \{b,s\}} \mathbf{s}^{(sup_\alpha)}, \quad \mathbf{s}^{(sup_\alpha)} = \left(s_k^{(sup_\alpha)} \right)_{k=1,\ldots,n_{nof}},$$

$$s_k^{(sup_b)} = \begin{cases} -\sum_{i=1}^{n_{sup_b}} \left(\left(\frac{\partial \mathbf{e}^{(sup_b,i)}}{\partial q_k} \right)^T \mathbf{S}^{(sup_b,i)} \mathbf{e}^{(sup_b,i)} + \left(\frac{\partial \dot{\mathbf{e}}^{(sup_b,i)}}{\partial \dot{q}_k} \right)^T \mathbf{D}^{(sup_b,i)} \dot{\mathbf{e}}^{(sup_b,i)} \right) & \text{if } q_k \in \left\{ \mathbf{q}^{(b)} \right\} \\ 0 & \text{otherwise} \end{cases}$$

$$s_k^{(\sup_s)} = \begin{cases} \sum_{i=1}^{n_{sup_s}} \left(\left(\frac{\partial \mathbf{e}^{(sup_s,i)}}{\partial q_k} \right)^T \mathbf{S}^{(sup_s,i)} \mathbf{e}^{(sup_s,i)} + \left(\frac{\partial \dot{\mathbf{e}}^{(sup_s,i)}}{\partial \dot{q}_k} \right)^T \mathbf{D}^{(\sup s,i)} \dot{\mathbf{e}}^{(\sup s,i)} \right) & \text{if } q_k \in \left\{ \mathbf{q}^{(b)}, \psi^{(c,1)} \right\} \\ -\sum_{i=1}^{n_{sup_s}} \left(\left(\frac{\partial \mathbf{e}^{(sup_s,i)}}{\partial q_k} \right)^T \mathbf{S}^{(sup_s,i)} \mathbf{e}^{(sup_s,i)} + \left(\frac{\partial \dot{\mathbf{e}}^{(sup_s,i)}}{\partial \dot{q}_k} \right)^T \mathbf{D}^{(\sup s,i)} \dot{\mathbf{e}}^{(\sup s,i)} \right) & \text{if } q_k \in \left\{ \mathbf{q}^{(s)} \right\} \\ 0 & \text{otherwise} \end{cases}$$

Based on the formulation presented above, a computer program was developed using the Visual C++ environment. The standard Runge-Kutta method of the fourth order was applied, with a constant time step 10^{-4}s .

3.1. Crane operation scenarios and load cases

The crane motion sequence is assumed as presented in Fig. 3. At time $t = 0$s the load is resting on a platform. After two seconds, the load is lifted up by increasing angle of the jib. Then crane column rotates (reaching $90°$ at $t = 5$s), and simultaneously the telescopic motion begins at time $t = 5$s . For the final column rotation angle, $\psi_{dr}^{(c,1)} = 180°$, the telescopic motion stops reaching minimum length at $t = 9$s and the cycle finishes with the load positioned down to a platform on the opposite side.

Crane loading conditions are:

1) empty grab (E) - operation with unloaded crane,

2) crane with load (F) - operation with $m^{(c,7)}$ load mass.

The empty load case scenario is considered to have identical driving functions, with reverse order/values - returning to the pick-up position same as in $t = 0$s .

As indicated in Fig. 1, the distance d represents offset between the load's center of gravity and axis of the joint and drive $\mathbf{t}^{(d,7)}$. Influence of this distance was considered as one of important parameters in the analysis. Each working cycle, in practical condition, will

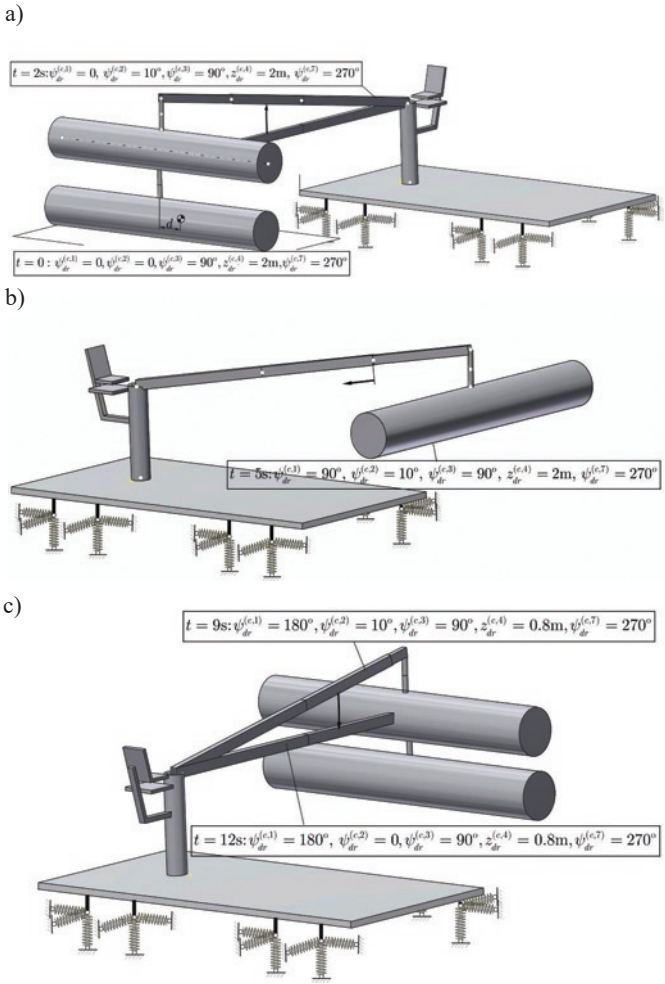


Fig. 3. Crane operation sequences

be characterized by different value of d, caused by not ideal mass distribution of trunks and misalignment of length, initial position of the load on a storage platform an many other reasons. For the performed study presented in this work, the range of $d = \pm 20\text{cm}$ is assumed as typical.

Load cases analyzed in this work are listed in Fig. 4. „Empty" crane cases are defined for the same sequence as loaded - just with no load attached to the grab. „Loaded" cases are performed with the tree trunk mass $m^{(c,7)} = 970\text{kg}$.

$x = 0$ (no damping)
$x = 1$ (nominal damping)

$L = F$ (loaded crane)
$L = E$ (empty hook)

$Dx - Fy - dz - L$

$y = 1$ (friction coefs. "Set-1")
$y = 2$ (friction coefs. "Set-2")

$z = 0$ (no load CoG offset)
$z = 10$ (10cm offset of load CoG)
$z = 20$ (20cm offset of load CoG)

Fig. 4. Analyzed cases and symbol assignation

The cases without damping in seat mounts are analyzed just for comparison of the damping effect on operator vibration level. Analysis of all possible cases, leading to a more general evaluation of the particular crane design, is a large task and will not be presented in details.

3.2. Main parameters of the system

All the crane mass components have assigned properties according to geometry properties (sections) as indicated in Fig. 1. Operator's seat mass is assumed as a combined mass of the operator (one single

Table 1. Parameters of the crane supports

	1	2	3	4	5	6	7	8
$x_{E_b^{(i)}}^{(b)}$ [m]	1.5	1.5	1.5	1.5	–1.5	–1.5	–1.5	–1.5
$y_{E_b^{(i)}}^{(b)}$ [m]	0	1.0	8.0	9.0	9.0	8.0	1.0	0
$z_{E_b^{(i)}}^{(b)}$ [m]					–0.57			
$s_{\alpha}^{(sup_b,i)}\Big\|_{\alpha\in\{x,y\}}$ [Nm^{-1}]					$3\cdot10^6$			
$s_{Z}^{(sup_b,i)}$ [Nm^{-1}]					$1\cdot10^7$			
$d_{\alpha}^{(sup_b,i)}\Big\|_{\alpha\in\{x,y\}}$ [Nsm^{-1}]					$5\cdot10^4$			
$d_{Z}^{(sup_b,i)}$ [Nsm^{-1}]					$9\cdot10^4$			

Table 2. Parameters of the seat supporting elements

	1	2	3	4
$x_{E_s^{(i)}}^{(s)}$ [m]	–0.25	0.25	0.25	–0.25
$y_{E_s^{(i)}}^{(s)}$ [m]	0.25	0.25	–0.25	–0.25
$z_{E_s^{(i)}}^{(s)}$ [m]			–0.05	
$x_{E_s^{(i)}}^{(c,1)}$ [m]	–0.85	–0.35	–0.35	–0.85
$y_{E_s^{(i)}}^{(c,1)}$ [m]	0.25	0.25	–0.25	–0.25
$z_{E_s^{(i)}}^{(c,1)}$ [m]			1.4	
$s_{\alpha}^{(sup_s,i)}\Big\|_{\alpha\in\{x,y\}}$ [Nm^{-1}]			10^3	
$s_{Z}^{(sup_s,i)}$ [Nm^{-1}]			$9\cdot10^3$	
$d_{\alpha}^{(sup_s,i)}\Big\|_{\alpha\in\{x,y\}}$ [Nsm^{-1}]			40	
$d_{Z}^{(sup_s,i)}$ [Nsm^{-1}]			120	

Table 3. Friction parameters

		1	2	3	4	5	6	7
Set-1	$\mu_{k,i}^{(j)}$	0.15	0.10	0.10	0.10	0.15	0.15	0.15
	$\mu_{s,i}^{(j)}$	0.20	0.15	0.20	0.20	0.20	0.20	0.20
Set-2	$\mu_{k,i}^{(j)}$	0.07	0.07	0.05	0.20	0.20	0.20	0.10
	$\mu_{s,i}^{(j)}$	0.10	0.10	0.15	0.30	0.35	0.35	0.15
	$\dot{q}_{S,i}^{(c)}\left[\text{rad s}^{-1},\text{ms}^{-1}\right]$				0.005			
	$\sigma_{0,i}^{(j)}\left[\text{Nmrad}^{-1},\text{Nm}^{-1}\right]$		10^5		10^7		10^5	
	$\sigma_{1,i}^{(j)}\left[\text{Nmsrad}^{-1},\text{Nsm}^{-1}\right]$				5			
	$\sigma_{2,i}^{(j)}\left[\text{Nmsrad}^{-1},\text{Nsm}^{-1}\right]$				0			

mass 80kg) and the seat self-mass equal to 25kg ; hence the total mass of the seat-operator is assumed to be $m^{(s)}=105\text{kg}$.

The geometrical parameters of the base supports are contained in Tab.1. The assumed stiffness and damping coefficients are also presented. The mounting point locations and spring-damping elements connecting the crane column and seat body are specified as presented in Tab. 2. The friction parameters are defined in Tab. 3. Two sets are defined in order to distinguish different conditions of the joints, i.e. normal (**Set-1**) well lubricated, and poorly greased joints (**Set-2**).

3.3. Vibration assessment

Human perception of discomfort is not unique to every person, i.e. perceived comfort depends on many factors. The reference standards used in the industry are, e.g. BS 6841 and ISO 2631-1 – Fig. 5. These standards were used in this work to assess the level of vibration and discomfort as perceived by the operator. A similar approach can be applied for vehicle dynamics and an estimation of the ride comfort, which was also applied in the optimisation routines yielding the desired minimum discomfort [31].

In general, the body is exposed to vibration in combination of all 6 directions (translations and rotations), but in present work only x,y,z axial signals are considered. Seated crane operator will perceive vibration at the back and at the feet (as well as hands and the head may also be of importance). The assigned locations investigated and appropriate filters are indicated in Fig. 6.

Each considered location for comfort evaluation is characterized by the following formulas:

$$PV_{RMS}^{(\alpha)}\Big|_{\alpha\in\{O_b,O_s,O_f\}}=\sqrt{\sum_{\beta\in\{x,y,z\}}\left(f_\beta^{(\alpha)}RMS_\beta^{(\alpha)}\right)^2},\qquad(3.1)$$

$$PV_{VDV}^{(\alpha)}\Big|_{\alpha\in\{O_b,O_s,O_f\}}=\sqrt[4]{\sum_{\beta\in\{x,y,z\}}\left(f_\beta^{(\alpha)}VDV_\beta^{(\alpha)}\right)^4}.\qquad(3.2)$$

A single values describing comfort level for the whole construction, are defined as the sum of all location values. Proposed approach will enable us to summarize the comfort as one single values, which can be compared between different designs or different operations. The following definitions apply:

$$SV_{RMS}^{(c)}=\sqrt{\sum_{\alpha\in\{O_b,O_s,O_f\}}\left(PV_{RMS}^{(\alpha)}\right)^2},\qquad(3.3)$$

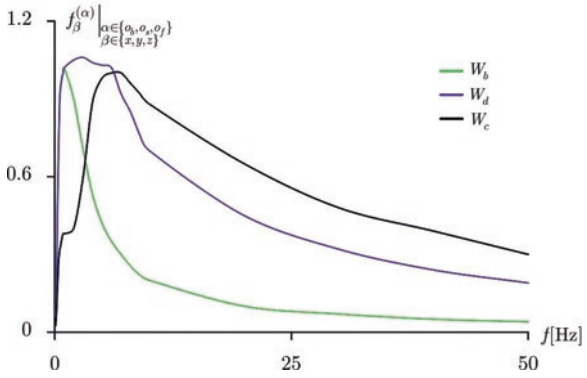


Fig. 5. Frequency-weighting filters to be applied in comfort assessment

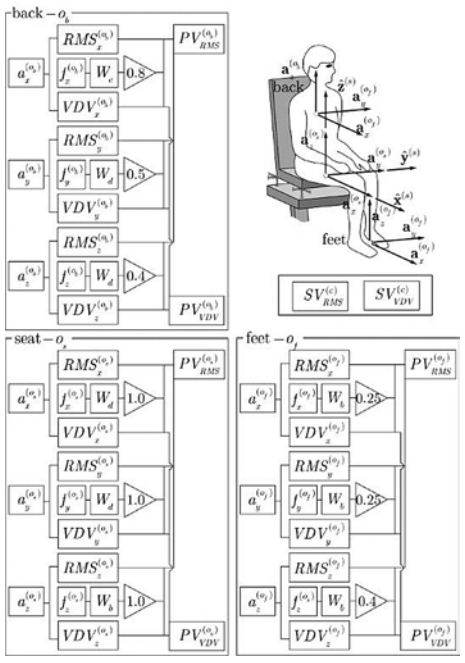


Fig. 6. Locations of interest and frequency-weighting filters

$$SV_{VDV}^{(c)}=\sqrt{\sum_{\alpha\in\{O_b,O_s,O_f\}}\left(PV_{VDV}^{(\alpha)}\right)^2}.\qquad(3.4)$$

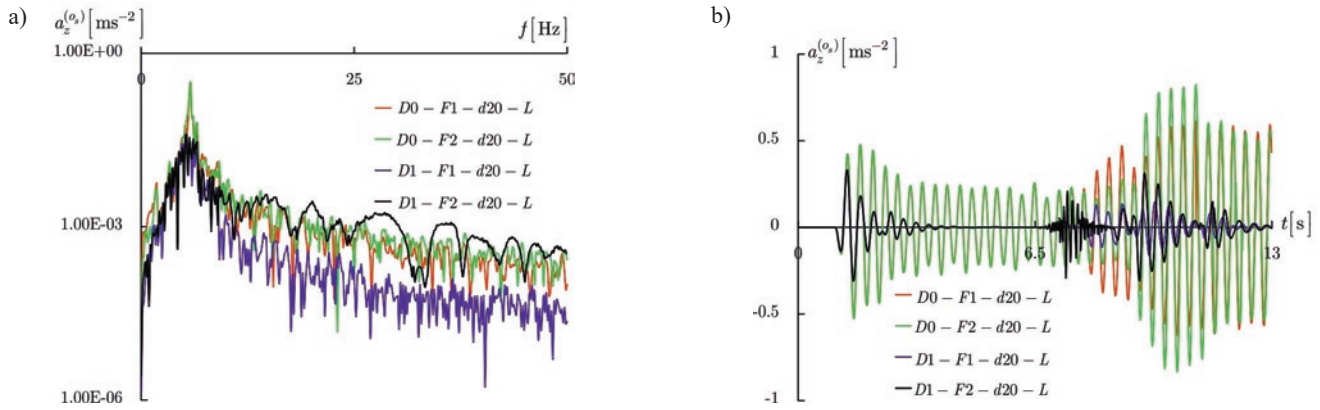


Fig. 7. Accelerations calculated for seat position in z direction; loaded crane with and without damping in seat mount elements. Frequency plots are on the left; time histories are on the right.

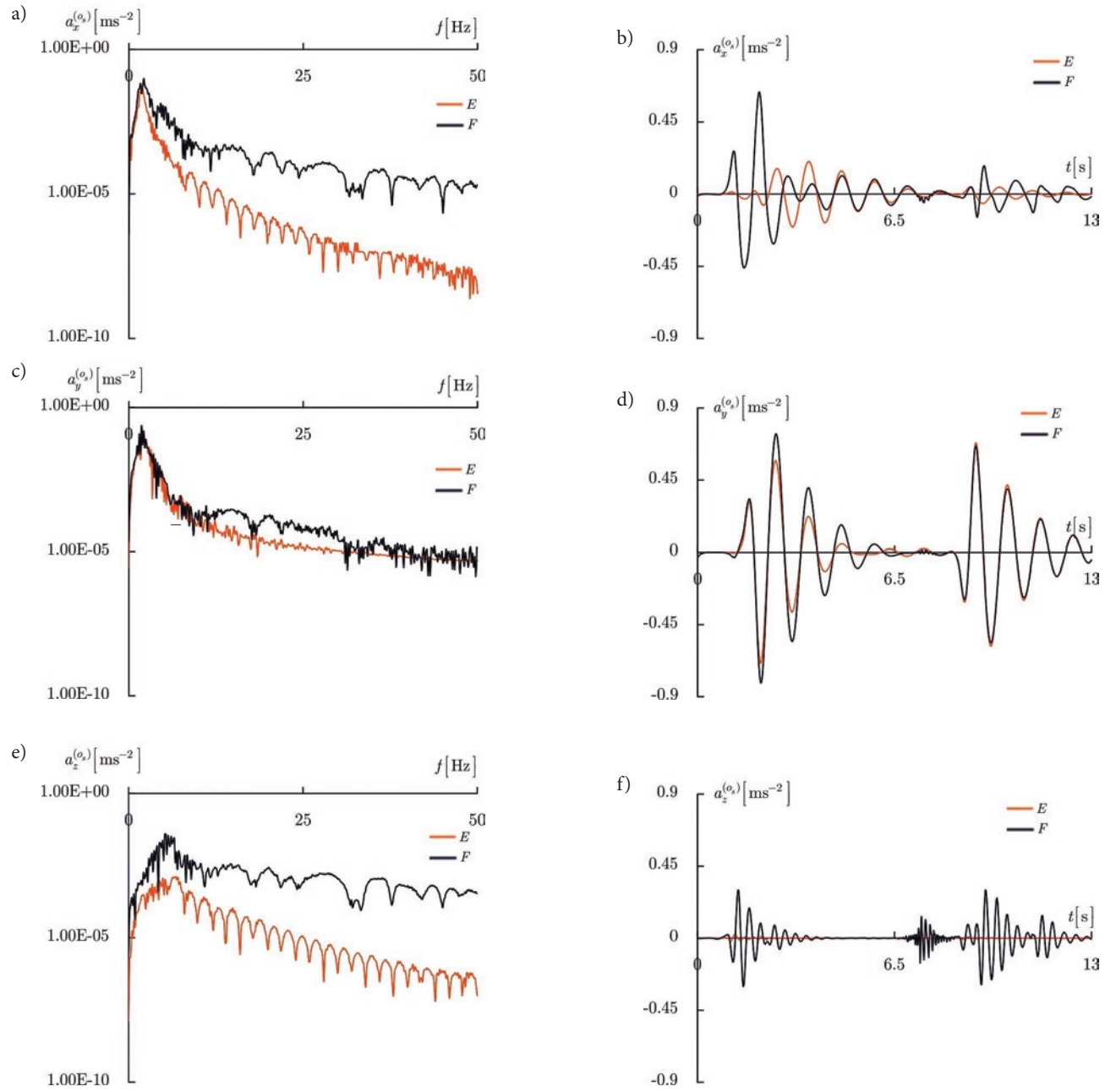


Fig. 8. Accelerations (filtered) calculated for seat position in x, y, z direction operation; with empty and loaded crane (load cases $D1-F2-d0-E$ and $D1-F2-d20-F$). Frequency (filtered) plots on the left; time histories on right.

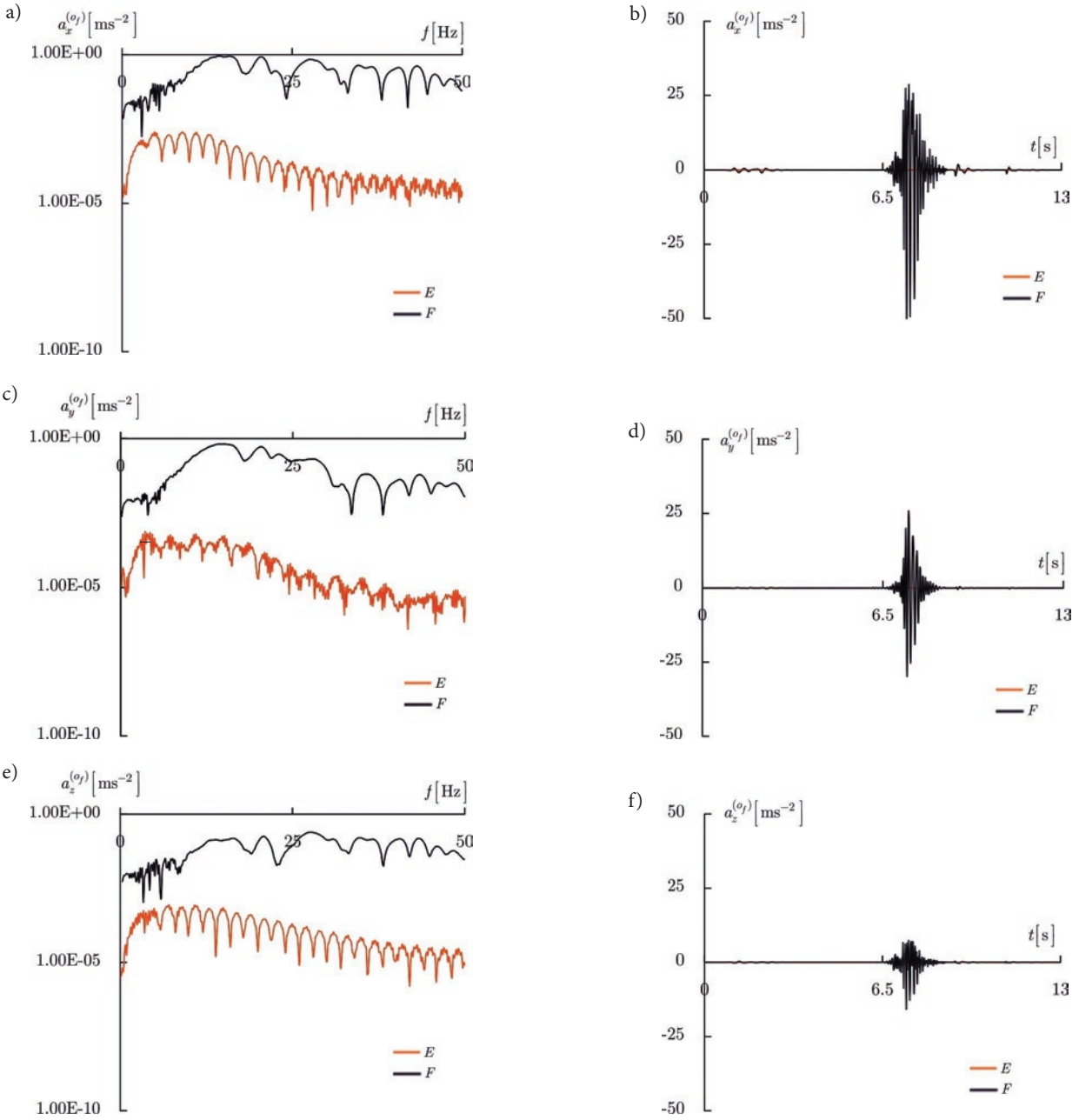


Fig. 9. Accelerations (filtered) calculated for feet rest position in x, y, z direction; operation with empty and loaded crane (load cases $D1-F2-d0-E$ and $D1-F2-d20-F$). Frequency (filtered) plots on the left; time histories on the right.

Other definitions (such as the „running *RMS*" (BS 6841) and peak-to-peak could also have been applied to assess the effect on vibration discomfort [3]. The selection of most appropriate parameter will be of designer choice and should be done based on type of operation performed by the crane. This could yield to the optimization routines resulting in minimization of the discomfort, but changing for example the mount characteristics or locations of the support points.

4. Simulation results

4.1. Example time histories

Some example results are shown in Fig. 7 – all series show the time histories and frequencies calculated for the vertical acceleration of the seat (unfiltered results are shown). The damping effect on the seat mounting points is examined. Typically, accelerations for the system without damping would be too conservative, even if a simple seat system for some poor designs may not have any damping elements (only structural damping).

Accelerations calculated for friction parameters Set-2 and for empty hook operation (unloaded crane) are presented in Fig. 8. A similar set of results for the feet rest is presented in Fig. 9. The friction effect on the transnational motion of the jib (drive activated during rotation) is well evident on the feet.

The influence of the load centre of gravity ($d = 0\text{cm}$, $d = 10\text{cm}$, $d = 20\text{cm}$) is presented in Fig. 10. The results show the strong influence of the load centre of gravity (and induced moments) on crane dynamics. A high friction force is generated in the jib (during the telescopic phase of motion) and the peaks are strongly visible. Most of

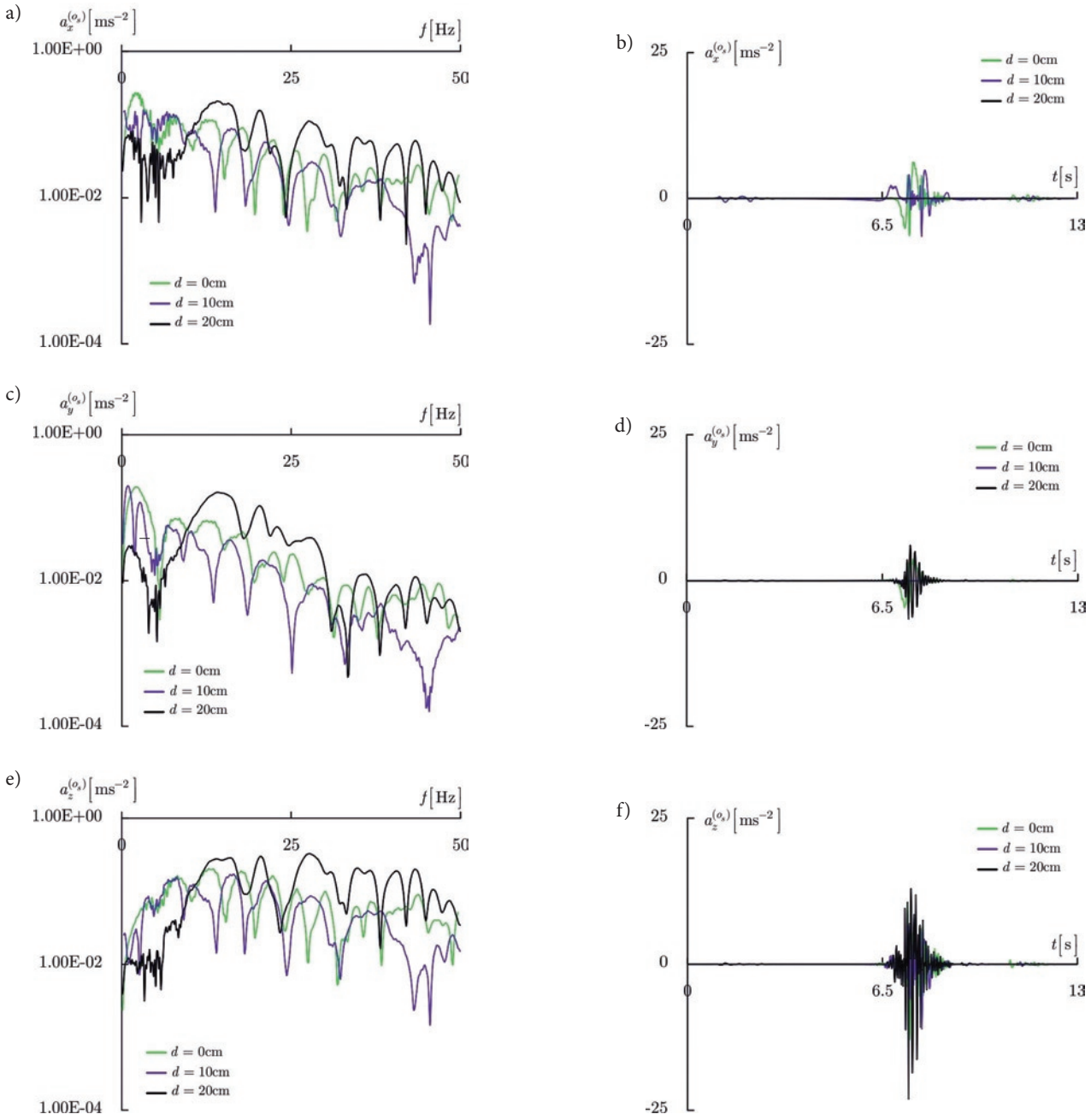


Fig. 10. Accelerations (filtered) calculated for seat base position in x, y, z direction; operation with loaded crane and different d value. Frequency (filtered) plots are on the left; time histories are on the right. Load cases: $D1-F2-d0-F$, $D1-F2-d10-F$ and $D1-F2-d20-F$

these peaks are transferred to the feet floor and seat base. Seat suspension, however, provides good isolation and the peaks visible between $6 \div 9$ s are not reflected in the seat points (such as the seat and backrest).

4.2. Vibration level - indexes

Various indexes related to perceived discomfort are shown in this section for the analysed crane operation scenarios.

Some indexes for load scenarios and design parameters listed in section 3.2 and 3.1 are presented in Tab. 6. Calculations have been performed with friction set **Set-2** (case $D1-F2-d0-E$).

Similarly, the results for crane handling with a full load considering two different friction coefficient sets **Set-1** and **Set-2** are listed in Tab. 7 and Tab. 8.

Some parameters defined in section 3.3, calculated for assumed friction parameters in crane joint as defined by **Set-1** and **Set-2** are presented in Fig. 11. Filtered RMS values, reduced to one single value, $SV_{RMS}^{(c)}$, are shown (calculated as indicated in Fig. 6). Results are calculated for the whole crane seat comfort when the seat suspension is included. For comparison, the results indicate also the level of RMS when rigid support would be assumed.

The time (in hours for **Set-1** and minutes for **Set-2**) required to accumulate desired level of dose value (i.e. $15\mathrm{ms}^{-1.75}$) is indicated

Table 6. Vibration discomfort parameters, case D2-F2-d0-E

Parameter	rigid seat			flexible seat			back			feet		
	x	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*
RMS	0.01	0.00	0.00	0.06	0.21	0.00	0.07	0.22	0.00	0.01	0.00	0.00
PV_{RMS}		0.01			0.22			0.12			0.00	
SV_{RMS}					0.25							
RMQ	0.01	0.00	0.01	0.10	0.32	0.01	0.10	0.33	0.01	0.03	0.01	0.01
VDV	0.02	0.00	0.01	0.19	0.61	0.01	0.20	0.63	0.01	0.05	0.01	0.02
PV_{VDV}		0.02			0.61			0.63			0.05	
K	20.51	12.68	62.06	5.44	5.21	20.25	5.44	5.03	5.68	54.81	24.95	68.61
C_f	6.90	5.32	11.66	3.21	3.28	6.99	3.20	3.18	3.09	10.64	8.37	11.17

Table 7. Vibration discomfort parameters, case D1-F1-d0-F

Parameter	rigid seat			flexible seat			back			feet		
	x	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*
RMS	0.19	0.04	0.19	0.13	0.24	0.06	0.16	0.25	0.03	0.50	0.09	0.15
PV_{RMS}		0.27			0.28			0.18			0.14	
SV_{RMS}					0.36							
RMQ	0.38	0.09	0.47	0.23	0.36	0.10	0.30	0.37	0.06	1.21	0.19	0.35
VDV	0.72	0.17	0.89	0.44	0.68	0.20	0.56	0.70	0.11	2.30	0.36	0.66
PV_{VDV}		0.97			0.71			0.48			0.58	
K	13.90	20.27	37.11	10.39	4.95	9.98	11.23	4.75	13.14	34.62	22.47	31.63
C_f	6.06	7.10									7.72	9.90

Table 8. Vibration discomfort parameters, case D1-F2-d20-F

Parameter	rigid seat			flexible seat			back			feet		
	x	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*
RMS	1.04	0.73	2.01	0.13	0.24	0.07	0.17	0.25	0.04	5.11	3.10	1.38
PV_{RMS}		2.38			0.28			0.18			1.59	
SV_{RMS}					1.63							
RMQ	1.67	1.07	2.45	0.24	0.37	0.15	0.31	0.37	0.07	6.53	2.28	1.64
VDV	3.18	2.04	4.65	0.45	0.69	0.29	0.59	0.71	0.13	12.39	4.33	3.11
PV_{VDV}		4.92			0.73			0.51			3.13	
K	29.38	29.10	53.16	11.69	4.88	9.10	12.21	4.69	8.28	64.23	65.46	44.08
C_f	8.91	8.56	11.87	5.10	3.36	4.59	5.47	3.20	4.30	13.02	12.46	11.09

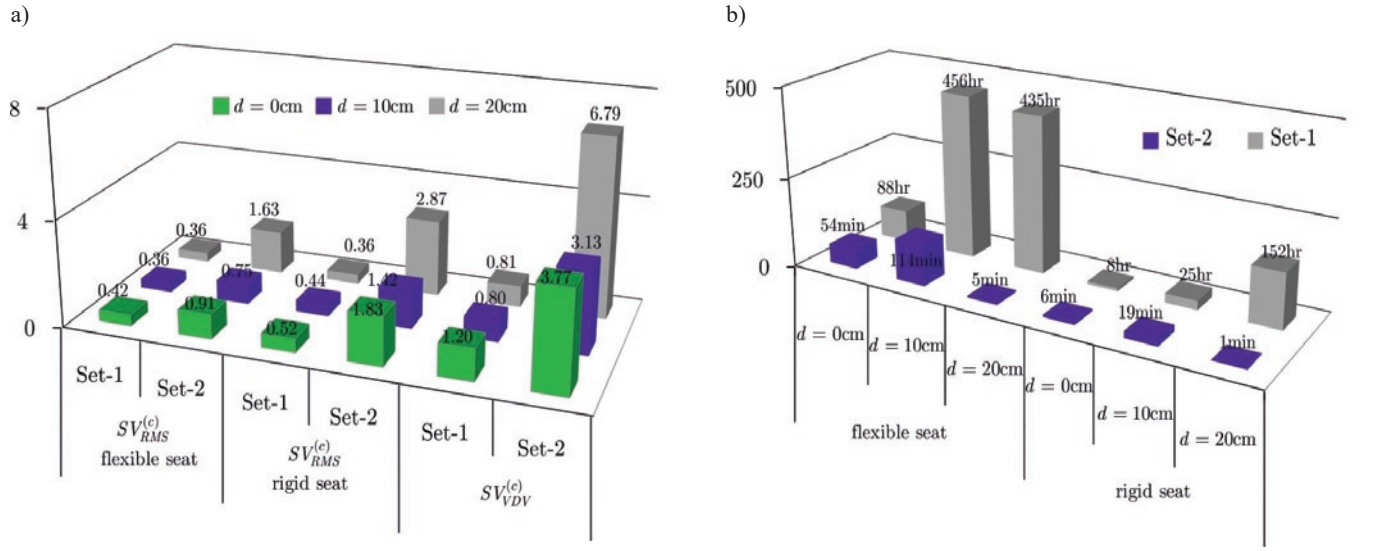


Fig. 11. $SV_{RMS}^{(c)}$ *and* $SV_{VDV}^{(c)}$ *values (left),* $T_{15}^{(c)}$ *time (right). Loaded crane operation, friction coefficients:* ***Set-1*** *and* ***Set-2****.*

in Fig. 11. This is the time calculated considering total vibration dose value from relation:

$$T_{15}^{(c)} = \left(\frac{15}{\sqrt{\sum_{\alpha \in \{O_b, O_s, O_f\}} \left(VDV_c^{(\alpha)} \right)^2}} \right)^4 t \qquad (4.1)$$

where $t = 13\text{s}$.

For the friction coefficients assumed in **Set-1**, the general level of discomfort can be estimated as "*a little uncomfortable*" when working with or without dampened seat. Also the exposure time is large, especially for the crane equipped with suspended seat. The values of $T_{15}^{(c)}$ show much greater difference between isolated and rigid seats. As this measure better suits for the characteristics of signals, it should be more important than the $SV_{RMS}^{(c)}$ index. Hence, much worse conditions for operator are expected when the seat is rigidly connected to the column. In the worst case, working 8hr in such conditions will be perceived as severe/huge discomfort and potentially dangerous for health.

Different tendency is obtained for **Set-2** friction coefficients: significantly worse results (bigger discomfort) were calculated when handling the trunk with larger offset between the grab axis and its center of gravity ($d = 20\text{cm}$). This conditions generated double $SV_{RMS}^{(c)}$ value and reduced the time $T_{15}^{(c)}$ to almost minimum (few minutes). For a seat with dampening elements, the RMS at level $0.8 \div 1.6$ can be classified as moderate discomfort, while for the case without such elements (rigid seat), the same operation will lead to very an uncomfortable level.

When considering the duration of such a level of vibration, the operator should not work on the „rigid seat", when some higher friction would occur. Assuming, that the trunks are not always transported in condition $d = 20\text{cm}$, but mostly around $d = (0 \div 10)\text{cm}$, operator could work on the crane within some limited time $1 \div 2\ \text{hr}$ until the VDV would become unacceptable. Considering the typical use of the forest crane (loading/unloading time is not dominating during the whole day, typically), estimated results shows how the current machine design could influence on health aspects.

Different results are obtained for operation with unloaded crane; for this conditions the vibrations generated due to operation of empty crane do not cause any significant discomfort for the operator, and the level of friction is also not important (for indexes used in the discomfort assessment). Dominated effect, in the case of unloaded crane, is the motion due to drives, and since there is no load applied (expect for the inertia induced forces), the crane operator perceived discomfort is practically identical. The combined value for the seat: $SV_{RMS}^{(O_s)} = 0.25$ (both friction sets) can be classified as „*not uncomfortable*".

Presented results cover only one configuration of the seat suspension. Several iterations are normally performed in order to find a good balance between desired comfort and design constrains.

5. Conclusions

The model and computer software developed here were applied to analyse the influence of friction on crane operators. The calculated responses can be useful for designers and early phase design can be examined. The results confirm that high friction can have a significant impact on human discomfort. This parameter, as well as many others, should be taken into account in the initial phase at crane design.

The model presented in this paper can be applied to many other aspects of typical working scenarios. Only the acceleration results are analysed in more detail, but the computer model allows us to investigate much more parameters, such as loads in specific components, the drive function effects on these loads, or optimisation of geometry and stiffness parameters. The model's simplicity and effectiveness are also important, especially for examining specific aspects of the system when many variant calculations are needed.

More advanced models taking into account, for example, the flexibility of the links can also be the correct direction for more detailed vibration/comfort analyses. This can be addressed in a similar way as presented in this work, i.e. just by extending the model. The disadvantage will be the time required for the analyses and the fact that some more complex input data are required.

References

1. Augustynek K, Urbaś A. Comparison of bristles' friction models in dynamics analysis of spatial linkages. Mechanics Research Communications 2017, https://doi.org/10.1016/j.mechrescom.2017.01.003.
2. Åström K J, Canudas-de-Witt C. Revisiting the LuGre model. IEEE Control Systems Magazine. Institute of Electrical and Electronics Magazine 2008; 28(6): 101-114, https://doi.org/10.1109/MCS.2008.929425.
3. Cann A P, Salmoni A W, Vi P, Eger T R. An exploratory study of whole-body vibration exposure and dose while operating heavy equipment in the construction industry. Applied Occupational and Environmental Hygiene 2003; 18: 999-1005, https://doi.org/10.1080/715717338.
4. Courtney-Pratt J S, Eisner E. The effect of a tangential force on the contact of metallic bodies. Proceedings of the Royal Society 1957; A: 529-550.
5. Craig J J. Introduction to robotics. Mechanics and control. Addison-Wesley Publishing Company, Inc., 1989.
6. BS 6841. Guide to measurement and evaluation of human exposure to whole-body mechanical vibration and repeated shock. British Standard, 1987.
7. Denavit J, Hartenberg R S. A kinematic notation for lower-pair mechanisms based on matrices. Journal of Applied Mechanics 1995; 23: 215-221.
8. Giacomin M, Hacaambwa T M. Performance of ISO2631 and BS6841 Comfort Criteria for Evaluating Automobile Road Vibrations, O1A1083, ATA 7th International Conference on the Role of Experimentation in the Modern Automotive Product Development Process, Florence, Italy, May 23-25, 2001.
9. Griffin M J. Handbook of Human Vibration. Elsevier Academic Press, 1990.
10. ISO 2631-1. Mechanical vibration and shock - Evaluation of human exposure to whole body vibration. Part 1: General requirements. International Organization for Standardization, 1997.
11. Jurevič E I (ed.). Dynamics of robot control. Nauka. Moscow: 1984. (in Russian)
12. La Hera P X, Morales D O. Non-linear dynamics modelling description for simulating the behavior of forestry cranes. International Journal of Modelling Identification and Control 2014; 21(2): 125-138, https://doi.org/10.1504/IJMIC.2014.060006.
13. La Hera P, Morales D O. Model-based development of control systems for forestry cranes. Journal of Control and Science and Engineering 2015; ID 256951, https://doi.org/10.1155/2015/256951.
14. Legnani G, Casalo F, Righettini P, Zappa B. A homogeneous matrix approach to 3D kinematics and dynamics – I. Theory, Mechanism and Machine Theory 1996; 31(5): 586-605, https://doi.org/10.1016/0094-114X(95)00100-D.
15. Legnani G, Casalo F, Righettini P, Zappa B. A homogeneous matrix approach to 3D kinematics and dynamics – II. Applications to chains of rigid bodies and serial manipulators. Mechanism and Machine Theory 1996; 31(5): 573-587, https://doi.org/10.1016/0094-114X(95)00100-D.
16. Mansfield N J. Human response to vibration. CRC Press LLC, Boca Raton, Florida, 2004, https://doi.org/10.1201/b12481.
17. Marques F, Flores P, Pimenta Claro J C, Lankarani H M. A survey and comparison of several friction force for dynamic analysis of multibody mechanical systems. Nonlinear Dynamics 2016; 86(3): 1407-1443, https://doi.org/10.1007/s11071-016-2999-3.
18. Morales D O, Westerberg S, La Hera P X, Mettin U, Freidovich L, Shiriaev A S. Increasing the level of automation in the forestry logging process with crane trajectory planning and control. Journal of Field Robotics 2014; 31(3): 343-363, https://doi.org/10.1002/rob.21496.
19. Papadopoulos E, Sarkar S. On the dynamic modeling of an articulated electrohydraulic forestry machine. in: Proceedings of the 1996 AIAA Forum on Advanced Developments in Space Robotics, WI, 1-2 August, 1996.
20. Papadopoulos E, Frenette R, Mu B, Gonthier Y. On the modeling and control of an experimental harvester machine manipulator. in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, Grenoble, France, 8-12 September, 1997,https://doi.org/10.1109/IROS.1997.656611.
21. Pennestri E, Rossi V, Salvini P, Valentini P P. Review and comparison of dry friction force models. Nonlinear Dynamics 2016; 83(4): 1785-1801, https://doi.org/10.1007/s11071-015-2485-3.
22. Posiadała B. Influence of crane support system on motion of the lifted load. Mechanism and Machine Theory 1997; 32(1): 9-20, https://doi.org/10.1016/0094-114X(96)00044-4.
23. Posiadała B. Modeling and analysis of the dynamics of load carrying system. in: Proc. of World Congress on Engineering and Computer Science, San Francisco, USA, 2012.
24. Posiadała B, Waryś P, Cekus D, Tomala M. The dynamics of the forest crane during the load carrying. International Journal of Structural Stability and Dynamics 2013; 13 (7), https://doi.org/10.1142/S0219455413400130.
25. Stribeck R. Die wesentlichen Eigenschaften der Gleit- und Rollenlager, Zeitschrift des Vereines Deutscher Ingenieure 2013; 46(36).
26. Urbaś A. Analysis of flexibility of the support and its influence on dynamics of the grab crane. Latin American Journal of Solids and Structures 2013; 10(1): 109-121, https://doi.org/10.1590/S1679-78252013000100011.
27. Urbaś A. Application of the Dahl friction model in the dynamics analysis of grab cranes, MATEC Web of Conferences 8347, 03008. doi: 10.1051/matecconf/201 CSNDD 2016 68303008, 2016.
28. Urbaś A. Computational implementation of the rigid finite element method in the statics and dynamics analysis of forest cranes. Applied Mathematical Modelling 2017; 46: 750-762, https://doi.org/10.1016/j.apm.2016.08.006.
29. Urbaś A, Harlecki A. Application of the rigid finite element method and the LuGre friction model in the dynamic analysis of the grab crane. in: Proceedings of 4th Joint International Conference on Multibody System Dynamics, Montreal, Canada, May 29-June 2, 2016.
30. Urbaś A, Szczotka M. Modelling friction phenomena in the dynamics analysis of forest cranes. Engineering Transactions 2016; 64(4): 393-400.
31. Szczotka M. Simulation and optimisation of the steering kickback performance. Journal of Theoretical and Applied Mechanics 2011; 49(1): 187-208.
32. Wittbrodt E, Szczotka M, Maczyński A, Wojciech S. Rigid Finite Element Method in Analysis of Dynamics of Offshore Structures. Ocean Engineering & Oceanography. Springer. Berlin-Heidelberg: 2013.

Andrzej URBAŚ
Department of Mechanical Engineering Fundamentals
University of Bielsko-Biala
Willowa 2, 43-309 Bielsko-Biala, Poland

Marek SZCZOTKA
Department of Transport
University of Bielsko-Biala
Willowa 2, 43-309 Bielsko-Biala, Poland

E-mail: aurbas@ath.bielsko.pl, mszczotka@ath.bielsko.pl

Article citation info:
CHUDZIK A, WARDA B. Effect of radial internal clearance on the fatigue life of the radial cylindrical roller bearing. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 211–219, http://dx.doi.org/10.17531/ein.2019.2.4.

Agnieszka CHUDZIK
Bogdan WARDA

EFFECT OF RADIAL INTERNAL CLEARANCE ON THE FATIGUE LIFE OF THE RADIAL CYLINDRICAL ROLLER BEARING

WPŁYW WEWNĘTRZNEGO LUZU PROMIENIOWEGO NA TRWAŁOŚĆ ZMĘCZENIOWĄ PROMIENIOWEGO ŁOŻYSKA WALCOWEGO*

The paper presents result of the research on influence of internal radial clearance in radial cylindrical roller bearing on its fatigue durability. By solving the Boussinesq problem for the elastic half-space and finite elements method, stress distributions were determined, necessary to estimate predicted fatigue life of the bearing. The calculations took into account geometrical parameters of the bearing: its radial clearance and shape of rolling parts. Predicted radical clearance was computed by using the Lundberg and Palmgren model. ANSYS program allowed to introduce the analysis of von Mises stress distribution in any plane of cooperating components. The outcome revealed, radial cylindrical roller bearing will have highest endurance with slight radial clearance.

Keywords**:** *rolling bearing; stress distribution; fatigue life; finite element method; radial clearance.*

W pracy przedstawiono wyniki badań wpływu wewnętrznego luzu promieniowego w promieniowym łożysku walcowym na jego prognozowaną trwałość zmęczeniową. Wykorzystując zagadnienie Boussinesqa dla półprzestrzeni sprężystej i metodę elementów skończonych, określono rozkłady naprężeń podpowierzchniowych niezbędne do oszacowania prognozowanej trwałości zmęczeniowej łożyska. W obliczeniach uwzględniono geometryczne parametry łożyska: jego luz promieniowy i kształt części tocznych. Do określenia wartości luzu promieniowego wykorzystano model Lundberga i Palmgrena. Zastosowanie programu ANSYS pozwoliło na przeprowadzenie analizy rozkładu naprężeń von Misesa w dowolnej płaszczyźnie współpracujących ze sobą elementów tocznych łożyska. Analiza obliczeń wykazała, że promieniowe łożysko walcowe będzie miało najwyższą wytrzymałość z niewielkim luzem promieniowym.

Słowa kluczowe*: łożysko walcowe, rozkład naprężeń, trwałość zmęczeniowa, metoda elementów skończonych, luz promieniowy****.***

Nomenclature

A	material constant
B	bearing width
C	basic dynamic load rating
D	bearing outside diameter
D_r	roller diameter
E	Young's modulus
F_r	radial load of the bearing
INT	integer function
L	fatigue life
L_r	roller length
Q_r	resultant normal force in the roller-main race contact
S_t	Stribeck's constant
U_z	local displacement inside the bearing elements
Z	depth of occurrence of maximal von Mises stresses along the *x* axis
Z_r	number of rollers in the bearing
b	half width of the contact
c	exponent in the equation determining the survival probability
d	bearing bore diameter
d_{bi}	diameter of the inner ring raceway
d_{bo}	diameter of the outer ring raceway
e	Weibull slope
g	radial clearance in the bearing
h	exponent in the equation determining the survival probability
i	inner raceway
j	number of roller
l	length of the roller-main race contact area
o	outer raceway
n	number of pairs of loaded items
r_b	radius of the main race
r_c	roller chamfer
u	number of load cycles per one revolution
δ_{max}	maximum elastic deformation
δ_c	total displacement of the bearing axis
ν	Poisson's constant
σ	maximal von Mises stress occurring along the *x* axis

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

ϕ survival probability of the bearing
ψ angle measured along the bearing circumference
ψ_j angle between rollers
ψ_ε angle of the loaded zone

1. Introduction

Researches of rolling bearing and forecasting their fatigue life were carried out over many years. The deciding factors, having direct influence on approximations of fatigue life are the phenomena occurring between roller elements and raceways. Analytical computational methods of fatigue life does not allow to consider many factors, e.g.:

- bearings radial clearance;
- shape of rolling element generators;
- bearing ring misalignment;
- axial load of radial cylindrical roller bearing.

Moreover analytical computational methods does not allow to predict result of inaccuracy of rolling bearing elements production, insufficient rigidness of development and deflection of bearing roller. FEM become commonly used tool in many domains of science and technology. Applying it allows to introduce parameters, which in analytical computations were unavailable until now. The goal of the numerical and analytical analysis was to research the phenomena occurring in rolling elements single interface and bearing's raceways. Due to the complexity of the researched model, to numerical computations author used solid model of bearing section from half of a bearing and relevant fragments of both rings.

Contact surface of cooperating elements with non-flat contact areas of all devices is very small. With relatively small loading forces, considerable surface pressures are generated causing subsurface stresses. Subsurface stresses have a significant effect on the durability of structural elements. This kind of problem occurs in gearing and all kinds of rolling bearings. H. Hertz [7] initialled the research on the contact surface under pressure. He presented general contact solution of a two elastic bodies to expose radial load. This paper was the basis for Lundberg's research [14], who provided an analytical solution of the model in which a cylinder is compressed between two planes. In 1907, Stribeck [23] showed that the bearing's external load is carried by a part of the rolling element and the load on the individual rolling element depends on their position and the bearing's internal clearance. The greater the clearance between rolling elements, the smaller the number of rollers that carry the load and the greater the force will act on the roller lying on the line of action of external force. Stribeck [23] stated, that the value of the force acting on the roller in case of zero internal clearance bearings is 4.37 of the mean roller bearing load . He suggested, that the value of this force was 5 times the average force, for a bearing with non-zero internal load. Palmgren [17] proposed that for a roller bearing with zero inner clearance, this value should be 4.08. For the other bearings with typical internal clearance, including cylindrical roller bearings, it should be 5. This constant was called the Stribeck's constant.

Rolling bearings are commonly used in engineering designs. These are elements which align the axis of the shaft that carry both radial and axial loads. Rolling bearing is required to sustain long working hours and have high durability. For many years, rolling bearings have been the object of interest for researchers who try to determine rolling bearing's working time up to their failure for various working conditions. That is, to determine the time of failure-free operation of a bearing. Lundberg and Palmgren [15, 16] initiated studies of the effect of internal load on durability of rolling bearings. There were also comprehensive research of different types of surface contact [9], including proposals for calculating the load distribution for rolling elements in various kinds of bearings [2, 3, 4, 5, 6, 8]. The simulations carried out on numerical models of bearings are a convenient and affordable tool for carrying out such tests. For this purpose, numerical models of bearings are built, loaded with external forces (transverse, longitudinal or their combination). By examining the model of the slewing bearing, Kania [10] presented numerical calculations characteristics: deformation - load.

Ricci [18] carried out numerical calculations of distribution of the load among elements of roller bearing. He presented the results of changes in the geometry of the bearing under the specified load. In 2011 Tang et al. [24] developed a numerical roller bearing model using the ANSYS software package. Contact pressure values and maximum contact stresses of elements were calculated. In the same year Laniado-Jacome [13] carried out the research to determine a distribution of the external load on the rolling elements of the rolling bearings loaded with radial force at variable rotational speed. The analysis was performed on a numerical and an analytical model. In Shah Maulik et al. [19] conducted research on the effect of modifying a roller profile on a pressure distribution in contact area. The cylindrical roller bearing durability were investigated using the ANSYS package. Shingala et al. [21] carried out numerical analysis of the contact of rolling elements in needle bearings used in a gearbox. The research which expanded numerical studies on the influence of flows on rolling elements were presented by Deshapande et al. [1] and Shaha Rohit [20].

One of the most widely used programs for numerical calculations is ANSYS. Further work indicates that research carried out using this package is constantly expanding and numerical models are refined. The presented work is an attempt to approximate the numerical model to the geometry and load conditions of actual bearing. The main objective of the research was to assess the effect of radial clearance of roller bearings on phenomena occurring in the contact area rolling elements with raceways. The article presents a methodology of conduct in the determination of subsurface stress necessary to calculate fatigue life by a numerical method.

2. Distribution of external loads

Cylindrical roller bearings are an important element in manufacturing mechanical equipment and are widely used in all sorts of fields. For most applications, cylindrical roller bearings insure a radial load. Cylindrical roller bearing of N or NU type will allow free movement in the axial direction of the shaft in relation to the casing. This allows to overcome the differences in thermal expansion of the materials used in the structure. The NJ type bearings allow to move light axial shaft loads in one direction. Cylindrical roller bearings have a relatively high radial load limits and can operate at high rotational speeds. Their permissible speeds exceed the speed of spherical roller bearings and taper bearings. Rollers in roller bearings are rarely ideal cylinders. In order to reduce stress accumulation at the ends of the roller caused by possible displacement of a shaft and a housing, their ends are rounded and cylinder shapes are corrected. Misalignment of the edge of a bearing, deformation of a shaft, inclination of inner or outer ring may cause tilting the rollers against the rings.

2.1. Distribution of loads in radially-loaded cylindrical roller bearings with radial clearance

The roller bearing consists of two rings (inner and outer) and from a few to several rollers placed between them depending on the size of the bearing. Their even distribution is ensured by the basket. The design of the outer ring of NU type bearing or the inner ring of N type bearing prevents the rollers from moving in the axial direction. In the NJ type bearings, the inner race design allows one-sided displacement

to be achieved with respect to the rollers. Cylindrical roller bearings are made with different radial clearances from C1 to C5. C1 bearings have the smallest radial clearance and C5 the largest. Most of the bearings are produced with normal radial clearance CN. Regardless of that, by selecting the appropriate fit, an initial load in the bearing can be obtained. Then the bearing operates with radial clearance less than zero ($g < 0$). Stribeck [3] demonstrated that the load distribution on rolling bearings with radial force is uneven. Load bearing area for the bearing with positive radial clearance is less than 180°. Figure 1 shows the unloaded bearing with $g = 0$, while Figure 2 shows the bearing loaded with radial force F_r.

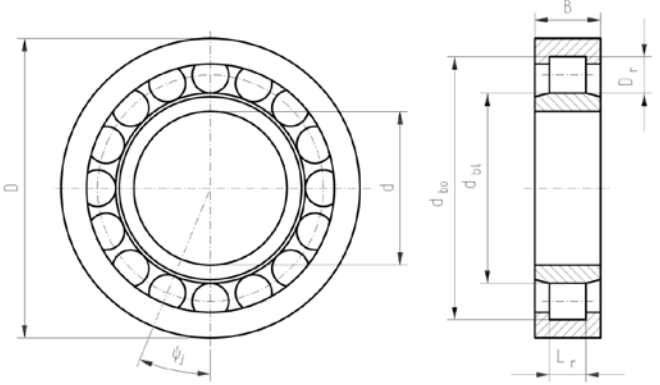


Fig. 1. Bearing with radial clearance g = 0 before loading with force F_r

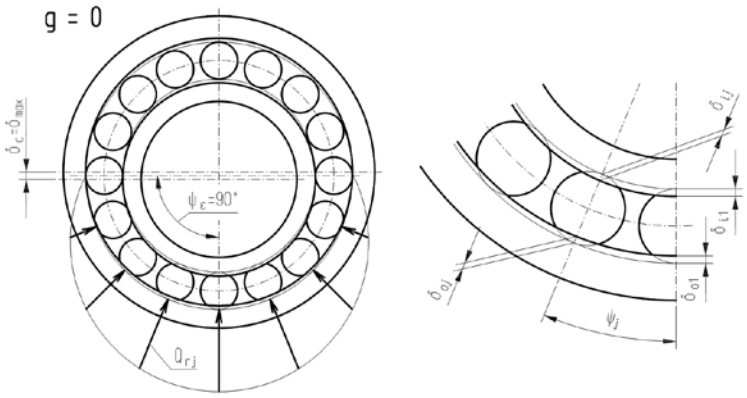


Fig. 2. Bearing with radial clearance g = 0 after loading with force F_r

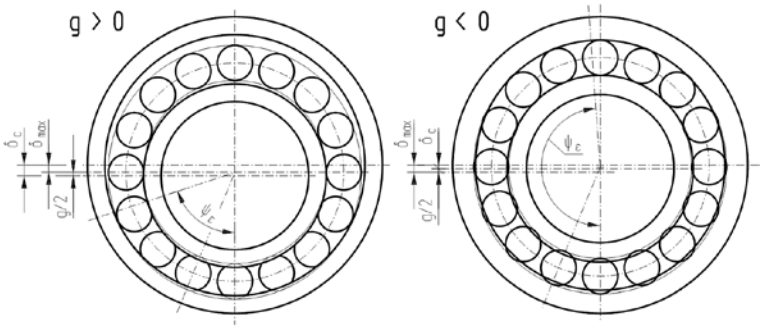


Fig. 3. Bearing with positive and negative radial clearance after loading with force F_r

The number of pairs of rolling elements carrying the load for $g = 0$ can be calculated from this formula:

$$n_{g=0} = INT\left(\frac{Z_r - 1}{4}\right). \quad (1)$$

The number of rolling elements carrying the load in the situations shown in Figure 2 is $2n_{g=0} + 1$. The angle between rolls is calculated from the formula:

$$\psi_j = \frac{2 \cdot \pi}{Z_r} \cdot \quad (2)$$

Figure 3 illustrates the loading bearing zone for positive ($g > 0$) and negative ($g < 0$) radial clearance. The angle of the loaded zone is given by:

$$\psi_\varepsilon = arccos\left(\frac{g}{2 \cdot \delta_c}\right), \quad (3)$$

$$\delta_c = \delta_{max} + g/2 \cdot \quad (4)$$

In the case when $g > 0$, the number of rollers subjected to loading is lower than in the case of $g = 0$. Whereas $g < 0$ it can reach the maximum value equal to the number of rollers in the bearing Z_r.

The equilibrium equation for the bearing loaded with radial force F_r has the form:

$$F_r = Q_{r1} + 2Q_{r2}cos\psi_2 + \ldots + 2Q_{rj}cos\psi_j + \ldots + 2Q_{rn}cos\psi_n \cdot \quad (5)$$

The roller No.1 subjected to the load Q_{r1} cause deformation δ_{max}, that is the sum of contact deformations between the roller and inner ring δ_{i1} and between the roller and outer ring δ_{o1}:

$$\delta_{max} = \delta_{i1} + \delta_{o1} \cdot \quad (6)$$

The value of deformation of the other rollers can be calculated from the formula:

$$\delta_{rj} = \delta_c \cdot cos\psi_j - g/2, \quad (7)$$

where:

$$\delta_{rj} = \delta_{ij} + \delta_{oj} \cdot \quad (8)$$

Knowing the magnitude of deformation, the resultant force Q_{rj} at the contact between the roller and the raceway can be determined according to the Palmgren formula [15, 16]:

$$Q_{rj} = 78000\delta_{i,oj}^{10/9} \cdot l^{8/9}, \quad (9)$$

where the length of the roller-main race contact area is as follows:

$$l = L_r - 2r_c \cdot \quad (10)$$

Equation (5) has been solved numerically using the ROLL1 computer program, built on the basis of the methodology described in the works [25, 26]. The results obtained enable determination of the Stribeck's constant:

$$S_t = Q_{r1} \cdot Z_r / F_r \cdot \quad (11)$$

3. Fatigue life prediction of the radial cylindrical roller bearing

The amount of radial clearance in the radial roller bearing determines the distribution of loads on the rolling elements, and thus has a significant impact on its fatigue life. To determine the fatigue life of the bearing, the methodology described in [25] was applied.

The survival probability of the stationary bearing ring takes the form:

$$ln\frac{1}{\varphi_o} = A \cdot u_o^e L_o^e \int_0^{2\pi} \int_0^l r_{box} \sigma_{oxy}^c Z_{oxy}^{1-h} dx d\psi \ . \quad (12)$$

The survival probability of the bearing ring rotating relative to the load is as follows:

$$ln\frac{1}{\varphi_i} = A \cdot u_i^e L_i^e 2\pi \left[\frac{1}{2\pi} \int_0^{2\pi} \left(\int_0^l r_{bix} \sigma_{ixy}^c Z_{ixy}^{1-h} dx \right)^{1/e} d\psi \right]^e . \quad (13)$$

In the above formulas L is the number of revolutions, u the number of load cycles per one revolution, $\sigma_{x\psi}$ is the maximal von Mises subsurface stresses and $Z_{x\psi}$ is the depth at which these stresses occur (Fig. 4). A is a material constant, which is equal to $4.5 \cdot 10^{-40}$ for the bearing being the object of consideration. The values of the exponents occurring in Eq. (12, 13) are assumed: $c = 31/3$, $h = 7/3$, $e = 9/8$.

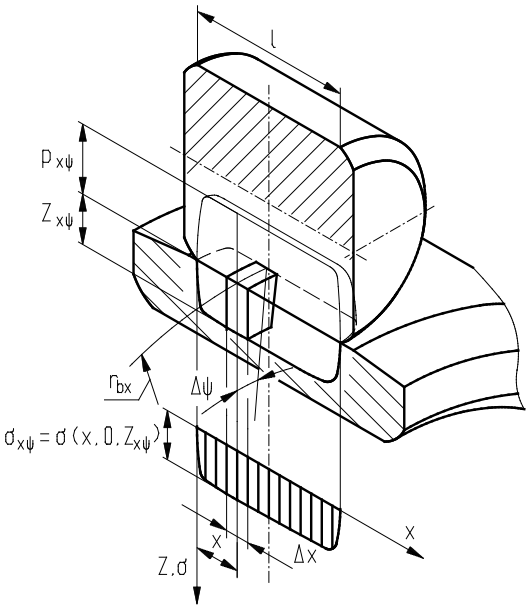


Fig. 4. Distributions of pressure and subsurface stresses deciding on fatigue life of radial cylindrical roller bearing

Using the formula (14), the fatigue life of the tested bearing was determined.

$$L = \left(L_o^{-e} + L_i^{-e} \right)^{-1/e} . \quad (14)$$

For the survival probability $\varphi_o = \varphi_i = 0.9$, the computed value L is identical with the basic rating life L_{10}. The L value is representing number of rotations the bearing can perform before it wears out.

The equations (12) and (13) were resolved by using a program ROLL2 [25]. Determining maximum values substitute subsurface stresses σ in the contact zone between rollers and raceways and estimating the depth Z of their occurrence was necessary to perform the calculations of the fatigue life.

Two methods were used to obtain the distraction of pressure and subsurface stresses:

- "half space" method using the Boussinesq solution describing the deformation of the elastic half-space subjected to pressure – ROLL4 computer program [25, 26];
- finite element method (FEM).

The FEM solid numerical model was developed using the ANSYS – Augmented Lagrange algorithm package. Using the conditions of symmetry, the numerical calculation was done for the half of a roller with raceways, which is shown in Figure 5. The numerical models under investigation were divided into 8-node solid elements of the SOLID185 type and elements which were in contact zone – CONTA174 and TARGE170. Additionally, the symmetry conditions and the degrees of freedom following from real working conditions were considered in the calculations. In order to increase the computational accuracy, a division into elements was increased in the area of the predicted contact and a non-uniform division of contact elements in this area was also applied. The distance between the mesh nodes of the contact area elements of the roller is 0.05 mm (with bearing diameter 15 mm). The distances between the node of the remaining raceways and the roller are adjusted at an angle eliminating component shape errors. As a result, the numerical model contains 200363 nodes in 88910 elements. Considering the character of the contact of a roller bearing with the raceways, the roller was treated as a contact surface. Raceways surfaces were assumed as target surfaces. The displacement towards every axis of the coordinate system for the outer surface of the ring was blocked. The inner ring was allowed to move only in the direction of the "z" axis. The coefficient of stiffness of FKN = 1.5 has been adopted. Using the symmetry conditions, the model was loaded with force $F_r / 2$.

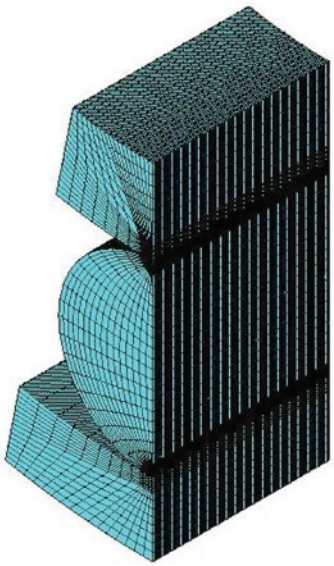


Fig. 5. FEM numerical model grid of the roller-raceways contact

4. Subject of study

The subject of study was the NU 213 ECP cylindrical roller bearing [22]. Parameters of the tested bearing are shown in Table 1.

The rollers were corrected according to the modified logarithmic correction proposed by Krzemiński-Freda [11]. The generator profile was the same as described in the paper [26] (Fig. 6). Parameters of the logarithmic correction profile of NJ 213 ECP bearing are not published by the manufacturer.

Table 1. Parameters of the NU 213 ECP cylindrical roller bearing [22]

Bearing bore diameter	d = 65 mm
Bearing outside diameter	D = 120 mm
Bearing width	B = 23 mm
Diameter of the outer ring raceway	d_{bo} = 108.5 mm
Diameter of the inner ring raceway	d_{bi} = 78.5 mm
Roller diameter	D_r = 15 mm
Roller length	L_r = 15 mm
Roller chamfer	r_c = 0.5 mm
Number of rollers in the bearing	Z_r = 16
Dynamic load rating	C = 122000 N

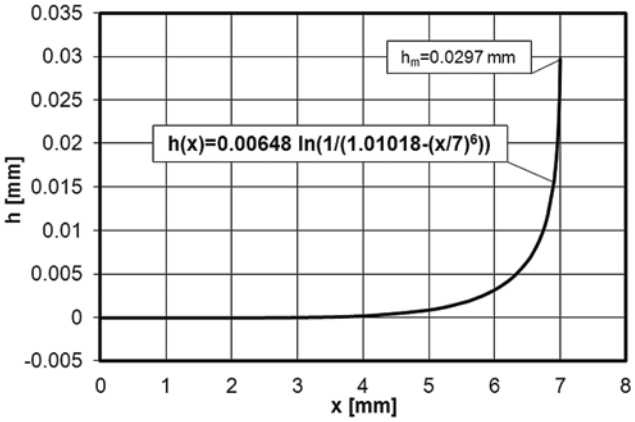


Fig. 6. Roller generator profile with a modified logarithmic correction [26]

5. Results

5.1. Rollers load

According to Krzemiński-Freda [12] cylindrical roller bearings loaded with radial force $F_r = 0.3\ C$ operate under hard operating conditions. Normal operating conditions are considered to be such that the load is: $F_r = (0.07–0.12)\ C$. The researched bearing was loaded with the force: $F_r = 0.3\ C = 36600$ N and $F_r = 0.1\ C = 12200$ N. Table 2 shows load distributions calculated with ROLL1 computer program for selected radial clearances. Figure 7 illustrates the distributions of radial load Q_r on the rollers for $F_r = 0.3\ C$.

Table 2. Radial load distributions on NU 213 ECP bearing rollers

		F_r = 36600 N				
g [mm]		-0.038	-0.019	0.0	0.019	0.038
ψ_ε [°]		180.0	113.36	90.00	78.49	71.32
Roller No.	ψ_j [°]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]
1	0	9447	8672	9342	10015	10673
2	22.5	9069	8148	8556	8962	9353
3	45	8001	6676	6357	6036	5706
4	67.5	6429	4535	3213	1944	752
5	90	4624	2141	0	0	0
6	112.5	2888	51	0	0	0
7	135	1495	0	0	0	0
8	157.5	631	0	0	0	0
9	180	349	0	0	0	0
S_t		4.13	3.79	4.08	4.38	4.67
		F_r = 12200 N				
g [mm]		-0.019	-0.0095	0.0	0.019	0.038
ψ_ε [°]		180.00	128.45	90.00	67.71	58.07
Roller No.	ψ_j [°]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]
1	0	3721	2875	3114	3701	4067
2	22.5	3598	2725	2852	3201	3344
3	45	3250	2304	2119	1821	1381
4	67.5	2735	1688	1071	12	0
5	90	2141	991	0	0	0
6	112.5	1562	343	0	0	0
7	135	1088	0	0	0	0
8	157.5	782	0	0	0	0
9	180	678	0	0	0	0
S_t		4.88	3.77	4.08	4.85	5.33

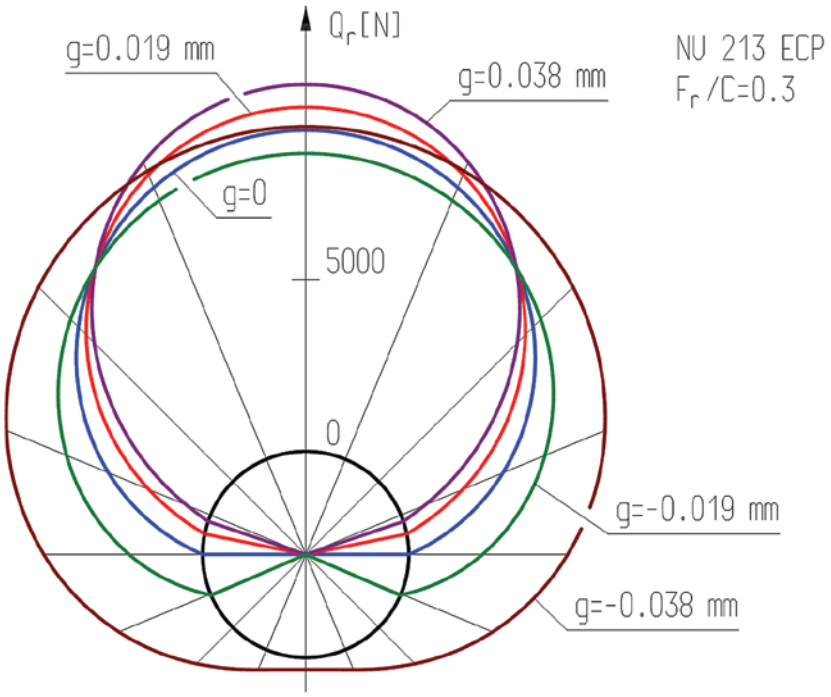


Fig. 7. The distributions of radial load on the rollers of NU 213 ECP bearing

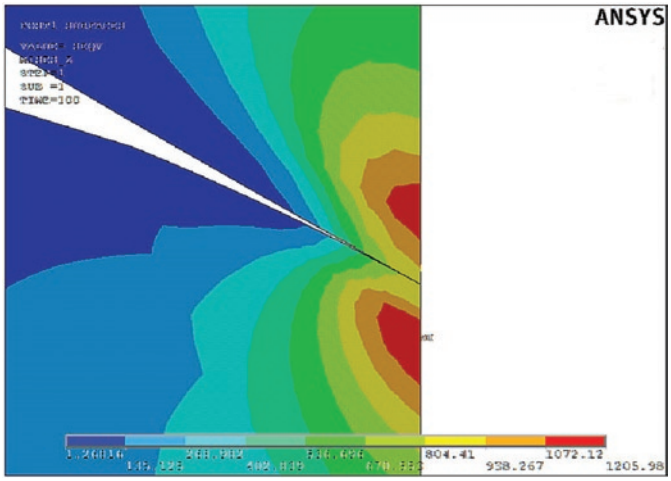


Fig. 8. Distributions of von Mises stresses below the roller-inner ring raceway contact surface

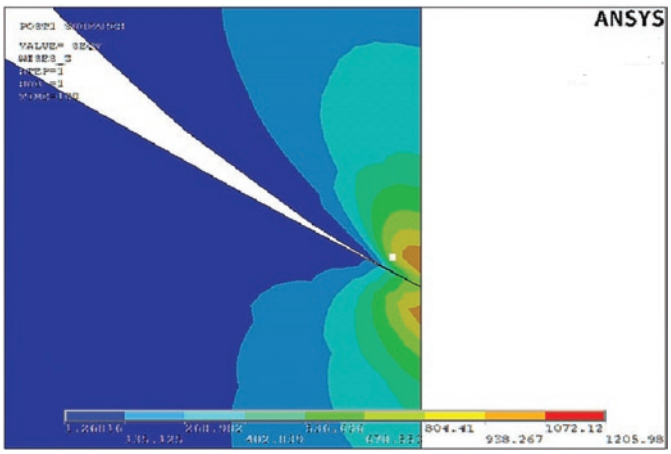


Fig. 9. Distributions of von Mises stresses below the roller-outer ring raceway contact surface

5.2. Surface pressure and von Mises stress distributions

As a result of numerical calculations, data necessary to analysis the phenomena occurring in the contact zone of the kinematic pair of the researched model were obtained. The obtained results of the bearing calculations NU 213 ECP are shown in the form of graphs and maps. The researched bearing is loaded with radial force F_r = 36600 N assuming that the axes of the rolling elements and the raceways of the bearing are parallel, the radial clearance is g = 0.038 mm. For the calculations of surface pressures and von Mises stresses with the "half space" method, it was assumed that bearing elements were made of elastic-ideally plastic material. The material properties were determined by Young's modulus E = 208 GPa, Poisson's constant ν = 0.3 and the tensile yield strength σ_o = 1950 MPa. The same assumptions as regards the material characteristics were made during the calculations of surface pressures and von Mises stresses with the FEM.

Figure 8 shows a graphical map of the distribution of maximum von Mises stress below the surface of the roller-inner ring raceway contact.

Figure 9 shows the distribution of maximum von Mises stress below the surface of the roller-outer race contact. Figures 10 and 11 show the contact surface of the roller with the inner and outer ring raceway with detail.

In the basic version, ANSYS gave only one maximum value and one minimum von Mises stress value and coordinates *(x, z)* of the place of occurrence of these stresses across the contact area. In this paper, the authors present the possibility of defining the von Mises stress distribution in a cross-section determined by any x-coordinate, which makes it possible to determine maximal values of stresses σ and the depth of their occurrence Z in the cross-section. The ANSYS package allows direct determination of the von Mises stress distribution in the plane selected by the user. Example of calculation results are presented in Figure 12, 13 and 14. The results obtained with this method enabled determination of the maximum von Mises stress dis-

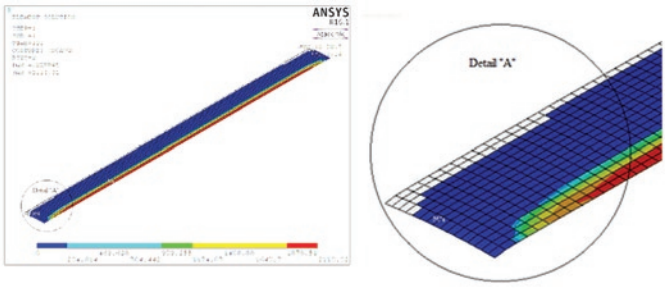


Fig. 10. Contact surface of the roller with the inner ring raceway

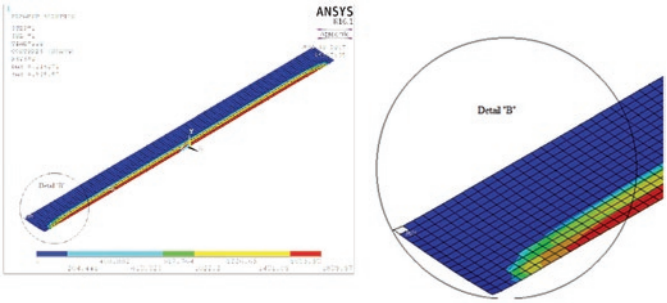


Fig. 11. Contact surface of the roller with the outer ring raceway

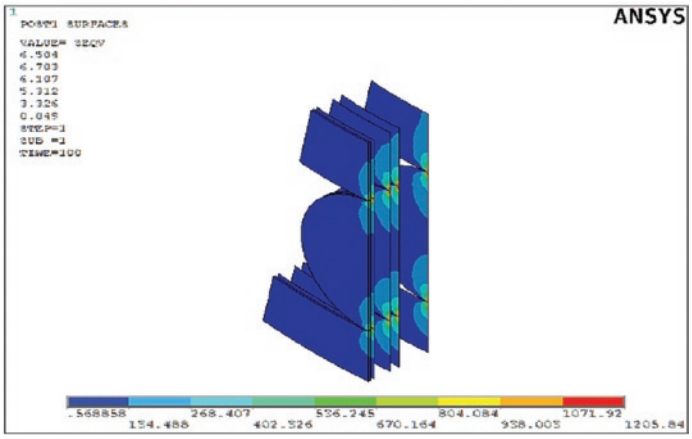


Fig. 12. The example values of von Mises stresses at the material depth

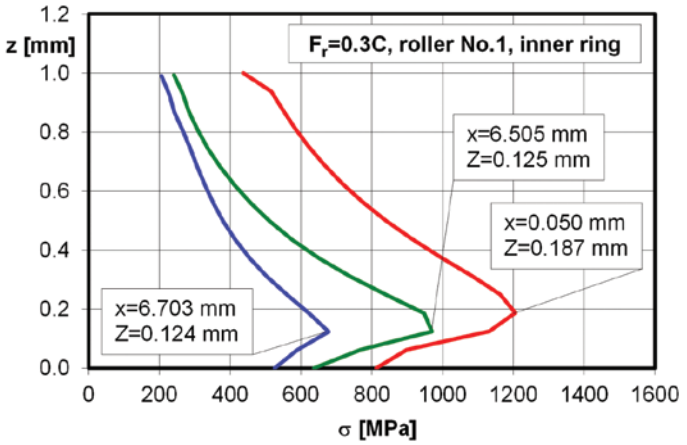


Fig. 13. Distribution of von Mises stresses in the three example cross-sections of the roller-inner ring raceway contact

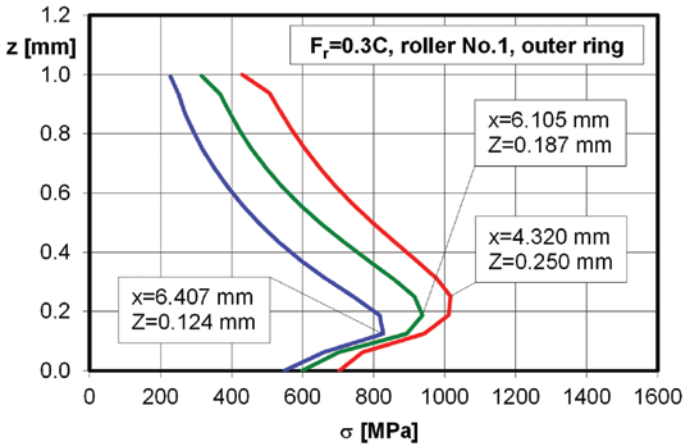


Fig. 14. Distribution of von Mises stresses in the three example cross-sections of the roller-outer ring raceway contact

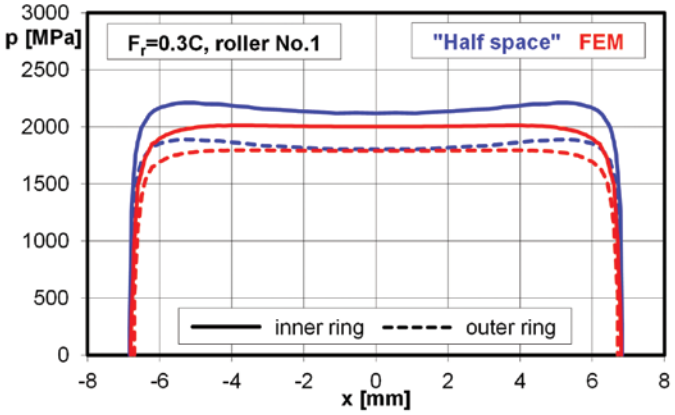


Fig. 15. Surface pressure distributions along the x axis of the most heavily loaded roller in contact with the inner and outer ring raceway

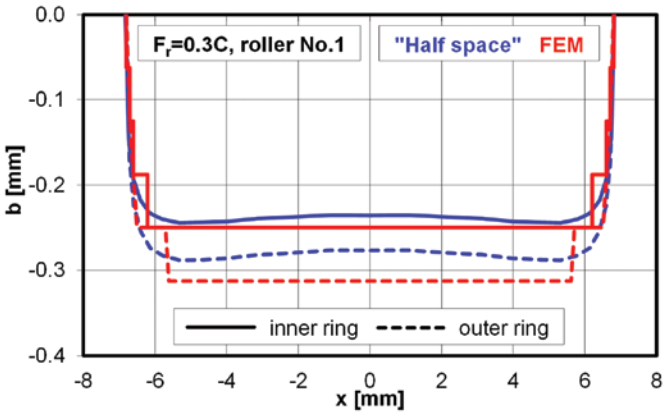


Fig. 16. Contact areas of a roller with raceways

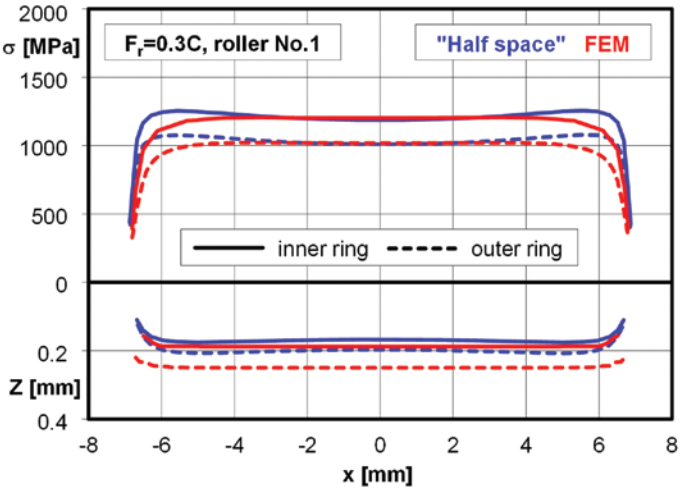


Fig. 17. Distributions of maximal von Mises stresses σ and the depth of their occurrence Z in the most heavily loaded roller in contact with the inner and outer ring raceway

tributions and the depth of their occurrence along the contact line. These distributions are necessary to determine the expected fatigue life of the bearing.

In Figure 15 a comparison of distributions of maximal surface stresses along the *x* axis of the most loaded roller (No. 1) in contact with the inner and outer ring raceway, obtained with ROLL4 (”half space” method) and with the FEM is presented. Figure 16 shows boundaries of the contact area of a roller with both raceways.

The results obtained using elastic “half space” solution and FEM are sufficiently convergent. The shape and size of the contact zones are similar. The only difference in the surface stresses distributions is only observed near the ends of the contact field. This is due to the small surface pressure exerted on this points and has a little effect on the estimations of fatigue life.

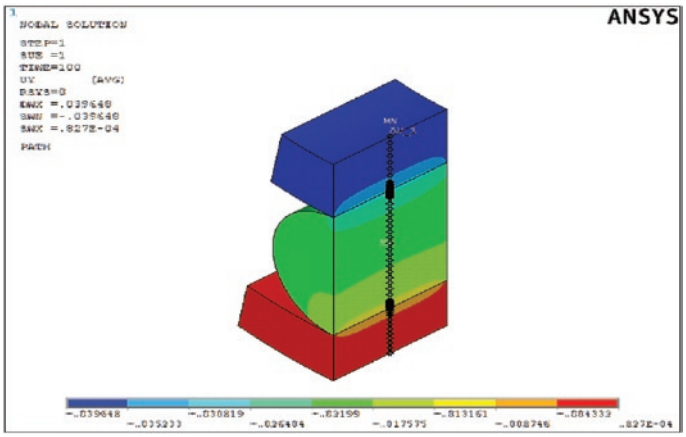


Fig. 18. Displacement of radial cylindrical roller bearing elements

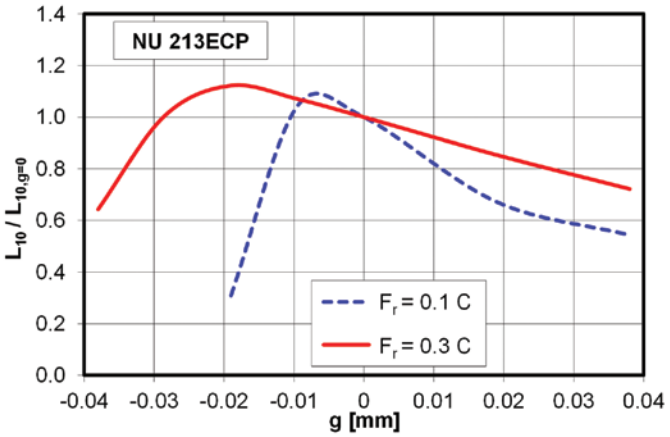


Fig. 20. Relative bearing life as a function of radial clearance

Figure 17 shows the stresses σ and the depths *Z* of the most heavily loaded roller in contact with the inner and outer ring raceway, obtained with both method.

The curves describing the depth at which the maximal von Mises stresses occur, determined by both the „half space" method and by FEM, have a shape similar to the von Mises stress distributions. The distributions of the depth *Z* determined with the FEM are characterized by slightly higher values if compared to the plots obtained with the "half space" method. These differences are due to the model used in the FEM which, in contrast to the "half space" method, simultaneously considers the contact of the roller with an inner and an outer ring.

The ANSYS program also allows to get a map of the displacements inside the contacting bodies. Figures 18 and 19 show the displacement of cooperating bearing elements due to the applied load. Figure 19 shows that the largest displacement is near the contact between the roller and raceways.

5.3. Fatigue life

Due to the long time necessary to solve the contact problem using FEM, only the distributions obtained using the "half space" method were used to calculate the expected fatigue life of the bearing.

In Figure 20, as a function of radial clearance, the characteristics of bearing fatigue life changes with respect to fatigue life for radial clearance equal to zero are presented. The course of the curves is a confirmation of the known fact that the bearing achieves the highest durability with a little radial interference [12].

In case of radial load $F_r = 0.3\ C$, the fatigue life of the tested bearing achieves the highest values (i.e. greater than the durability for $g = 0$) when radial clearances is between the range from $g = 0$ to

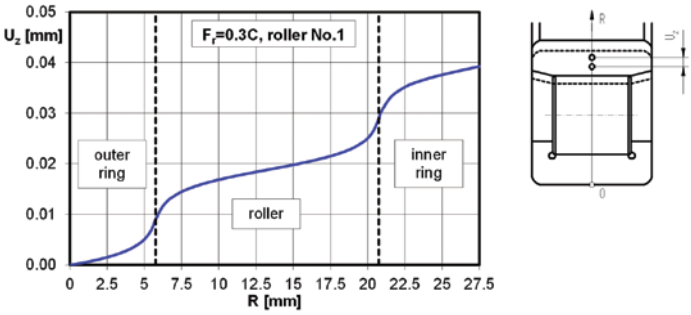


Fig. 19. Local displacement Uz inside the bearing elements

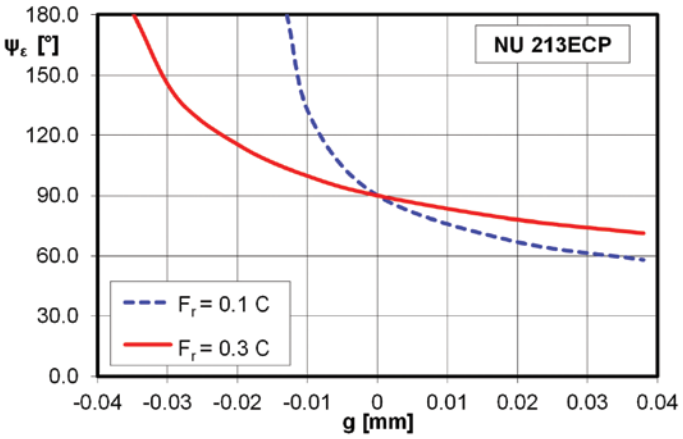


Fig. 21. Angle of the loaded zone as a function of radial clearance

$g = -0.029$ mm. In the case of radial load $F_r = 0.1\ C$, the highest fatigue life is achieved by the bearing with radial clearances in the range from $g = 0$ to $g = -0.011$ mm. In both cases, this corresponds to the angles of the loaded zone in the range of $90° \le \psi_\varepsilon \le 140°$ (Fig. 21).

The greatest increase in fatigue life in relation to fatigue life for zero radial clearance that can reach the bearing is equal to $L_{10} / L_{10,g=0} \approx 1.1$. This increase occurs at the angle $\psi_\varepsilon \approx 115°$, regardless of the radial load value. However, as the load increases, the amount of interference increases, at which the greatest increase in fatigue life is achieved. In case of radial clearances greater than zero and for interferences with values greater than the values given above, the fatigue life of the tested cylindrical roller bearing is less than the fatigue life for clearance $g = 0$. The reasons for this phenomenon are explained in Figure 7, presenting the load distributions for various radial clearances. As shown, in the range of clearance from $g = 0$ to $g = -0.029$ mm, the number of rollers subjected to the load is greater than in the case of radial clearance $g > 0$, but at the same time the values of forces acting on the most loaded rollers that determine the bearing fatigue life are lower. For radial clearances smaller than $g = -0.029$ mm and greater than $g = 0$ both the force values increase, which results in a decrease in fatigue life.

6. Conclusions

The applied methodology allows to predict the durability of any radial cylindrical roller bearing for a given load and radial clearance by introducing parameters into the calculations that could not be taken into consideration when doing the analytical computations. One of them is the correction of roller generators.

The methodology enabled to determine the distributions of radial load on the rolling elements at a predetermined radial clearance. It also allows to calculate the Stribeck's constant.

As a result of numerical calculations, the data necessary to analysis the phenomena occurring in the contact zone of the kinematic pair of the researched model were obtained. In case of testing of rolling bearings, it was necessary to build complex solid models of cooperating elements. Knowledge of subsurface stress distributions is necessary to determine the fatigue life of rolling bearings.

FEM allowed to estimate accurate results on the state of subsurface stresses. Similar outcome can be obtained by solving the Boussinesq problem for elastic half-space.

The ANSYS package allows to determine the von Mises stress distribution in the plane selected by the user. The possibility of defining the von Mises stress distribution in the section defined by any x coordinate is particularly important for predicting the fatigue life of radial cylindrical bearings subjected to a complex load. The complex load causes a significant tilting of the rolling elements of the bearing, which requires knowledge of the exact position of the maximum stress in the material.

The performed analyses showed that the radial cylindrical roller bearing achieves the highest durability with a little radial interference. The greatest increase in fatigue life in relation to fatigue life for zero radial clearance the bearing achieves for a radial interference at which the angle of the loaded zone is equal to $\psi_\varepsilon \approx 115°$, regardless of the radial load value. The increase in fatigue life in this case is equal to $L_{10} / L_{10,g=0} \approx 1.1$.

References

1. Deshpande H, Kulkarni S, Gandhare B S. Investigation on effect of defect on cylindrical roller bearing, by experimental and FEA approach. Iternational Journal of Emerging Technology and Advanced Engineering 2014; 4 (6).
2. Hamrock B J, Anderson W J. Rolling-element bearings. NASA RP_1105/REV1; 1983.
3. Harris T A. Rolling bearing analysis. Wiley-Interscience, New York, USA; 1991.
4. Harris T A, Kotzalas M N. Advanced concepts of bearing technology. Rolling bearing analysis. Fifth edition. CRC Press; 2006, https://doi.org/10.1201/9781420006582.
5. Harris T A, Kotzalas M N. Essential concepts of bearing technology. Fifth edition. CRC Press; 2006.
6. Harris T A, Kotzalas M N. Rolling Bearing Analysis. Fifth Edition - 2 Volume Set. CRC Press; 2007.
7. Hertz H. Uber die Beruhrung fester elastischer Korper. Gesammelte Werke (P. Lenard, ed.), Bd. 1, (J.A. Barth, Leipzig; 1895):155-173.
8. Jones A. A general theory for elastically constrained ball and roller bearing under arbitrary load and speedy conditions. Trans ASME 1960; 105: 591-595.
9. Jonson K L. Contact mechanics. Cambrige University Press, Cambrige; 1985, https://doi.org/10.1017/CBO9781139171731.
10. Kania L. Modelling of rollers in calculation of slewing bearing with the use finite elements. Mech Mach Theory 2006; 41: 1359-1376, https://doi.org/10.1016/j.mechmachtheory.2005.12.007.
11. Krzemiński-Freda H. Correction of the generators of the main working surfaces of roller bearings. Arch Mech Eng 1990; 37: 115–132.
12. Krzemiński-Freda H. Roller Bearing. PWN, Warsaw; 1985 (in polish).
13. Laniado-Jacome E. Numerical model to study of contact force in a cylindrical roller bearing with technical mechanical event simulation. J Mech Eng Autom 2011; 1: 1–7.
14. Lundberg G. Cylinder compressed between two plane bodies. SKF Reg. 4134 1949.
15. Lundberg G, Palmgren A. Dynamic capacity of rolling bearings. Acta Polytech Scand Mech Eng 1947; 1(3): 1-52.
16. Lundberg G, Palmgren A. Dynamic capacity of roller bearings. Acta Polytech Scand Mech Eng 1952; 2(4): 96-127.
17. Palmgren A. Ball and roller bearing engineering. Third ed., SKF Industries, Philadelphia, PA 1959.
18. Ricci M C. Internal loading distribution in statically loaded ball bearings subjected to a centric thrust load: alternative approach. World Academy of Science, Engineering and Technology 2010; 65: 641-649.
19. Shah Maulik J, Darji P H. Fatigue life improvement through reduction of edge pressure in cylindrical roller bearing using FE analysis. International Journal For Technological Research in Engineering 2014; 1(10): 1069-1074.
20. Shaha Rohit D, Kulkarni S S. Vibration analysis of deep groove ball bearing using finite element analysis. International Journal of Engineering Research and Applications 2015; 5(5): 44-50.
21. Shingala Niraj R, Sata Ankit V, Delvadiya Parth V, et al. Contact stress analysis of needle roller bearing used in synchromesh gear box. Trends in Machine Design 2018; 5(1): 5–20.
22. SKF General Catalogue; 2007.
23. Stribeck R. Ball bearings for various loads. Reports from the Central Laboratory for Scientific Technical Investigation, translation by K.W. Van Treuren et al., Trans ASME 1907; 29: 420-463.
24. Tang Zhaoping, Sun Jianping. The contact analysis for deep groove ball bearing based on Ansys. Procedia Engineering 2011; 23: 423-428, https://doi.org/10.1016/j.proeng.2011.11.2524.
25. Warda B, Chudzik A. Fatigue life prediction of the radial roller bearing with the correction of roller generators. International Journal of Mechanical Science 2014; 89: 299-310, https://doi.org/10.1016/j.ijmecsci.2014.09.015.
26. Warda B, Chudzik A. Effect of ring misalignment on the fatigue life of the radial cylindrical roller bearing. International Journal of Mechanical Science 2016; 111-112, https://doi.org/10.1016/j.ijmecsci.2016.03.019.

Agnieszka CHUDZIK
Department of Dynamics
Lodz University of Technology
Stefanowskiego 1/15, 90-537 Lodz, Poland

Bogdan WARDA
Department of Vehicles and Fundamentals of Machine Design
Lodz University of Technology
Stefanowskiego 1/15, 90-537 Lodz, Poland

E-mails: Agnieszka.Chudzik@p.lodz.pl, Bogdan.Warda@p.lodz.pl

Article citation info:
MARCZUK A, MISZTAL W, SAVINYKH P, TURUBANOV N, ISUPOV A, ZYRYANOV D. Improving efficiency of horizontal ribbon mixer by optimizing its constructional and operational parameters. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 220–225, http://dx.doi.org/10.17531/ein.2019.2.5.

Andrzej MARCZUK
Wojciech MISZTAL
Petr SAVINYKH
Nikolay TURUBANOV
Alexey ISUPOV
Dmitry ZYRYANOV

IMPROVING EFFICIENCY OF HORIZONTAL RIBBON MIXER BY OPTIMIZING ITS CONSTRUCTIONAL AND OPERATIONAL PARAMETERS

POPRAWA EFEKTYWNOŚCI PRACY POZIOMEGO MIESZALNIKA WSTĘGOWEGO POPRZEZ OPTYMALIZACJĘ PARAMETRÓW KONSTRUKCYJNYCH I EKSPLOATACYJNYCH

The homogeneity of mixing various feed mixture ingredients and the reliability of mixers directly depend on the constructional features of mixing devices and operational parameters of their work. This article presents a theoretical underpinning for the structural (design) specifications of a horizontal ribbon auger mixer (blender) with regard to the fulfilment of the condition when the material is transported by auger tapes in different directions, ensuring the efficient mixing of the feed mixture ingredients and timely material unloading from the mixer tank (hopper). The performed optimization of the structural and operational specifications as a result of experimental studies has allowed to increase the reliability of the mixer and to determine the homogeneity coefficient of the finished product, which reaches its the highest value of 94.13% with the forward direction of rotation of the mixer movable operating elements, when mixing time is 13 minutes and when the tank load is 52.4% of its volume.

***Keywords**: mixer reliability, ribbon auger, mixing parameters, optimization.*

Jednorodność mieszania różnych składników mieszanki paszowej i niezawodność mieszalników zależy bezpośrednio od cech konstrukcyjnych urządzeń mieszających i parametrów roboczych ich pracy. W artykule przedstawiono teoretyczne podstawy dotyczące parametrów konstrukcyjnych (projektowych) poziomego mieszalnika ślimakowego (mieszarki) wywierających wpływ na warunki przemieszczania (podczas transportu w różnych kierunkach) zapewniających efektywne mieszanie składników mieszanki paszowej i terminowy rozładunek materiału ze zbiornika mieszalnika (zasobnika). Przeprowadzona na bazie badań eksperymentalnych optymalizacja parametrów konstrukcyjnych i eksploatacyjnych pozwoliła na zwiększenie niezawodności mieszalnika i wyznaczenie współczynnika homogeniczności gotowego produktu, który osiąga najwyższą wartość 94,13% podczas przemieszczania przy normalnym kierunku obrotów ruchomych elementów mieszalnika, z czasem mieszania wynoszącym 13 minut, gdy ładunek stanowi 52,4% objętości zbiornika.

***Słowa kluczowe**: niezawodność mieszalnika, przenośnik ślimakowy, parametry mieszania, optymalizacja.*

1. Introduction

To increase the productivity of farm livestock (animals and poultry), it is necessary to provide them the right nutrients in the right doses [5, 7, 8, 23]. Achieving this is possible only with the use of mixed feed [1, 3, 5, 6, 11, 13, 25]. Mixing is an operation that is particularly important. It allows combining many components with different properties into one product that meets the assumed requirements [4, 5, 10, 20]. The effectiveness of this operation affects the success of many production processes carried out in enterprises belonging to various industries [2, 4, 10, 12, 13, 20, 26].

The aim and indicator of the effectiveness of mixing is the homogeneity of the final product [5]. However, due to the fact that mixing is a process in which components actually go to random places with random probability, there are some problems with obtaining it [10]. Among the causes of these problems are indicated, among others, differences in mixed fractions (eg shape and size of particles, density, etc.). In the case of feed, the cause of additional problems is the addition of drugs and vitamins [6].

High market requirements force enterprises to streamline their production processes in order to achieve a shorter production cycle time, while ensuring an adequate level of quality of manufactured products and reduction of general costs as well as energy expenses [10, 11].

The uniformity of mixing of various components of the feedstuff, mixing speed and the reliability of mixers directly depends on the constructional (design) features of mixing devices and operational parameters of their work [6, 7, 8, 10, 13, 23, 25].

To date, a large number of mixer constructions have been developed [2, 4, 10, 12, 13, 20]. A number of studies were also conducted to determine the impact of various factors (including the way of providing components) and the construction of mixers on the homogeneity and speed of mixing [1, 2, 3, 4, 10, 16, 26]. Attempts have also been made to achieve uniformity through dilution, or ensuring adequate humidity, the proper temperature and proportion of ingredients [6].

But despite this, the processes occurring in them, which affect the reliability of operation and performance indicators when mixing components, are insufficiently studied [1, 2, 4, 13, 15, 20, 25].

Today the organization of an effective process of mixing components, which depends on the operational performance indicators of the mixers, leading to the production of mixtures corresponding to the zootechnical requirements, is an urgent task [25].

From the zootechnical perspective (considering the safety and correctness of nutrition), it is important not only to introduce the ingredients provided by the ration into the feed mixture composition at the required ratio, but it is also necessary to have them all evenly distributed throughout the entire volume of the mixture. The mixture homogeneity provides the uniform nutritional value of the feed in all parts of its volume [1, 3, 5, 7, 8, 10, 11, 12, 16, 20, 23, 26].

2. Study goals and objectives

The purpose of theoretical and experimental research is to increase the reliability of a horizontal ribbon mixer due to optimization of its constructional and operational parameters by determining the values of the most significant factors affecting the quality of mixing, reliability of equipment and energy intensity of the working process.

3. Techniques

Studies were performed in the livestock breeding mechanization laboratory. The research process have been carried out using monitoring and metering systems cooperating with the computer (in accordance with All-Union States Standards - GOST 15.101-98) [19].

During the research process, a material was used in which the basic factor was a mixture of barley (80%) and rye (20%), the density of which was 742 kg/m^3, and the control factor was pea with a density of 812 kg/m^3 [19]. The sampling procedure was consistent with GOST R ISO 6497-2011 [9] and suited to the procedures implemented in previous research [12, 19, 24].

During the evaluation of the quality of operational efficiency of mixer, homogeneity (understood as a state in which the content of components in any part of its volume corresponds to a given blend composition) was assumed as the main criterion of the final product [19, 21, 22, 24].

Theoretical substantiation of constructional parameters of a horizontal ribbon mixer

Final preparation of feed mixtures is carried out in mixers of various types. The mechanical specifications of movable operating elements of the mixing devices directly affect the reliability and their operational features (performance parameters) [21, 22, 25]. We will be considering a ribbon auger in the horizontal mixer. The outer diameter of the external auger is given in the constructional features of the mixer and is equal to $D_1 = 1$ m. The diameter of the middle D_2 and internal D_3 augers is taken from the condition of uniform material flow without the formation of stagnant zones and its effective mixing: $D_2 = 0.75$ m, $D_3 = 0.4$ m [17]. The number of cycles required to obtain a high-quality mixture will be set as N = 3; mixing time as t = 5 minutes. Given this, we are going to define the pitch of each auger [18, 25].

Let us calculate the amount of material in the mixer per one operating cycle [18, 25]:

$$M = V \cdot \phi_0 \tag{1}$$

where: M – mass of the mixture,kg;
φ_0 – specific weight of the material, kg/m^3.

Let us determine the volume of the working space in the mixer [18, 25]:

$$V = \frac{\pi \cdot D^2}{4} \cdot l \tag{2}$$

where: D is the diameter of the outer edge of the auger, m;
l – the mixer length, m.

The amount of material transported per one operating cycle of the mixer should be [18, 25]:

$$Q = N \cdot \frac{M}{t} \cdot \gamma_1 \tag{3}$$

where: Q is the capacity of a ribbon screw auger of the mixer, kg/s;
N – number of cycles of material flow by the auger for one mixing cycle;
t – mixing time, s;
γ_1 – load factor of the mixer.

Let us determine the speed of the axial flow of material v, m/s [18, 25]:

$$v = \frac{40 \cdot Q}{\pi \cdot \left(D^2 - d^2\right) \cdot \psi \cdot \phi_0 \cdot c_0} \tag{4}$$

where: d – the diameter of the inner edge of the auger, m;
ψ – fill factor of inter-turn space;
c_0 – coefficient, which takes into account the leakage of the material between the body, the material and the auger flight (screw) surface.

The fill factor of the inter-turn space should not exceed the loading capacity coefficient of the conveyor [18, 25]:

$$\psi = \psi_1 \cdot \psi_2 \tag{5}$$

where: ψ_1 – coefficient taking into account the number of auger revolutions
ψ_2 – coefficient taking into account the auger rotation angle.

The loading capacity coefficients of the screw conveyor are determined using the following empirical formulas:

$$\psi_1 = 1 - 0{,}0006 \cdot n \tag{6}$$

$$\psi_2 = 1 - 0{,}0005 \cdot \beta \tag{7}$$

where: n is the number of screw revolutions per minute;
β – screw axis tilt angle to the horizon, degrees.

Let us determine the auger flight (screw) pitch S, m [18, 25]:

$$S = \frac{60v}{n} \tag{8}$$

Angular screw rate is ω , s^{-1}:

$$\omega = \frac{\pi \cdot n}{30} \tag{9}$$

Let us calculate the rotation angle of a layer of material in the direction of the auger rotation [18, 25]:

$$\phi' = arctg\left(f_2 \cdot tg\left(\alpha + p_1\right)\right) \tag{10}$$

where: f_2 – the coefficient of friction of the material against the steel surface of the body when in motion;
α – the angle of elevation of a screw thread of the auger along the outer edge;
p_1 – the angle of friction of the material against the steel surface of the auger in motion.

$$\alpha = arctg\left(\frac{S}{\pi \cdot D}\right) \tag{11}$$

$$p_1 = arctg \cdot f_1 \tag{12}$$

where: f_1 – is the coefficient of friction of the material against the steel surface of the auger when in motion.

According to the accepted values [18, 25], after the calculations made we get: the design pitch distance for each auger: external S_1 = 0.264 m, middle S_2 = 0.496 m, and internal one S_3 = 0.48 m [18]. Based on the findings of the experimental research presented in [15, 19 and 24], and proceeding from the continuity of the mixing process, constructional (design) possibilities of making ribbon augers, reliability and operating performance, as well as the uniform flow of material by each auger taken separately, we assume the pitch distance for each mixer auger as follows: the external S_1 = 0.3 m, the middle S_2 = 0.4 m, the internal S_3 = 0.24 m.

4. Experimental research

4.1. Findings of the experimental research of the mixer

In order to determine the most effective mixing of components in a horizontal ribbon mixer, experimental studies of the direction of rotation of the mixer auger shaft for its operation cycle were carried out. In addition, the reliability of the mixer ribbon auger was assessed depending on the direction of its rotation and operational performance indicators. As the research criteria under consideration, the following rotation directions were chosen: forward (direct), backward (reverse), forward-and-backward with a 1-minute interval. According to the results of experimental studies, dependencies of the finished product homogeneity coefficient on the mixing time have been built (Fig. 1).

When conducting the experimental studies, pea was used as a control component (its specific density is 812 kg/m^3) along with the base (a mixture of barley (80%) and rye (20%) with a specific density of 742 kg/m^3) [19]. The pea weight was 100 kg, the base weight was 700 kg, the percentage ratio of the components was 12.5% - pea, 87.5% - the base, which meets the zootechnical requirements when feeding with leguminous crops. Forward (direct) rotation of the auger was considered to be the rotation when the external auger shifts the material to the centre of the mixer. Backward (reverse) rotation was considered to be the auger rotation when the external auger shifts the material from the centre to the edges (ends) of the mixer hopper. Forward-and-backward (direct-reverse) is the alternate rotation of the auger with a 1-minute interval in the forward and backward directions. The point corresponding in the dependences presented in Figure 1 to 1 minute of time, shows the value of the finished product homogeneity coefficient at the moment when material loading is finished. Then sampling from the mixer was carried out with a 1-minute interval for all selected directions of auger rotation. The selected operating modes of the mixer ribbon auger also allow evaluating its reliability.

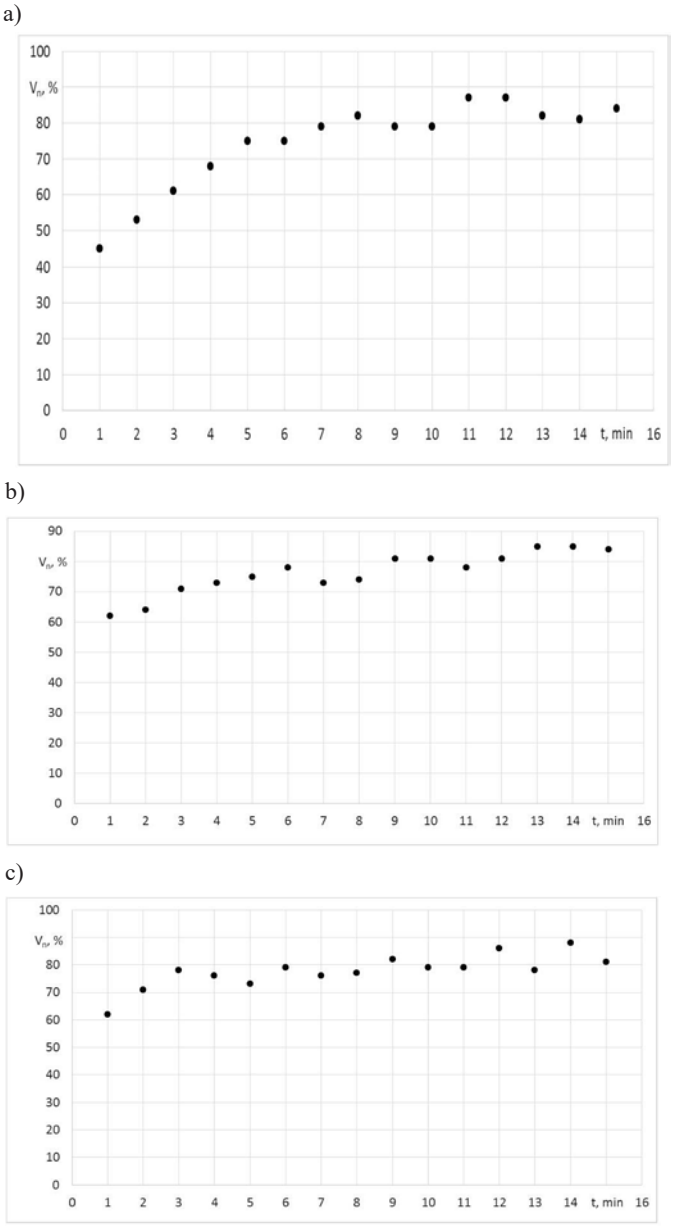


Fig. 1. Dependencies of the finished product homogeneity coefficient on the auger rotation time: a) forward, b) backward, c) forward-and-backward

Analysis of Figure 1a shows that with forward rotation of the auger within the interval from the 1st to the 12th minute of mixing, a smooth increase in the homogeneity coefficient of the finished product to its maximum of 86.7% occurs. After the 12th minute, its decrease and relative independence from the mixing time is observed, which indicates the achievement of the maximum value of the homogeneity coefficient.

Analysis of Figure 1b shows that with backward (reverse) rotation of the auger within the interval from the 1st to the 14th minute of mix-

Table 1. Mixture ratio

Degree of tank filling, %	45	55	65	75	85	95
Mixture weight, total, kg	450	550	650	750	850	950
Pea weight, kg	54	66	78	90	102	114
Basis weight, kg	396	484	572	660	748	836

ing, the homogeneity coefficient of the finished product increases to its maximum of 86%.

Analysis of Figure 1c shows that with forward-backward (direct-reverse) auger rotation within the interval from the 1st to the 15th minute of mixing, there is a slight increase in the finished product homogeneity coefficient to its maximum of 86%. That is, the value of the homogeneity coefficient during the entire time of the experiment varies insignificantly (from 73.6% in the 2nd minute to 86% in the 14th minute) and has a pronounced cyclical "more-less" trend, depending on the direction of the auger rotation.

Thus, the analysis of dependencies presented in Figure 1 shows that in further experimental studies forward auger rotation will be used, since this direction of rotation leads to the highest homogeneity coefficient of the finished product at the level of 86.7% in the shortest period of time of 11 minutes. At the same time, the throughput capacity of the mixer is 0,939 kg/s, and the specific energy consumption is $9{,}612 \cdot 10^{-3}$ MJ/kg.

Specific design features and operating modes of the ribbon auger also allow us to conclude that forward rotation of the auger shaft provides for the highest reliability of the mixer.

In further studies, we are going to determine the dependence of the homogeneity coefficient of the finished product and the reliability of the mixer auger on the filling level of the tank (hopper). Filling the mixer hopper was carried out according to Table 1. Percent control component (pea) in the amount of material in the hopper is 12% across all experimental procedures. Mixing time, without reference to loading time, is 10 minutes.

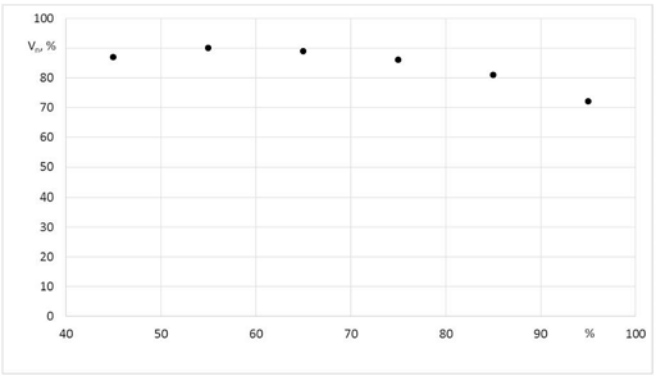


Fig. 2. Dependency of the finished product homogeneity coefficient on the degree of mixer tank filling

According to the experimental findings, the dependence of the homogeneity coefficient of the finished product on the filling degree of the mixer hopper (Figure 2) is constructed.

Analysis of the dependence presented in Figure 2 shows that the finished product homogeneity coefficient reaches its maximum when the tank is 55% full, so it is 90.5%.

When tank (hopper) is filled at the level from 60% to 95%, the value of the finished product homogeneity coefficient is reduced to 72.3%. It stems from the formation of zones where the transportation of material is not as intense due to the congestion of the mixer auger and, therefore, it affects the homogeneity of the finished product.

When the hopper is filled less than 55%, the amount of material is insufficient for its effective transportation with all three ribbon augers, therefore, we can observe that the quality of mixing is decreased.

Reducing the amount of material in the mixer hopper significantly increases the reliability of its operation by reducing the load on the operating elements.

Thus, the analysis of the experimental studies allows us to draw a conclusion that the value of the homogeneity coefficient of the finished product reaches the maximum of 90.5% with the forward direction of rotation of the mixer shaft, with 10 minutes of mixing, and with the tank load at the level of 55% of its volume. At the same time, the throughput capacity of the mixer is 1,072 kg/s, and the specific energy consumption is $7{,}308 \cdot 10^{-3}$ MJ/kg.

4.2. Optimization of constructional and operational parameters

In order to establish the optimal constructional parameters of the mixer after the implementation of single-factor experiments, studies were carried out using the multifactor experiment planning method [19, 22]. According to the results of the single-factor experiments, the following factors were selected for examination: x_1 - mixing time, min; x_2 - rotation direction of the mixer auger shaft; x_3 - the amount of material in the mixer hopper as a percentage of its possible maximum load, %. The following indicators were selected as optimization criteria: y_1 - mixture homogeneity coefficient ν_n, %; y_2 - specific energy consumption q, MJ/kg; y_3 - throughput capacity Q, kg/s. During the tests, Box-Behnken design matrix was implemented (Table 2).

The implementation of the multifactor experiment allows obtaining approximate mathematical models of the process, which link together all the factors taken into account. The experimental studies allow determining the numerical values of coefficients in equations of the mathematical models, and based on them one can estimate the degree of influence of the corresponding factors [14, 22]. The experimental findings were processed using a computer, and the following regression equations were obtained (with insignificant factors excluded):

$$y_1 = 90.49 + 4.12 \cdot x_1 - 0.16 \cdot x_2 + 0.17 \cdot x_3 - 0.75 \cdot x_1^2 - 1.18 \cdot x_1 \cdot x_2 - 1.62 \cdot x_1 \cdot x_3 - 5.12 \cdot x_2^2 - 2.93 \cdot x_3^2; \quad (13)$$

$$y_2 = 3.33 + 0.53 \cdot x_1 - 0.48 \cdot x_3 + 0.03 \cdot x_1^2 + 0.18 \cdot x_3^2; \quad (14)$$

$$y_3 = 2.36 - 0.63 \cdot x_1 + 0.44 \cdot x_3 + 0.13 \cdot x_1^2 - 0.12 \cdot x_1 \cdot x_3. \quad (15)$$

Analysis of the obtained regression equations (13-15) (based on the significance of the coefficients of the regression equations) allows us to draw a conclusion that the rotation direction of the mixer auger shaft (x_2) does not affect the optimization criteria under consideration. The mixture homogeneity coefficient ν_n (y_1), the specific energy consumption q (y_2) and the throughput capacity Q (y_3) are predominantly influenced by the mixing time of the material (x_1).

The analysis of the regression equations (13-15) and two-dimensional sections of the response surfaces (Figure 3) allow us to draw the following conclusions.

In Figure 3 it can be seen that the maximum homogeneity coefficient of the finished product ν_n = 94.13 % is achieved with the forward direction of rotation of the mixer auger shaft (x_2), 13 minutes of mixing the components (x_1) and the amount of material in the tank (hopper) equal to 52.4% of the maximum mixing chamber capacity. The through-

Table 2. Box-Behnken design matrix, intervals and levels of factors variation

Parameters	Factors		
	Mixing time, minutes	Rotation direction of the mixer auger shaft	Amount of material, %
	x_1	x_2	x_3
Upper level (+)	13 minutes	forward-backward	65
Basic level (0)	10 minutes	forward	55
Lower level (-)	7 minutes	backward	45

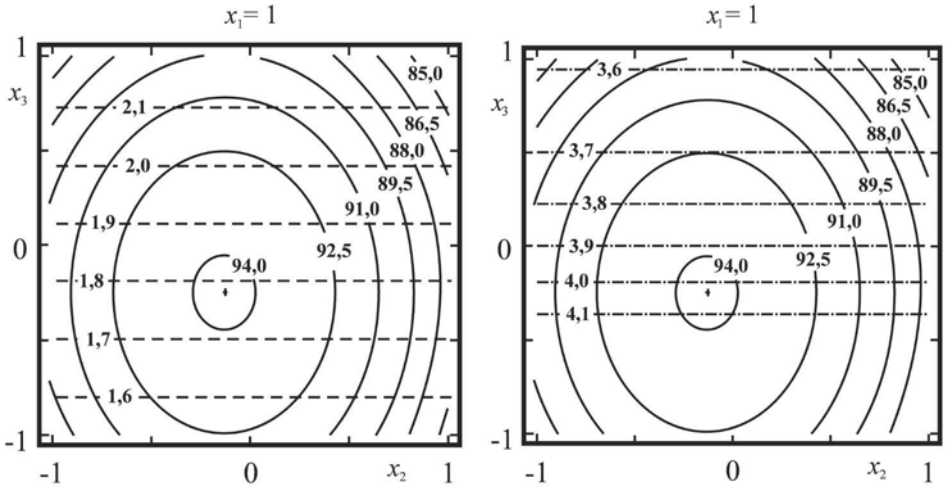


Fig. 3. Two-dimensional sections of the response surface characterizing the influence of the mixing time (x1), the rotation direction of the mixer auger shaft (x2) and the amount of material in the mixer hopper (x3) on the: a) finished product homogeneity coefficient vn (———— - y1), throughput capacity of the mixer Q (— — — — - y3); b) finished product homogeneity coefficient vn (———— - y1), specific energy consumption q (- · - · - · - · - y2)

put capacity of the mixer Q is 0,494 kg/s (1.78 t/h), and the specific energy consumption is $15{,}012 \cdot 10^{-3}$ MJ/kg (4.17 kWh/t).

When the amount of material in the tank is increased (x_3) from 45 to 65%, the throughput capacity of the mixer Q increases from 0,444 to 0,583 kg/s (1.6 - 2.1 t/h) and the specific energy consumption q decreases from $14{,}76 \cdot 10^{-3}$ to $12{,}96 \cdot 10^{-3}$ MJ/kg (4.1 - 3.6 kWh/t).

The rotation direction of the ribbon auger mixer shaft (x_2) does not affect the performance indicators.

The throughput capacity of the mixer Q reaches its maximum of 1,022 kg/s (3.68 t/h) when mixing time (x_1) is 7 minutes and the amount of material in the mixer hopper (x_3) equals 65%.

The specific energy consumption q reaches its minimum of $9{,}144 \cdot 10^{-3}$ MJ/kg (2.54 kWh/t) when mixing time (x_1) is 7 minutes and the amount of material in the mixer hopper (x_3) = 65%.

5. Conclusions

1. A theoretical substantiation of the consructional parameters of a horizontal ribbon auger mixer was carried out taking into account the fulfilment of the condition of transporting the material by auger ribbons in different directions in order to ensure effective mixing of the mixture components and timely unloading of the material from the mixer tank (hopper). On the basis of theoretical studies, and proceeding from the continuity of the mixing process, constructional (design) possibilities of making ribbon augers, reliability and operating performance, as well as the uniform flow of material by each auger taken separately, the pitch distance for each mixer auger was assumed as follows: the external S_1 = 0.3 m, middle S_2 = 0.4 m, internal S_3 = 0.24 m.
2. Structural specifications and operating modes of the ribbon auger also allow us to conclude that forward (direct) rotation of the auger shaft provides the highest mixer reliability, while the decrease in the amount of material in the mixer tank (hopper) greatly increases the reliability of its operation by reducing the load on the working elements.
3. Optimization of the constructional (design) and operational parameters as a result of the experimental studies made it possible to determine that the finished product homogeneity coefficient reaches its highest value of 94.13% when there is forward rotation direction of the mixer operating elements, mixing time is 13 minutes and the tank load is 52.4% of its volume. At the same time, the mixer throughput capacity is 0,494 kg/s (1.78 t/h), and the specific energy consumption is $15{,}012 \cdot 10^{-3}$ MJ/kg (4.17 kWh/t).

1. Balami A A, Adgidzi D, Mua`zu A. Development and testing of an animal feed mixing machine. International Journal of Basics and Applied Sciences 2013: 1(3): 491-503, https://doi.org/10.17142/ijbas-2012.1.3.4.
2. Chan E L, Washino K, Ahmadian H, Bayly A, Alam Z, Hounslow M J, Salman A D. Dem investigation of horizontal high shear mixer flow behaviour and implications for scale-up. Powder Technology 2015; 270(B): 561-568.
3. Çiftçi I, Ercan A. Effects of diets of different mixing homogeneity on performance and carcass traits of broilers. J. Anim. Feed Sci. 2003; 12(1):163-171, https://doi.org/10.22358/jafs/67693/2003.
4. Desai A T, Kumbhar M G, Deokar R H, Mandhare A M. Design and analysis of powder mixing ribbon blender – A review. IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE). 1st National Conference on Recent Innovations in Mechanical Engineering (NCRIME) 2018: 39-43.
5. Djuragic O, Levic J, Serdanovic S, Lević L. Evaluation of homogeneity in feed by method of microtracers. Archiva Zootechnica 2009; 12(4): 85-91.
6. Dušica J, Djuragic O, Kokic B. Two-step mixing in feed premixes production. Journal on Processing and Energy in Agriculture 2009; 13(3): 280-282.
7. Flizikowski J, Tomporowski A. Charakterystyki ruchowe wielotarczowego rozdrabniacza ziaren zbóż. Przemysł Chemiczny 2013; 92(4): 498-503.
8. Flizikowski J, Sadkiewicz J, Tomporowski A. Charakterystyki użytkowe sześciowalcowego mielenia uziarnionych surowców dla przemysłu chemicznego i spożywczego. Przemysł Chemiczny 2015; 94(1): 69-75.
9. GOST R ISO 6497-2011 Korma dlia zhivotnykh. Otbor prob - M . Standardinform, 2012; 18.

10. Krolczyk J B. Homogeneity assessment of multielement heterogeneous granular mixtures by using Multivariante Analysis of Variance. Tehnicki Vjesnik 2016; 23(2): 383-388.
11. Krolczyk J B. The effect of mixing time on the homogeneity of multi-component granular systems. Transactions of FAMENA 2016; XL-1: 45-56.
12. Kulat P S, Chadge R B. 3D motion mixer for material. IJPRET 2014; 2(9): 97-102.
13. Makange N R, Rarmar RP, Sungawa N. Design and Fabrication of an Animal Feed Mixing Machine. Advances in Life Sciences 2016; 5(9): 3710-3715.
14. Marczuk A, Caban J, Savinykh P, Turubanov N, Zyryanov D. Maintenance research of a horizontal ribbon mixer. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (1): 121–125, http://dx.doi.org/10.17531/ein.2017.1.17.
15. Muzzio F J, Llusa M, Goodridge Ch L, Duong N-H, Shen E. Evaluating the mixing performance of a ribbon blender. Powder Technology 2008; 186: 247-254, https://doi.org/10.1016/j.powtec.2007.12.013.
16. Robinson M, Cleary P W. Flow and mixing performance in helical ribbon mixers. Chemical Engineering Science 2012; 84: 382-398, https://doi.org/10.1016/j.ces.2012.08.044.
17. Savinykh P A, Turubanov N V, Romaniuk W, Kiselyov A S, Chernyatyev N A. Pat. 2488434 Rossiyskaya Federatsiya, MPK7 B 01 F 7/8. Smesitel / Applicant and patentee - State Scientific Institution Zonal Research Institute of Agriculture of the North-East named after N.V. Rudnitskiy (Russian Federation) and the Technological and Natural Science Institute (Poland). – No.2012104920/05, applied on February 13, 2012; published on July 27 2013, Inform. letter. No. 21. - 5 pp.: incl. illustrations.
18. Savinykh P A, Turubanov N V, Zyryanov D A. Obosnovaniye konstruktsionno-tekhnologicheskikh parametrov lentochnogo smesitelia. Vestnik Vserossiyskogo nauchno-issledovatelskogo institute mekhanizatsii zhivotnovodstva 2015; 3 (19): 76-80.
19. Savinykh P A, Turubanov N V, Zyryanov D A. Rezultaty eksperimentalnykh issledovaniy protsessa smeshivaniya v gorizontalnom lentochnom smesitele. Tractors and agricultural machinery 2016: (7): 32-36.
20. Singhai S K, Chopra V S, Nagar M, Jain N, Trivedi P. Scale up factor determination of V blender: An overview. Der Pharmacia Lettre 2010; 2(2): 408-433.
21. Sysuev V A, Aleškin A V, Savinykh P A, Marczuk A, Wrotkowski K, Misztal W. Studium agregatów mieszalniczych pasz z transportem pneumatycznym. Uniwersytet Przyrodniczy w Lublinie, 2014.
22. Sysuev V A, Aleškin A V, Savinykh P A, Marczuk A, Wrotkowski K, Misztal W. Badanie mobilnych rozdrabniaczy oraz rozdrabniaczy-mieszarek pasz. Uniwersytet Przyrodniczy w Lublinie, 2016.
23. Tomporowski A, Flizikowski J, Kruszelnicka W. Nowa koncepcja młyna walcowo-płytowego. Przemysł Chemiczny 2017: 96(8): 1750-1755, https://doi.org/10.15199/62.2017.8.29.
24. Turubanov N V, Medvedev O Yu, Isupov A Yu. Issledovaniya smeshivaniya materiala v gorizontalnom lentochnom smesitele. In the Proceedings: SOCIETY, SCIENCE, INNOVATION (Research-to-Practice Conference RPC-2016). Collection of articles, 2nd edition, revised and amended. Vyatka State University 2016; 864-870.
25. Turubanov N V. Raschet osnovnykh konstruktsionnykh parametrov shneka smesitelya. In the Proceedings: Society, Science, Innovation (RPC - 2015) All-Russian Annual Research-to-Practice Conference: Proceedings: All-University Section, BF, HF, FSA, FAM, ETF, FAVT, FPMT, FEM, FGSN, YuF. Federal State Educational Institution of Higher Professional Education "Vyatka State University" 2015; 502-507.
26. Yeow S T. Shahar A, Adbul Aziz N, Anuar M S, Yusof Y A, Taip F S. The influence of operational parameters and feed preparation in a convective batch ribbon powder mixer. Drug Design, Developmnet and Therapy 2011; 5: 465-469.

Andrzej MARCZUK
Wojciech MISZTAL
Department of Agricultural, Forest and Transport Machines
University of Life Sciences in Lublin
ul. Głęboka, 20-612 Lublin, Poland

Petr SAVINYKH
Nikolay TURUBANOV
Alexey ISUPOV
Dmitry ZYRYANOV
North-Eastern Agricultural Scientific-Research Institute, Kirov, Russia
Russian Academy of Agricultural Sciences
Lenin Str., 610017 Kirov, Russia

E-mails: andrzej.marczuk@up.lublin.pl, wojciech.misztal@up.lublin.pl, peter.savinyh@mail.ru, nikolaytu@mfil.ru, dimonzyryanov@vfil.ru

Article citation info:
MACIÁN V, TORMOS B, HERRERO J. Maintenance management balanced scorecard approach for urban transport fleets. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 226–236, http://dx.doi.org/10.17531/ein.2019.2.6.

Vicente MACIÁN
Bernardo TORMOS
Jesús HERRERO

MAINTENANCE MANAGEMENT BALANCED SCORECARD APPROACH FOR URBAN TRANSPORT FLEETS

ZARZĄDZANIE UTRZYMANIEM RUCHU W ZAKŁADACH KOMUNIKACJI MIEJSKIEJ W OPARCIU O ZRÓWNOWAŻONĄ KARTĘ WYNIKÓW

Attending the important role of maintenance function in any production or service provider company, the measurement and assessment of maintenance performance is crucial for competitiveness and future survival. That situation is even more critical in urban transport fleets where some specific boundary conditions and special characteristics will affect maintenance policy and implementation. This paper presents a deep review of different studies worldwide to define the most proper and effective maintenance performance indicators, selecting and refining the most important ones to obtain a reduced maintenance management balanced scorecard. That balanced scorecard is proposed as a main tool for urban transport fleet maintenance managers to assess efficiency and effectiveness of maintenance processes and will be used as a basis for a future benchmarking process for this type of companies.

***Keywords**: KPIs, balanced scorecard, urban transport fleet, maintenance management.*

Biorąc pod uwagę ważną rolę jaką pełni utrzymanie ruchu w firmach produkcyjnych i usługowych, pomiar i ocena wydajności eksploatacji ma kluczowe znaczenie dla konkurencyjności tych firm i ich przetrwania na rynku. Sytuacja ta jest szczególnie ważna w zakładach komunikacji miejskiej, w których pewne szczególne warunki brzegowe i szczególne cechy floty transportowej mają wpływ na politykę utrzymania ruchu i jej realizację. W niniejszym artykule przedstawiono dokładny przegląd różnych badań prowadzonych na całym świecie w celu określenia najbardziej odpowiednich i skutecznych wskaźników efektywności utrzymania ruchu, wybierając najważniejsze z nich i i udoskonalając je tak aby uzyskać zrównoważoną kartę wyników zarządzania utrzymaniem ruchu z ograniczoną liczbą parametrów. Zrównoważona karta wyników może być stosowana przez specjalistów utrzymania ruchu zakładów komunikacji miejskiej do oceny wydajności i skuteczności procesów konserwacji i stanowić będzie podstawę przyszłych analiz porównawczych dla tego typu przedsiębiorstw.

***Słowa kluczowe**: kluczowy wskaźnik efektywności, zrównoważona karta wyników, komunikacja miejska, zarządzanie utrzymaniem ruchu.*

1. Introduction

Maintenance is one of the largest expenditures for the urban transport companies together with fuel (or energy) costs and drivers (personnel) [16], but is the most important one from the view of controllability, attending that fuel and labour costs are more externally driven (crude prices volatility, taxes, personnel policies and salaries, etc.). A proper maintenance policy, managerial processes and planning and optimization of maintenance decisions, scheduling and execution of work can lead to reduce costs, improve vehicle effectiveness, reliability and performance. Consequently, maintenance function is therefore vital for sustainable performance of any urban transport fleet.

Attending the responsibility of ensuring that urban fleet achieves the expected performance, maintenance managers requires a tracking system for maintenance operations performance and results [3, 6, 20, 21]. Furthermore, it is in the interest of maintenance managers to know the relationship between the input of the maintenance process and the outcome in terms of total contribution to vehicle fleet performance and strategic objectives. The measurement of maintenance performance is indeed an essential requirement for any industry of today. This tracking action can be done through development and implementation of a proper and accurate performance measurement system and indicators that are able to quantify important elements of maintenance function performance [5, 15]. Moreover, without having a formal measurement system for maintenance performance, it is difficult to control, plan and improve the maintenance process and consequently can be considered that tracking the performance operations must be a key management issue in any industrial organization.

In summary, a proper maintenance performance measurement system shall contribute to:

- Assess the maintenance function to the strategic company targets.
- Detect the strengths and weaknesses on the maintenance strategy applied.
- Using quantitative and qualitative data for helping to define and stablish a continuous improvement process for maintenance.
- Let us to apply a maintenance benchmarking analysis within and outside the business related with urban fleet transport.

Different frameworks have been defined and used in the past for measuring performance and until the 80s in last century mostly based on financial measures. The original balanced scorecard defined by Kaplan and Norton [14] had four perspectives for analysing: financial aspects, customers, internal processes and innovation and learning. Later, different approaches has been defined by other authors considering other non-financial measurements and intangible assets to achieve competitive advantages [2, 4, 25].

Last, but not least, benchmarking is used for business development and also for improving efficiency and effectiveness of maintenance processes in any type of industry. The analysis performed on this work provides a basis for learning from the top class business companies and offers a road map for performance improvement [12]. As a prior requirement to begin a benchmarking analysis obviously is required a set of proper, reliable, accurate and well defined performance indicators for the industrial sector considered, as it has been previously mentioned.

This paper presents a deep revision of key performance indicators for maintenance management in the specific and very important sector of urban transport fleets in the section 2. In section 3 authors propose a reduced selection of key parameters that can be considered the most important for this application grouped into three main categories. Later, in section 4, those selected parameters are developed and it is presented how can be calculated and managed. Finally, in section 5 the balanced scorecard proposal is presented combining the previous parameters defined.

2. State of the art regarding KPIs for maintenance management in transport fleets.

In order to perform a review of the state of the art regarding maintenance management KPIs, authors have focused firstly on more general documents, mainly international standards, with a more general approach to that problem and later have focused on more specific literature regarding public transport fleets. In this way, publications and documents coming from UITP (Union International des Transports Publics or International Association of Public Transport), and other information from international transport associations have been managed. Next, a complete summary of the review performed is presented.

2.1. EN 15341 Maintenance - Maintenance Key Performance Indicators

This is a European standard [9] and is the most general standard referred to maintenance KPIs. The Spanish version is the UNE EN 15341:2008. As it is stated on the introduction, this standard: "provides Maintenance Key Performance Indicators to support management in achieving *maintenance excellence* and utilize technical assets in a competitive manner. The majority of these indicators *apply to all industrial and supporting facilities* (buildings, infrastructure, *transport,* distribution, networks, etc.)".

This standard defines a set of indicators structured on a three different levels: economics, technical and organizational and mainly are presented such a relation between factors (numerator and denominator) related with activities, resources or events. Maintenance performance and consequently key performance indicators can be affected by internal factors such as group, company, factory, plant that are outside of the maintenance manager control but inside of the company management control. On the other side, external factors such as location, market, laws and regulation, etc. are variable conditions outside company management control also affecting maintenance performance.

Most indicators can be used at different levels depending on whether they are used to measure the performance of plant production, one production line, or a given equipment or item, i.e. are structured from a more general to a more detailed point of view.

On this standard are summarized 24 economic indicators, 21 technical indicators and 26 organizational indicators. Each company must select the most proper indicators attending own characteristics and objectives expected.

As it has been mentioned, that standard is very general and can be applied to any type of industry and consequently perhaps have not into account specific characteristics more related with a urban transport fleet company; but after the analysis of that standard, the Key Parameters Indicators more proper for that companies would be:

Economic indicators: E3, E14, E15, E16 and E17.

Technical indicators: T1. T2, T6 and T21.

Next, as a reminder, general definition of that indicators are presented on Table 1.

2.2. EN 13816- Transportation-Logistics and services – Public passenger transport service quality definition, targeting and measurement.

This standard [8] is mainly focused to promote a quality approach to public transport and focus interest on customers' needs and expectations. This standard collect the quality criteria representing customer view of the service provided divided into eight categories. It can be considered that there are three of them directly related with vehicle maintenance: comfort, security and environmental impact. That general quality criteria are subdivided on more specific items and for those selected previously the next Table 2 present the most important ones from the point of view of authors related with maintenance.

Table 1. Definition of KPIs (following EN 15341) more suitable for urban transport fleets

KPI	Economic	KPI	Technical
E3	$\frac{Total\ Maintenance\ Cost}{Quantity\ of\ output}$	T1	$\frac{Total\ Operating\ time}{Total\ Operating\ time + Downtime\ due\ to\ maintenance}\cdot 100$
E14	$\frac{Total\ Maintenance\ Cost}{Total\ Energy\ Used}$	T2	$\frac{Achieved\ up\ time\ during\ required\ time}{Required\ time}\cdot 100$
E15	$\frac{Corrective\ Maintenance\ Cost}{Total\ Maintenance\ Cost}\cdot 100$	T6	$\frac{Total\ Operating\ time}{Total\ Operating\ time + Downtime\ related\ to\ failures}\cdot 100$
E16	$\frac{Preventive\ Maintenance\ Cost}{Total\ Maintenance\ Cost}\cdot 100$	T21	$\frac{Total\ time\ to\ restoration}{Number\ of\ failures} = MTTR$
E17	$\frac{Condition\ based\ Maintenance\ Cost}{Total\ Maintenance\ Cost}\cdot 100$		

Table 2. Quality criteria and sub-items more related with Maintenance aspects in urban transport fleets

Quality criteria	Sub-items
#6 Comfort	6.1 Usability of passenger facilities
	6.3 Ride comfort
	6.4 Ambient conditions
#7 Security	7.2 Freedom from accident
#8 Environmental impact	8.1 Pollution
	8.2 Natural resources

2.3. Others indicators at interna tional level

A deep review has been performed on scientific papers published, consultancy works, Transport associations, good practices, etc. Next, a summary of the most relevant documents found related specifically with urban transport fleets are presented on Table 3.

Most documents reviewed present a lot of indicators for a complete fleet management, ranging from general company

Table 3. Summary of the most relevant documents managed for that study

Title	Origin / Authors	Brief description	Year	Country
Field study on bus depots and bus maintenance. Similarities between and singularities of different engine technologies and their impact on infrastructure, quality and maintenance. [10]	UITP (Union Internationale des Transports Publics)	UITP study regarding different propulsion technologies (diesel, hybrid, CNG, electric, etc.) and a comparative assessment under different concepts.	2013	Belgium
Managing for Results in America's Great City Schools. [18]	Council of the Great City Schools	This report defines and presents an extensive array of statistical indicators developed by the Council of the Great City Schools and its member urban school districts to measure performance on a broad range of operational functions, including business services, finances, human resources and technology.	2017	USA
A framework for urban Transport benchmarking. [13]	The World Bank	This report summarizes the findings of a study aimed at exploring key elements of a benchmarking framework for urban transport.	2011	USA
The Fleet Management Key Performance Indicators (KPIs) (https://knowledge.fleetforum.org/knowledge-base/article/assessing-your-fleet-management)	Fleet Forum	Excel spreadsheet developed by Fleet Forum (association with more than 40 members including UN; different NGO, academic institutions, donnors and corporative partners).	2013	Switzerland
MBTA Bus Maintenance Efficiency Study. [19]	MBTA (Massachusetts Bay Transportation Authority)	MBTA approach to identify alternatives and develop recommendations for maintenance efficiency improvement.	2016	USA
Fleet owner. KPI's for maintenance: 15 suggestions for numbers that matter. [24]	Fleet Owner	KPIs proposal coming from different associated companies.	2016	USA
Performance Metrics for the New Fleet Manager. Wheels & Wings - Online Newsletter for the Federal Motor Vehicle and Aviation Communities. [11]	Larry Fredrich, Mercury Associates	Paper from a consultant where some fleet maintenance manager performance indicators are summarized.	2014	USA
Developing optimum KPI system for Public Transport Organizations. [1]	SIGMA journal	Egyptian researcher work presenting a study regarding public transport companies' performance indicators.	2016	Egypt
Useful Key Performance Indicators for Maintenance. [23]	www.lifetime-reliability.com	Set of indicators proposed by this Australian web for maintenance management improvement.	2009	AUSTRALIA

International Bus System Benchmarking: Performance Measurement Development, Challenges, and Lessons Learned. [22]	International Bus Benchmarking Group (IBBG)	The IBBG is a comprehensive programme of benchmarking urban bus operations. Currently the consortium is made up of 16 medium and large bus organizations located around the world.	2004	UK

Table 4. Summary of indicators presence on different international studies. Higher presence is a clear clue of relevance and meaningful

Indicator	Number of studies where is mentioned
Average Fleet age *	5
Averaged service speed *	4
Annual mileage *	8
Fleet availability	10
Cost per kilometer (or mile)	10
MTBF	6
CO_2 emission (Tons)	5
Energy Consumption	9
MTTR	3
Accidents per kilometer (or mile)	7
Work orders (WO) per year	2
Fleet Reserve %	1
Preventive Maintenance Program Fulfillment	3
Life Cycle Cost (LCC)	2
Absenteeism	3
Parts list	1
Kilometers (miles) lost service	1

Note (*). These parameters are fleet operational data but are required for assessment and monitoring of other key indicators.

management, customer satisfaction, security, finance, environmental aspects and maintenance. Some of them are very similar, with just a little variation on the definition attending mainly to the data availability in order to obtain the indicator.

For this work, authors have just considered those related with maintenance at economic and technical level. Next, a summary can be observed in Table 4 with those indicators mainly considered and a number representing how many times appears on the ten previous literature sources considered:

It is necessary to mention that the Life Cycle Cost (LCC) more than an indicator is a tool in order to help the decision makers to define if a fleet renovation is required based on real data.

3. Key Parameters Selection

After a deep revision of the state of the art, it is necessary to take into account that KPIs selected must led us to quantify the maintenance process to control and assess its performance and must contribute to process improvements and also for helping decision makers. With that initiative, it would be possible to apply a continuous improvement policy and define control boundaries and "non conformity" limits, with cause's analysis and solutions definitions.

Some major characteristics must be accounted for KPIs selection and definition in order to assure future usability and validity and consequently to obtain a clear, relevant and reliable indicator:

- Easy and quick procurement of data involved on the indicator.
- Proper frequency of data procurement.
- Clarity of the info obtained and showed by the indicator.
- High info interest.
- Graphic trends analysis.
- Easy definition of target or optimal value for such indicator (or limits levels).

Attending previous characteristics, it is proposed to define 3 main indicator groups:

- Structural or own company indicators: this are indicators referring to main fleet general characteristics that are important to take into account for the assessment of others indicators groups, that is, this are boundary conditions affecting maintenance department.
- Basic indicators: That are the indicators defined for technical and economical fleet maintenance management assessment. That indicators led us to evaluate the maintenance execution.
- Advanced indicators. That are the indicators that led us to assess the maintenance repercussion or those consequences derived from our maintenance management.

Next, on Table 5, it is presented a brief summary of each group including indicators considered:

As it can be seen, authors have tried to avoid a vast number of indicators that consequently requires an extensive amount of human resources and financial budget and which may not be feasible for a long-term process. Furthermore, the indicators defined as advanced can be used or bring into play after a first approach with more simple indicators and training people involved on that process, in order to assure success on that more difficult performance metrics.

Table 5. Main parameters defined for BSC proposed

Structural indicators	Basic Indicators	Advanced indicators
Total mileage per year	Maintenance cost / kilometer	Energy consumption kW/km (per powertrain typology: diesel, electric, hybrid, CNG, so on).
Average Fleet age	Availability	Total CO_2 emissions
Averaged service speed	Failures / 10000 Km (related to maintenance)	MTTR
Number of vehicles per maintenance personnel	Accidents / 10000 Km (related to maintenance)	Maintenance delay
Externalization percentage	Maintenance programme fulfilment.	Optimal vehicle period replacement (by LCC)
	Maintenance personnel absenteeism	

4. Indicators definition

Right now, the proposed indicators are going to be defined in a detailed way in order to be comparable between different fleet companies in future benchmarking activities. Attending that data can be obtained by different ways, different alternatives will be defined for a same indicator (if required) that will be noted with different sub index.

For an easier process, we are going to use a record card that will help for simple calculus and future auditing.

4.1. Structural parameters

4.1.1. Total mileage performed

Name	KM-total mileage performed
Definition	Addition of total mileage performed by entire fleet monthly
Formulae	$KM_1 = (\sum total\ mileage\ of\ each\ vehicle\ of\ the\ fleet)km$
Units	Total kilometers (whole number, no decimals: xxx xxx km)
Target value	Unnecessary
Calculus frequency	Monthly measurement and annual monitoring and assessment
Additional comments	Total mileage can be obtained per vehicle models, vehicles types or fuel in order to obtain a more specific indicator.

4.1.2. Averaged fleet age

Name	ED-Average fleet test
Definition	Addition of the age of each vehicle divided by the number of vehicles considered.
Formulae	$ED_1 = \left(\frac{\sum age\ of\ each\ vehicle}{Number\ of\ total\ vehicles} \right) years$
Units	Years with two decimals (xx.xx years)
Target value	Target value should be between 6 and 7 years depending on the optimal fleet replacement value obtained by the LCC analysis.
Calculus frequency	Annual
Additional comments	Average age can be obtained per vehicle models or types and even define different target values depending on that differences assuming a specific LCC analysis.

4.1.3. Average service speed

Name	VM-average fleet speed (
Definition	Addition of average speed of each vehicle per service (urban, interurban, ..) divided by the number of vehicles considered
Formulae	$VM_1 = \left(\frac{\sum average\ speed\ of\ each\ vehicle}{Number\ of\ total\ vehicles} \right) km/h$
Units	Kilometers per hour with two decimals (xx.xx km/h)
Target value	Unnecessary

Calculus frequency	Monthly measurement and annual monitoring and assessment
Additional comments	That average speed can be calculated for each type of service: urban, interurban, mix or others.

4.1.4. Number of vehicles per maintenance personnel

Name	VP-Number of vehicles per maintenance personnel
Definition	Total number of vehicles divided by the number of maintenance personnel (direct maintenance workers, managers and administrative personnel in Maintenance area – own or externalized)
Formulae	$$VP_1 = \left(\frac{Total\ number\ of\ vehicles}{Number\ of\ maintenance\ personnel} \right) vehicles\ /\ person$$
Units	Vehicles per person with one decimal (xx.x vehicles/person)
Target value	Depending on company size and externalization level
Calculus frequency	Annual
Additional comments	That indicator could be calculated just considering direct maintenance workers.

4.1.5. Externalization percentage

Name	PC-Externalization percentage
Definition	Externalized or subcontracted maintenance cost versus total maintenance cost obtained in the indicator of total maintenance cost per kilometer.
Formulae	$$PC_1 = \left(\frac{\sum externalized\ maintenance\ cost}{Total\ Maintenance\ cost} \right) x100\ \%$$
Units	Percentage with one decimal (xx.x %)
Target value	Depending company structure and size
Calculus frequency	Annual
Additional comments	That parameter must be considered for the assessment of other indicators.

4.2. Basic Parameters

4.2.1. Maintenance cost per kilometer

Name	CM-Maintenance cost per kilometer
Definition	That is the cost for manpower, parts and legal inspections fees devoted to maintenance. It would include: MOD- Direct manpower. Manpower directly related with corrective, preventive, condition monitoring or modification in vehicles. It could be included vehicles cleaning and refueling although usually are externalized tasks it is highly recommended consider it in other specific item. MOI- Indirect manpower. Manpower cost considered but not directly referred to a vehicle such as: maneuvers drivers, interchanging parts repair, so on. TEX- External workshops. Costs related with maintenance actions performed on external workshops and not included in other items (such tires or legal inspections) MAT-Parts. Parts costs related directly to vehicles (not included fuel and ad-blue that are considered operational costs). MAC-General parts. General parts costs not related with specific vehicles such as screws, rags, air conditioner refrigerant charge, etc. ACN-Lubricants, cooling liquid and tires. ITG-Legal required inspections. Cost relates with complimentary legal inspections to be performed on vehicles plus other ones required for some specific design (for instance, high pressure deposit inspection for CNG vehicles). LRC-Vehicle cleaning, refueling and liquids level control. Usually that costs are externalized and not considered in any other previous parameter. GST-management cost. Costs related with management personnel and intermediate level including: maintenance manager, workshop manager, engineering manager, administrative, team manager, warehouse personnel, maintenance purchasing personnel, so on.
Formulae	$$CM_1 = \left(\frac{\sum (MOD + MOI + TEX + MAT + MAC + ACN + ITG + LRC + GST}{Total\ fleet\ mileage} \right) € / km$$
Units	Euros (€) per kilometer including 4 decimals (xx.xxxx €/km)

Target value	Attending that it will depend on several factors such as: fleet age, type of vehicles, etc.; it should be obtained trends and minimum and maximum values to define a target value.
Calculus frequency	Monthly measurement and annual monitoring and assessment attending the indicator seasonality.
Additional comments	That is a parameter that led us to assess the maintenance efficiency and could be very useful for abnormal situation detection considering that there a lot of different factors affecting it. That parameter of maintenance cost per kilometer and all of the costs considered sorted by different items such as models or vehicle technology can be very useful to help in strategic company decisions.

4.2.2 Availability

Name	DS-Availability
Definition	Percentage indicator representing the time that vehicle is available to perform as and when required for fleet service.
Formulae 1	$DS_1 = 1 - \left(\frac{\sum total\ time\ for\ vehicles\ inmobilization\ in\ hours}{\sum total\ vehicles\ \ fleet\ required\ time\ in\ hours} \right) x100\ \%$
Formulae 2	$DS_2 = 1 - \left(\frac{\sum total\ time\ for\ vehicles\ inmobilization\ in\ days}{\sum total\ vehicles\ \ fleet\ required\ time\ in\ days} \right) x100\ \%$
Units	Percentage with one decimal (xx.x %)
Target value	Target value should be higher than 90% and can be considered optimum at 95%.
Calculus frequency	Monthly
Additional comments	That is a great indicator to assess the maintenance policy and management efficiency. Additionally, some variations could be defined for considering vehicles on reserve, vehicles on demand, etc.

4.2.3. Failures per 10.000 km

Name	KA-Failures per 10.000 km
Definition	That is a typical indicator for transport fleet representing the inverse of MTBF. Failures computed are all those that send the vehicle to the workshop and are no related with preventive or predictive maintenance activities.
Formulae 1	$KA_1 = \left(\frac{\sum failures}{\sum km} \right) x10000$
Units	Failures per 10000 km with two decimal (xx.xx failures/10000 km)
Target value	Target value should be lower than 2 and can be considered optimum between 1 and 1.5 failures/10000 km.
Calculus frequency	Monthly but annual monitoring and assessment annual attending that there is a seasonality effect.
Additional comments	That indicator can be customized o modified to obtain more specific info such as: failure type (mechanical, electric, etc.); by vehicle model or powertrain, for a specific period of time or mileage, so on.

4.2.4. Accidents per 10 000 km

Name	SN-Accidents per 10 000 km
Definition	That indicator computes all the accidents reports performed during service for the complete fleet. Typical indicator must take into account all the accidents reports; but for maintenance assessment it can be considered a slightly modification, considering just those reports related with accidents derived from a vehicle failure such as: brakes, direction and suspension, so on.
Formulae 1	$SN_1 = \left(\frac{\sum accident\ reports}{\sum km} \right) x10000$
Formulae 2	$SN_2 = \left(\frac{\sum accident\ reports\ coming\ \ from\ vehicle\ \ failure}{\sum km} \right) x10000$
Units	Accidents per 10 000 km with two decimal (xx.xx accidents/10 000 km)
Target value	Target value should be 0
Calculus frequency	Monthly but annual monitoring and assessment annual attending that there is a seasonality effect.
Additional comments	That indicator can be specified for more accurate analysis by parameters such as: vehicles models, service line, for a specific period of time or mileage, so on.

4.2.5. Maintenance plan fulfilment

Name	CM-Maintenance plan fulfilment
Definition	That indicator represents the preventive maintenance program execution versus planning, giving an indication of how far is the real situation versus ideal or complete fulfilment of that program.
Formulae 1	$$CM_1 = \left(\frac{\sum Preventive\ Maintenance\ WO\ performed}{\sum Preventive\ Maintenance\ WO\ scheduled} \right) x100\ \%$$
Units	Percentage with one decimal (xx.x %)
Target value	Target value should be higher than 95% and can be considered optimum at 100%.
Calculus frequency	Monthly.
Additional comments	That indicator should be analyzed considering also other parameters such as: maintenance WO delay, WO execution time expected vs real, so on.

4.2.6. Maintenance personnel absenteeism

Name	AB- Maintenance personnel absenteeism rate
Definition	That indicator is the relation between absenteeism hours for maintenance work force and total labor hours for maintenance work force.
Formulae 1	$$AB_1 = \left(\frac{\sum Maintenance\ WF\ absenteeism\ hours}{\sum Maintenance\ WF\ labor\ hours} \right) x100\ \%$$
Formulae 2	$$AB_2 = \left(\frac{\sum Maintenance\ WF\ absenteeism\ days}{\sum Maintenance\ WF\ labor\ days} \right) x100\ \%$$
Units	Percentage (hours or days) with one decimal (xx.x %)
Target value	Target value should be lower than 4%. Averaged value for Spain in 2000-2016 period has been 4.5% based on data from Statistical National Service.
Calculus frequency	Monthly but annual monitoring and assessment annual.
Additional comments	Absenteeism rate is a classical indicator for any type of industry. Usually is computed considering different shifts of the company and normally is higher for the night shift, which is a very common situation on urban transport fleets.

4.3. Advanced Parameters

4.3.1. Energy consumption kWh/100 km (Diesel, CNG, Hybrid or Electric)

Name	CE- Energy consumption
Definition	That indicator must be obtained following the EN 16258 standard (Methodology for calculation and declaration of energy consumption and GHG emissions of transport services (freight and passengers)). It should be considered energy consumption at local level that is referred as tank to wheel and using conversion factors depending on the fuel in use.
Formulae CE_1 (Diesel and hybrids)	$$CE_1\left(\frac{kWh}{km}\right) = \left(Average\ diesel\ fuel\ consumption\left(\frac{l}{100\,km}\right) \div 100 \right) \cdot 35.9 \frac{MJ}{l} \cdot \frac{1000}{3600} kJ/s \left(\frac{kWh}{km}\right)$$
Formulae CE_1 (CNG)	$$CE_1\left(\frac{kWh}{km}\right) = \left(Average\ CNG\ fuel\ consumption\left(\frac{Nm^3}{100\,km}\right) \div 100 \right) \cdot 45.1 \frac{MJ}{kg} \cdot \rho_{NG} \frac{kg}{m^3} \cdot \frac{1000}{3600} kJ/s \left(\frac{kWh}{km}\right)$$
Formulae CE_1 (electrics)	$$CE_1\left(\frac{kWh}{km}\right) = \frac{Electricity\ consumption\ electric\ buses\ kWh}{Electric\ fleet\ mileage\ performed\ km}\left(\frac{kWh}{km}\right)$$
Units	Energy consumption per kilometer with two decimals (xx.xx kWh/km)
Target value	Target value should be defined as a trend, trying to obtain energy consumption reduction as a consequence of fleet renewal or fuel consumption reductions initiatives.
Calculus frequency	Monthly measurement and annual monitoring and assessment attending the indicator seasonality.
Additional comments	That parameter also can be estimated sorting by vehicles models or manufacturers, and additionally can be referred not just to mileage and also to users or passengers transported leading to possible benchmarking comparison with similar companies.

4.3.2. Total CO_2 emissions

Name	EM- Total CO_2 emissions
Definition	That indicator must be obtained following the EN 16258 standard (Methodology for calculation and declaration of energy consumption and GHG emissions of transport services (freight and passengers)). It should be considered emissions at local level that is referred as tank to wheel and using conversion factors depending on the fuel in use.
Formulae EM_1 (Diesel)	$EM_1(kg\,CO_2)=\left(Average\,diesel\ fuel\,consumption\left(\frac{l}{100\,km}\right)\right)\cdot\left(\frac{Diesel\ fleet\,mileage}{100}\right)\cdot 2.67 kg\,CO_2\,/\,l$
Formulae EM_1 (CNG)	$EM_2(kg\,CO_2)=\left(Average\,CNG\ fuel\,cons.\left(\frac{Nm^3}{100\,km}\right)\cdot\left(\rho_{CNG}\,\frac{kg}{m^3}\right)\right)\cdot\left(\frac{CNG\ fleet\,mileage}{100}\right)\cdot 2.68\,kg\,CO_2\,/\,kg$
Units	Total CO_2 kilograms emissions (kg CO_2)
Target value	Target value should be defined as an annual reduction target depending of the fleet renewal and other programs for fleet fuel efficiency improvements.
Calculus frequency	Monthly measurement and annual monitoring and assessment attending the indicator seasonality.
Additional comments	That parameter assess the environmental fleet impact and can be sorted by vehicles types and/or models and also be referred to mileage performed or travelers transported for future benchmarking activities.

4.3.3. Mean Time To Repair (MTTR)

Name	TR-MTTR
Definition	That indicator is computed in a discrete way adding all repair times (in hours or days) used on corrective maintenance and divided by the number of failures. Repair time have to consider parts unavailability time. Preventive, predictive or modification activities are not computed as a repair.
Formulae TR_1	$TR_1=\left(\frac{\sum TTR(h)}{Number\ of\ failures}\right)hours$
Units	Hours with two decimals (xx.xx h)
Target value	Attending that this parameter can be affected by so many factors such as: vehicle age, vehicle type, etc.; it is suggested to perform a trending analysis.
Calculus frequency	Monthly measurement and annual monitoring.
Additional comments	MTTR is a technical indicator for maintenance management and led us to obtain as estimation of vehicles maintainability sorting by models or vehicles types.

4.3.4. Maintenance delay

Name	RT- Maintenance delay
Definition	That indicator quantifies the delay regarding real preventive maintenance actions and the theoretical referred to the base reference period for preventive maintenance action expressed in terms of engine oil drain period.
Formulae TR_1	$RT_1=\left(\frac{\sum real\,mileage\,between\ preventive\,maintenance\,actions-\sum theoretical\,mileage}{\sum theoretical\,mileage}\right)x100$
Units	Percentage with one decimal (xx.x %)
Target value	Target value must be lower than 15% and could be optimal lower than 5%.
Calculus frequency	Monthly.
Additional comments	That indicator should be assessed in combination with other ones such as: maintenance fulfilment.

4.3.5. Optimal vehicle period replacement

Name	ER- Optimal vehicle period replacement
Definition	That indicator determines the optimal moment for a vehicle replacement using a Life Cycle Cost Analysis (LCC).
Formulae TR_1	For this indicator, calculus cannot be performed in just one equation. Authors suggest some bibliography for development. [7, 17]
Units	Years with two decimals (xx.xx years)
Target value	Value obtained by LLC analysis.
Calculus frequency	Monthly measurement and annual monitoring and evaluation.
Additional comments	That indicator must be obtained for each vehicle model on the fleet, attending that there are differences between them than can led to different optimum value for each model.

5. Balance scorecard proposal

Attending the previous parameter definition and some important comments, next it is presented a balance scorecard proposal for fleet companies (Table 6).

Table 6. Final structure for the BSC defined

Indicator	Target value	Actual Month	Previous Month	Monthly variation (%)	Assessment	Corrective actions
Structural						
KM						
ED						
VM						
VP						
PC						
Basic						
CM						
DS						
KA						
SN						
CM						
AB						
Advanced						
CE						
EM						
TR						
RT						
ER						

6. Conclusions

This work has presented a balanced scorecard approach for maintenance management in urban transport fleets. Although the BSC defined has not presented the KPIs explicitly in the classical classification attending: financial, costumers, environment, so on, they are presented in an implicit manner. Attending the modern society requirements for a sustainable mobility and the huge importance for that related with urban transport companies in cities, that approach is a contribution step for reaching expected targets. This proposal must be understood such a basis for a subsequent benchmarking approach based on the indicators proposed leading to exchange good practices and collaborations in areas of common interest between different urban transport companies both private and publics.

The final target of that work must be understood as defined by Wireman "performance measurements, when used properly, should highlight opportunities for improvement, detect problems and help find solutions" [26].

Acknowledgement

Authors want to acknowledgement to EMT de Valencia and other ATUC company members their collaboration and involvement on works performed during the project for KPIs definition and selection. Special thanks to Engineer Santiago Ballester for sharing efforts and knowledge to develop that work.

References

1. Abbas K. Developing optimum KPI system for Public Transport Organizations. Sigma Journal of Engineering and Natural Sciences 2016; 7 (1): 31-41.

2. Ahn H. Applying the Balanced Scorecard Concept: An Experience Report. Long Range Planning 2001; 34: 441-461, https://doi.org/10.1016/S0024-6301(01)00057-7.
3. Albert H, Tsang C, Jardine Andrew K S, Kolodny H. Measuring maintenance performance: a holistic approach. International Journal of Operations & Production Management 1999; 19 (7): 691-715, https://doi.org/10.1108/01443579910271674.
4. Alsyouf I. Measuring maintenance performance using a balanced scorecard approach. Journal of Quality in Maintenance Engineering 2006; 12 (2): 133-149, https://doi.org/10.1108/13552510610667165.
5. Bakhtiar A, Purwanggono B, Metasari N. Maintenance Function's Performance Evaluation Using Adapted Balanced Scorecard Model. International Journal of Industrial and Manufacturing Engineering 2009; 3(10): 1255-1259.
6. Crespo A, Moreu de León P, Gómez Fernández J F, Parra C, López Campos M. The maintenance management framework: A practical view to maintenance management. Journal of Quality in Maintenance Engineering 2009; 15 (2): 167-178, https://doi.org/10.1108/13552510910961110.
7. De Sa-Riechi JL, Macian V, Tormos B, Avila C. Optimal fleet replacement: A case study on a Spanish urban transport fleet. Journal of the Operational Research Society 2017; 68(8): 886-894, https://doi.org/10.1057/s41274-017-0236-1.
8. EN 13816: 2002 Transportation-Logistics and services – Public passenger transport service quality definition, targeting and measurement.
9. EN 15341: 2007 Maintenance - Maintenance Key Performance Indicators.
10. Field study on bus depots and bus maintenance. Similarities between and singularities of different engine technologies and their impact on infrastructure, quality and maintenance. UITP Bus Committee. 2013. Internal Report.
11. Fredrich L. Performance Metrics for the New Fleet Manager. Wheels & Wings - Online Newsletter for the Federal Motor Vehicle and Aviation Communities. 2014. Available at https://gsablogs.gsa.gov/wheelsandwings/2014/04/08/performance-metrics-for-the-new-fleet-manager/
12. Georgiadis G. The Role of Benchmarking in Public Transport: The Case of Thessaloniki, Greece. Procedia - Social and Behavioral Sciences 2012; 48: 2577 – 2587, https://doi.org/10.1016/j.sbspro.2012.06.1228.
13. Henning T, Dalil Essakali M, Eun Oh J. A framework for urban Transport benchmarking. The World Bank. 2011. Downloadable from https://openknowledge.worldbank.org/handle/10986/12847
14. Kaplan R S, Norton D P. The Balanced Scorecard – Measures that Drive Performance. Harvard Business Review 1992; 70(1): 71-79.
15. Kumar U, Galar D, Parida A, Stenström C, Berges L. Maintenance performance metrics: a state-of-the-art review. Journal of Quality in Maintenance Engineering 2013; 19 (3): 233-277, https://doi.org/10.1108/JQME-05-2013-0029.
16. Macián V, Tormos B, Ruiz S, Riechi J. Urban bus fleet maintenance costs: comparative analysis between diesel vs CNG fuelled vehicles. EuroMaintenance 2014 Congress Proceedings ISBN 978-952-67981-3-4. Helsinki (Finland), May 2014.
17. Macián V, Tormos B, Riechi J. Time replacement optimization model: comparative analysis of urban transport fleets using Monte Carlo Simulation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (2): 151-157, https://doi.org/10.17531/ein.2017.2.1.
18. Managing for Results in America's Great City Schools: A Report of the Performance Measurement and Benchmarking Project. 2017. Downloadable at https://www.cgcs.org/Page/660
19. MBTA Bus Maintenance Efficiency Study. Final Report. Massachusetts Bay Transportation Authority. March 2016. Downloadable from https://www.mbta.com/search?query=Bus%20Maintenance%20Efficiency%20Study&facets=&showmore=
20. Neely A, Gregory M, Platts K. Performance measurement system design: A literature review and research agenda. International Journal of Operations & Production Management 1995; 15 (4): 80-116, https://doi.org/10.1108/01443579510083622.
21. Parida A, Kumar U. Maintenance performance measurement (MPM): issues and challenges. Journal of Quality in Maintenance Engineering 2006, 12 (3); 239-251, https://doi.org/10.1108/13552510610685084.
22. Randall E, Condry B, Trompet M. International Bus System Benchmarking: Performance Measurement Development, Challenges, and Lessons Learned, Proceedings of the 86th Transportation Research Board Annual Meeting, January 2007, Washington.
23. Sondalini M. Useful Key Performance Indicators for Maintenance. Downloadable at https://www.lifetime-reliability.com/cms/free-articles/maintenance-management/
24. Stuart D. KPI's for maintenance: 15 suggestions for numbers that matter. 2016. Available at https://www.fleetowner.com/maintenance/kpis-maintenance-15-suggestions-numbers-matter
25. Tubis A, Werbińska-Wojciechowska S. Balanced Scorecard use in Passenger Transport Companies Performing at Polish Market. Procedia Engineering 2017; 187: 538-547, https://doi.org/10.1016/j.proeng.2017.04.412.
26. Wireman T. Developing Performance Indicators for Managing Maintenance. New York: Industrial Press, 1998.

Vicente MACIÁN
Bernardo TORMOS
Instituto Universitario CMT-Motores Térmicos
Universitat Politècnica de València
Campus de Vera. Edificio 6D. 46022, Valencia. Spain

Jesús HERRERO
ATUC- Asociación de Empresas Gestoras de los Transportes Urbanos Colectivos
C\ Princesa, 31 Piso 5 Oficina 1, 28008, Madrid. Spain

E-mails: vmacian@mot.upv.es, jherrero@atuc.es, betormos@mot.upv.es

Article citation info:
KSIAZKIEWICZ A, JANISZEWSKI J. Welding tendency for selected contact materials under different switching conditions. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 237–245, http://dx.doi.org/10.17531/ein.2019.2.7.

Andrzej KSIAZKIEWICZ
Jerzy JANISZEWSKI

WELDING TENDENCY FOR SELECTED CONTACT MATERIALS UNDER DIFFERENT SWITCHING CONDITIONS

TENDENCJA SCZEPIANIA WYBRANYCH MATERIAŁÓW STYKOWYCH W RÓŻNYCH WARUNKACH ŁĄCZENIOWYCH

The flow of significant current through electric contacts may lead to contact welding. In a.c. circuits this phenomena is not only dependent on properties of contact material (i.e. resistance to welding) but on the phase in which current is switched on. Welding tendency for contact materials made from AgNi, AgCdO and $AgSnO_2$ was evaluated based on selected phase at which make operation took place. The test circuit was protected by overcurrent apparatus to simulate real working environment. It is observed that welding tendency for the selected contact materials is contingent to current phase at which make operation is done.

***Keywords**: relays, contact materials, contact welding, surface erosion.*

Przepływ znacznego prądu przez styki elektryczne może prowadzić do ich sczepienia. W obwodach prądu przemiennego to zjawisko jest nie tylko zależne od właściwości materiału stykowego (tj. odporności na sczepianie), ale od fazy, w której prąd jest załączany. Zbadana została tendencja do sczepiania się styków, w zależności od fazy załączania prądu, wykonanych z następujących materiałów stykowych: AgNi, AgCdO i $AgSnO_2$. Obwód probierczy zabezpieczony był przez zabezpieczenia nadprądowe, w celu symulacji normalnych warunków pracy. Zauważono, że tendencja do sczepiania styków, dla wybranych materiałów stykowych, jest zależna od fazy, w której wykonuje się załączenie obwodu.

***Słowa kluczowe**: przekaźniki, materiały stykowe, sczepianie styków, erozja powierzchni.*

1. Introduction

The relays which are intended for connecting the electrical load are prone to some disadvantageous phenomena. These may include making of overload currents and short-circuit currents, which may lead to shortening the time of maintenance of relays or, in extreme cases, to their complete destruction. Long term exposure to higher temperature may lead to relays degradation, for example change of its contact resistance and opening and closing times [18, 19]. The research already performed by Morin [12], Neuhaus [13] and Doublet [4], who independently undertook work for similar contact materials, concentrate on low-current circuits of direct current and small amperage. There are also articles related to the processes of making circuits of alternating current of average voltage and amperage of several kA [1, 6, 14]. The operation of making significant currents may lead to contact bounces. Altogether with the increase of values for amperage of contacting current, there is the increase of loss of the mass of contact rivet [5, 14, 17]. Apart from that, discharge arc, accompanying commutation of electric circuits, may cause strong, local heating of an arc root, even above the temperature of melting for a certain contact material [7]. If at least on the surface of one contact the material is melted and the contacts close, then welding occurs [7, 9–11]. Primary properties of a contact material may change if the surface of contact or the composition of the material is changed [20]. The composition may change through thermal influence of current. Such conditions are true mostly for making currents of significant amperage, of several kiloamperes. The area of deformation for a contact surface depends on the resistance of a certain material to welding. The lower resistance, the bigger surface of contact is changed [20]. The article describes the influence of the short-circuit current phase on the welding tendency for selected contact materials, which are: AgNi, AgCdO and $AgSnO_2$. Research in this filed is focused mainly on d.c. circuits with low amperages, no exceeding hundreds of amps, or with a.c circuits and amperage level of several kA. Presented results are therefore new in this area as are focused on a.c. circuits and medium amps value with contact materials tested that are widely exploited in industry. Switching a damaged circuit will result in a short-circuit current flow which can be many times higher than the rated current of a relay. This current flow will generate significant heat generation in contacts that may result in contact welding. However the tendency for contact welding in such a.c. circuit is not only related to contact material but also to the phase of current at which the contacts are closed.

2. Contact welding tendency and force of the weld

Contact welding can occur inside the contact areas of the contact surface A_S. When the switching arc is present, it is assumed that the welding takes place inside the area where the A_m surfaces, melted due to the electric arc, meet. Each arc root spot is surrounded by a molten contact material, located completely or partially inside or outside of the contact surface. The pinching force of the contacts depends on the size of area A_0 [13, 16].

The welding force of the contacts also depends on the contact material used. Some contact materials show a higher tendency to welding than others [12]. If the contact material is characterized by a high welding tendency, also the resulting welds will form as strong. Pure silver has the worst properties in this respect, it has a higher tendency for welding [8] and this is one of the reasons why it is not used as a contact material. AgCdO is characterized by slightly better param-

eters. It has a lower tendency for welding, but they create welds with higher strength.

Contact materials $AgSnO_2$ and AgNi, used in the tested relays, have similar properties in the range of welding tendency and their strength. In this case, the first of these forms stronger joints, but exhibits a lower tendency to form them. The influence on the welding force will also have the type of the load. A resistive-inductive load leads to welds with lower weld force than with a purely resistive load. The greatest welding forces occur with the resistive-capacitive load [13, 16]. In the case under consideration, i.e. the short circuit, the load has a resistive character. The inclusion of a current of higher intensity leads to the welding of the contacts of higher weld strength. The shorter duration of the contact bounce leads to the formation of welds with a greater force than the bounces of a longer duration. According to Rieder and Neuhaus [15] the explanation of this process is related to different contact bounce heights. With a shorter duration of the bounce on the contact surface, a deep arc spot appears with a smaller surface than with a longer bounce. When a bounce occurs for a longer duration, the contact surface is melted, although only on the surface itself, which leads to the formation of a weld of lower strength. Welds with significant force occur with bounces lasting less than 100 μs. The contact bounces of a longer duration lead to splashing the molten contact material away from the contact point [2]. Increasing the content of oxides in the contact material ($AgSnO_2$ In) reduces the strength of welds. Contact welding will be more frequent with time, and they will produce stronger welds [3].

3. Materials and methods

The most difficult working conditions for electromagnetic relays can occur at the moment of closing the damaged circuit in which a short-circuit current occurs. With significant short-circuit currents, processes such as contact heating above the melting point of the material, contact bounces and also welding may occur. Attempts have been made to connect a short circuit with the expected short-circuit current of 320 A. Contact materials that were tested are: AgNi, AgCdO and $AgSnO_2$. Each of them was composed of 90 % silver and a 10 % addition of nickel, cadmium-oxide and tin-oxide respectively. The $AgSnO_2$ was tested in two variations. For the first one contact rivets were made in the process of internal oxidation and refereed in the article simply as $AgSnO_2$. For the second one the rivet was designed to withstand higher inrush currents (up to 80 A for 20 ms) and refereed in the article as $AgSnO_2$ P. The system is designed to test short-circuit currents with prospective values up to $20I_n$ in circuits equipped with typical overcurrent protection apparatus, used in low voltage electrical installations, with rated current 16 A. In addition, the system is equipped with an external device synchronizing the moment of relay activation with the phase of the supply voltage. This system is intended for switching the relay contacts in the selected phase of the mains voltage. Measurements were carried out for two switching on currents:

- switching on the short-circuit current occurs at the moment when the voltage between the contacts reaches zero (case A),
- switching on short-circuit current occurs at the moment when the voltage between the reaches the maximum value (case B).

For case A short-circuit current increases from zero, and in case B shortcircuit current starts from the maximum value. The effects of such switching on the short-circuit current will have an impact on the electrode processes occurring in the contacts during the short-circuit. Switching on the circuit at the moment when both current and voltage reach zero, puts milder conditions for the operation of the relay. There shouldn't be a preliminary electric discharge between the contacts, thus the contact surface may undergo less erosion. The short-circuit current increase is slower, so that electrodynamic contact bounces may be shorter or won't occur at all. Switching on the circuit at the moment when both voltage and current reach peak value can cause drastically different effects. A sudden increase in the short circuit current value may lead to the appearance of significant electrodynamic forces in the initial phase of the closing process. In consequence, this will lead to the occurrence of contact bounce during the current flow and thus to the ignition of the electric arc. This arc will contribute to further significant erosion of the contact surface.

4. Results

An example oscillogram of short-circuit current and the voltage between the relay contacts when the short-circuit circuit is closed when the voltage reaches zero is shown in Fig. 1. The graph shows that when the supply voltage reaches zero, the short-circuit current begins to flow. It can be noticed that after about a time of 4 ms since the beginning of the current flow, the current decreases slightly, and after the time 5 ms there is a significant reduction in its values. This limitation is related to the effect of overcurrent protection in the circuit. The interruption of the fault circuit occurred after less than 7 ms since the occurrence of a short circuit.

Eight attempts were made for each contact material. The current and voltage between the relay contacts were recorded. After each test, the contacts were checked for welding. On the basis of the oscilloscope recording, it was determined whether in the given test a contact bounce occurred. A summary of the results obtained are presented in table 1. It can be seen that for none of the attempts contact bounce b or welding s was observed. This leads to the conclusion that for switching the circuit at the moment when the current increases from zero value does not cause negative effects related to the relay rivets.

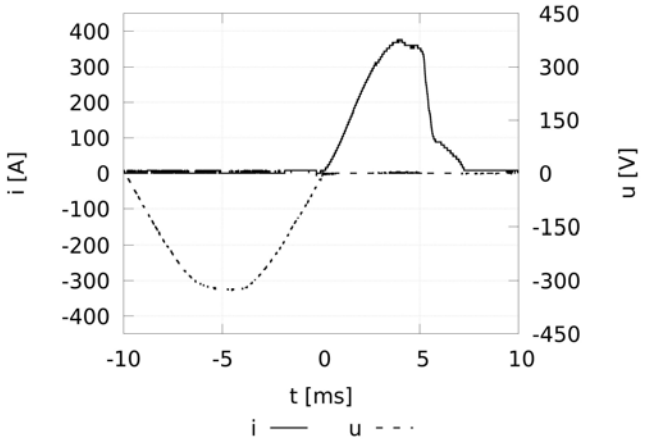


Fig. 1. An exemplary oscillogram of current and voltage waveform between the relay contacts when the circuit is closed when voltage between contacts reaches zero (case A); contact material – AgNi

Table 1. The results of switching the circuit when current reaches zero (case A); n – number of tries, s – number of welds, b – number of contact bounces, s∧b – simultaneous occurrence of contact bounce and contact weld

Lp.	Contact material	n	s	b	s ∧b
1	AgNi	8	0	0	0
2	$AgSnO_2$	8	0	0	0
3	$AgSnO_2$ P	8	0	0	0
4	AgCdO	8	0	0	0

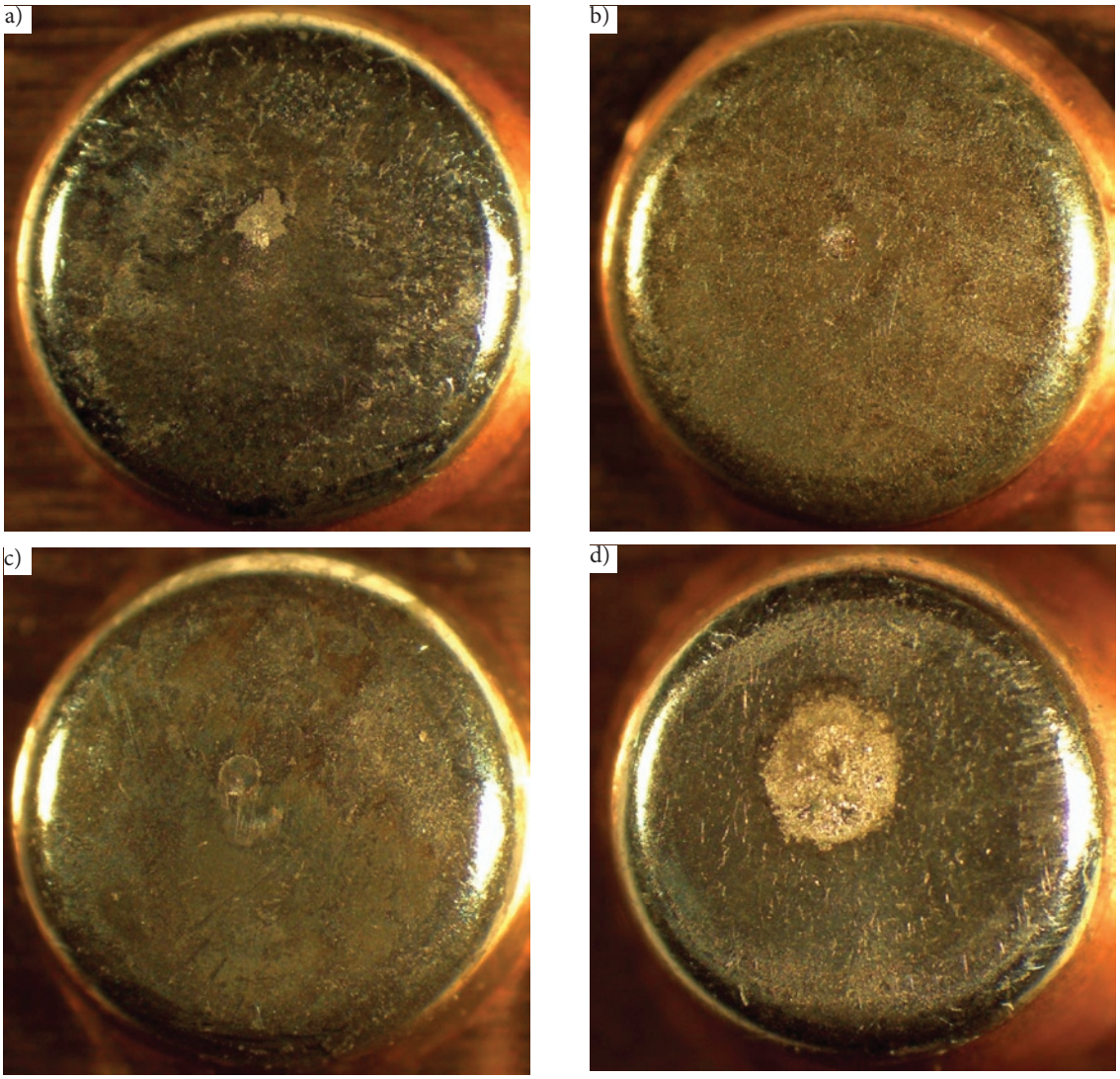


Fig. 2. Contact surface after short-circuit current switched on at zero value (case A), for individual contact materials; a) AgNi, b) $AgSnO_2$, c) $AgSnO_2$ P, d) AgCdO

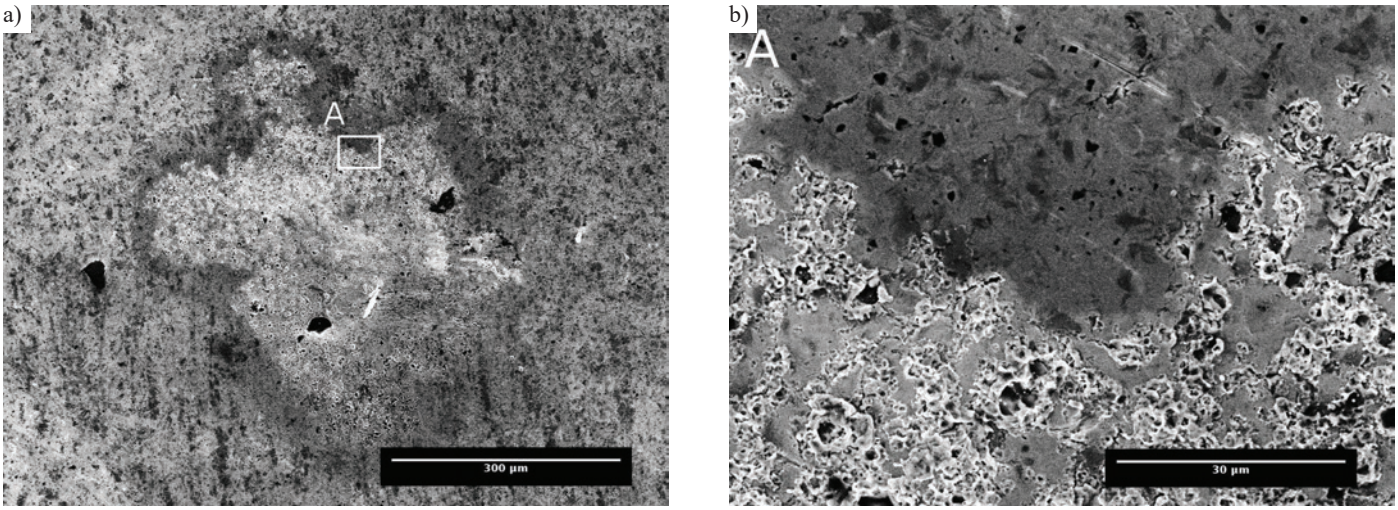


Fig. 3. View of the contact surface under the electron microscope, when the short-circuit current is switched at zero (case A), for contacts made of AgNi

Photographs of the contact surfaces are shown in the Fig. 2. A single erosive contact area is visible on the contact surface. This area has a shape that is clearly similar to circular. Each time it was found that the degraded surface is located almost centrally on the contact, with no visible discolouration. On figures 3, 4, 5 and 6 the view of the eroded areas under the scanning microscope is shown, successively for the contact materials AgNi, $AgSnO_2$, $AgSnO_2$ P and AgCdO. All the eroded surfaces have a distinct round shape, except for the AgNi. Individual contact areas are exposed, where erosive changes in contact surfaces occurred. The sizes of the eroded areas are shown in

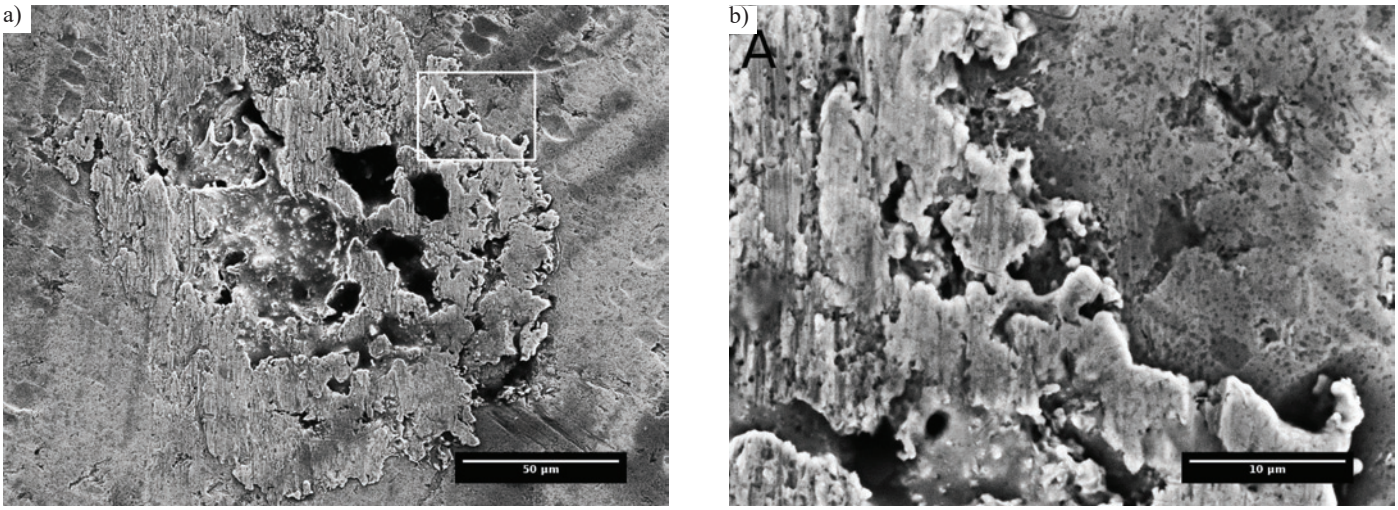


Fig. 4. View of the contact surface under the electron microscope, when the short-circuit current is switched at zero (case A), for contacts made of $AgSnO_2$

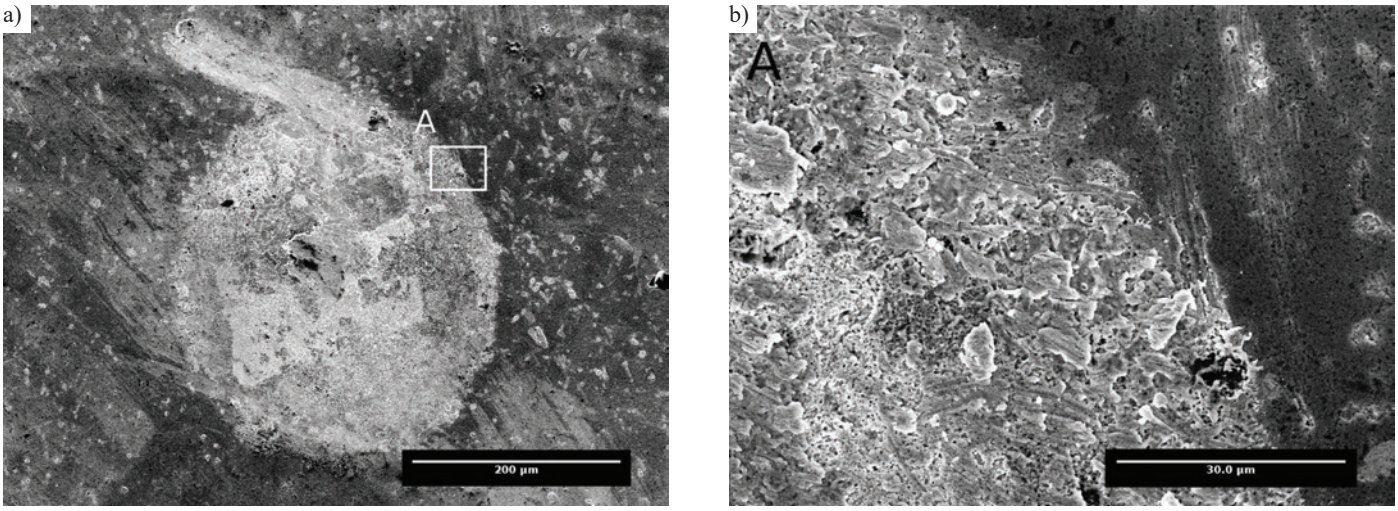


Fig. 5. View of the contact surface under the electron microscope, when the short-circuit current is switched at zero (case A), for contacts made of $AgSnO_2$ P

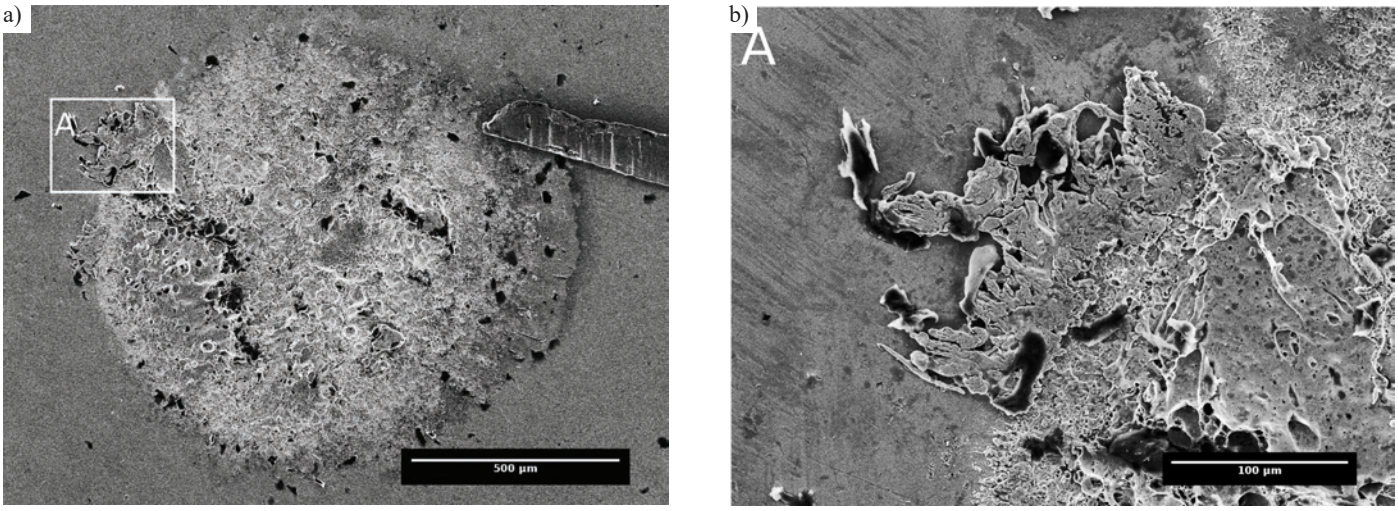


Fig. 6. iew of the contact surface under the electron microscope, when the short-circuit current is switched at zero (case A), for contacts made of AgCdO

table 2. Clearly visible is the boundary between the eroded and unchanged contact surface. For AgNi, the visible change is characterized by a greater graininess of the eroded contact material. For other materials, these changes are characterized by the occurrence of larger flat surfaces. In each case, there are no noticeable splashes of the contact material. This means that there was no melt discharge when closing the contacts. It was eroded only after they were closed under the influence of high temperature generated by the current flow. An exemplary oscillogram of short-circuit current and voltage between the relay contacts at the moment when the voltage reaches the maximal value is shown in Fig. 7. After closing the contacts of the relay, the current starts flowing and reaches a peak value of 432 A after 1 ms.

Table 2. The size of eroded contact surface when switching on the short-circuit current at zero value (case A)

Size of the eroded contact surface [mm^2]			
AgNi	$AgSnO_2$	$AgSnO_2$ P	AgCdO
0,13	0,017	0,093	0,864

The significant steepness of the current rise in the circuit results from the negligible value of the inductive reactance of the circuit. At the same time, the process of switching off the current by short-circuit protection apparatus begins. When the current passes through zero, the current flow is interrupted.

In the same way as in the previous case, each time the contact was checked and whether or not a bounce occurred during the time of the given test. The results are presented in the table 3. Almost for every registered sample there was a contact bounce. This bounce, as already discussed earlier, can be a dominant contributor to the welding process. In the scope of the short-circuit current switching experiments with the expected value of 320 A, when closing the relay contacts in the maximum supply voltage, the tested contact materials showed their different susceptibility to welding. They were noted for two contact materials: AgNi and AgCdO. The other two materials showed higher resistance to welding despite contact bounces. Therefore, one

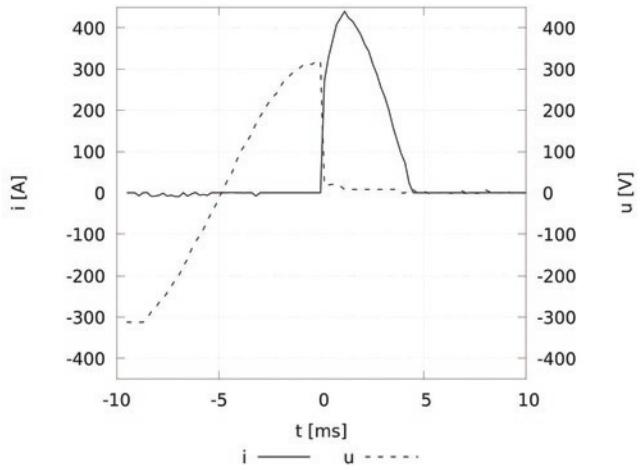


Fig. 7. Exemplary current and voltage waveform between the relay contacts when the short circuit circuit is switched on at peak voltage (case B), contact material – $AgSnO_2$ P

should also pay attention to the convergence of these two processes (s ∧ b) as shown in Tab. 3. The appearance of the bounce almost or each trial ended with contact weld. Although the contacts made with AgNi should show a higher resistance [12], according to the presented results it did not. It has been shown that in the unfavourable switching

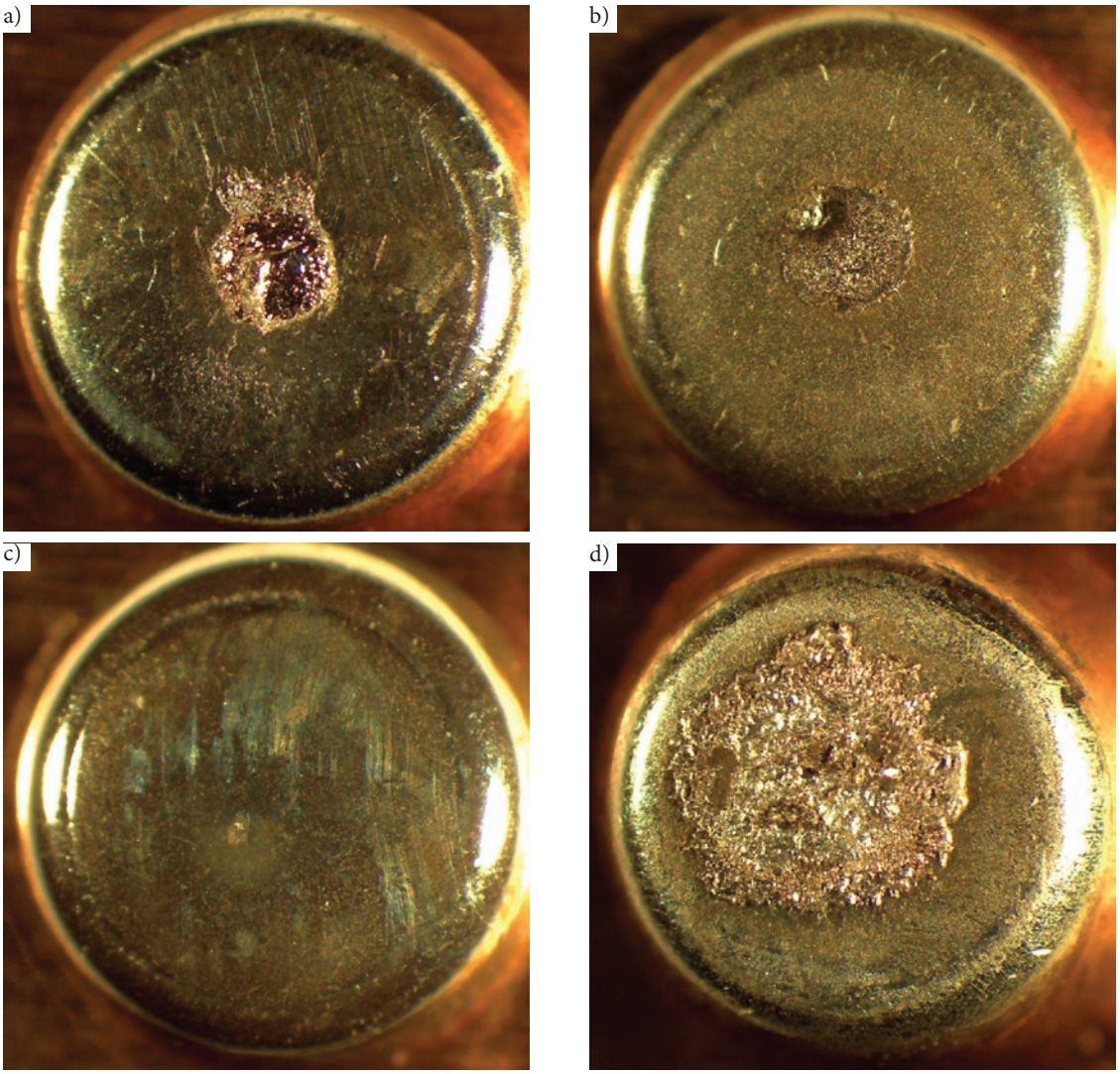


Fig. 8. Contact surface after short-circuit current switched on at peak value (case B), for individual contact materials: a) AgNi, b) $AgSnO_2$, c) $AgSnO_2$ P, d) AgCdO

Table 3. The results of switching the circuit when current reaches peak value (case B); n – number of tries, s – number of welds, b – number of contact bounces, s ∧ b – simultaneous occurrence of contact bounce and contact weld

Lp.	Contact material	n	s	b	s ∧b
1	AgNi	8	7	7	7
2	$AgSnO_2$	8	0	5	0
3	$AgSnO_2$ P	8	0	6	0
4	AgCdO	8	8	4	4

conditions that occur in operation - switching on damaged circuits, the contacts made of AgNi and AgCdO are characterized by low resistance to the welding process. Photographs of the contact surfaces are shown in Fig. 8. Clearly very different erosion of the contact surfaces can be observed (Tab. 4). In particular, this applies to the contact materials for which the welding process has been observed. For them, erosive changes relate to a significant part of the surface. The smallest surface changes occurred for $AgSnO_2$ P. For this material (and $AgSnO_2$), no welds were observed, despite the occurrence of bounces.

In the figures 9, 10, 11 and 12, photographs of eroded areas made with scanning microscope are presented, successively for contact materials AgNi, $AgSnO_2$, $AgSnO_2$ P and AgCdO. Due to thermal processes occurring during the switch-on time (mainly due to the electric arc) part of the contact material has been melted. When closing the contacts, this material is thrown out of the contact area. It can be seen that the direction of ejection of this material is centrifugal from the point of contact. The ejection range of the molten material reaches, in the extreme case, for AgNi even 300 μm counting from the border of the periphery of the region in the shape of a circle. For $AgSnO_2$ P, changes of a similar nature are virtually imperceptible.

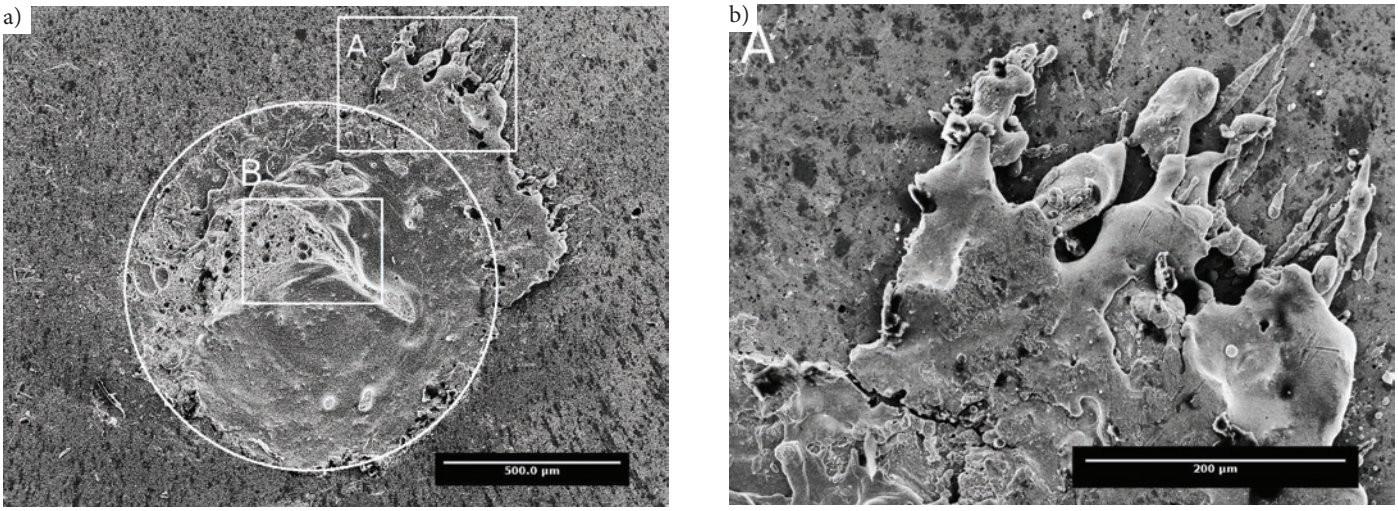


Fig. 9. View of the contact surface under the electron microscope, when the short-circuit current is switched at peak value (case B); contact material AgNi

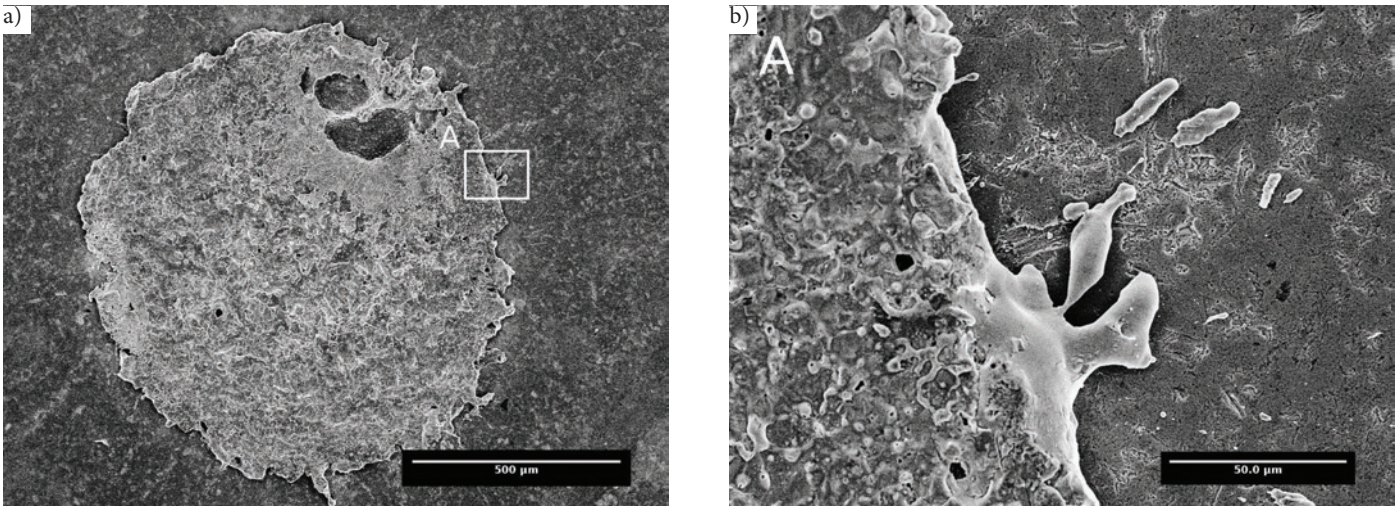


Fig. 10. View of the contact surface under the electron microscope, when the short-circuit current is switched at peak value (case B); contact material $AgSnO_2$

Table 4. The size of eroded contact surface when switching on the short-circuit current at peak value (case B)

Size of the eroded contact surface [mm^2]			
AgNi	$AgSnO_2$	$AgSnO_2$ P	AgCdO
1,029	0,838	0,022	3,525

In figure 13 the state of the contact surface is presented, for a selected case in which the contacts have been welded. Two areas clearly stand out. In the middle of the figures, porous changes are visible on the contact surfaces. These are the places where the contacts have interfered and then mechanically broken. These surfaces are characterized by an elevation (convexity) in relation to the base of contact surface. Both for AgNi and AgCdO the areas of torn bonding joints

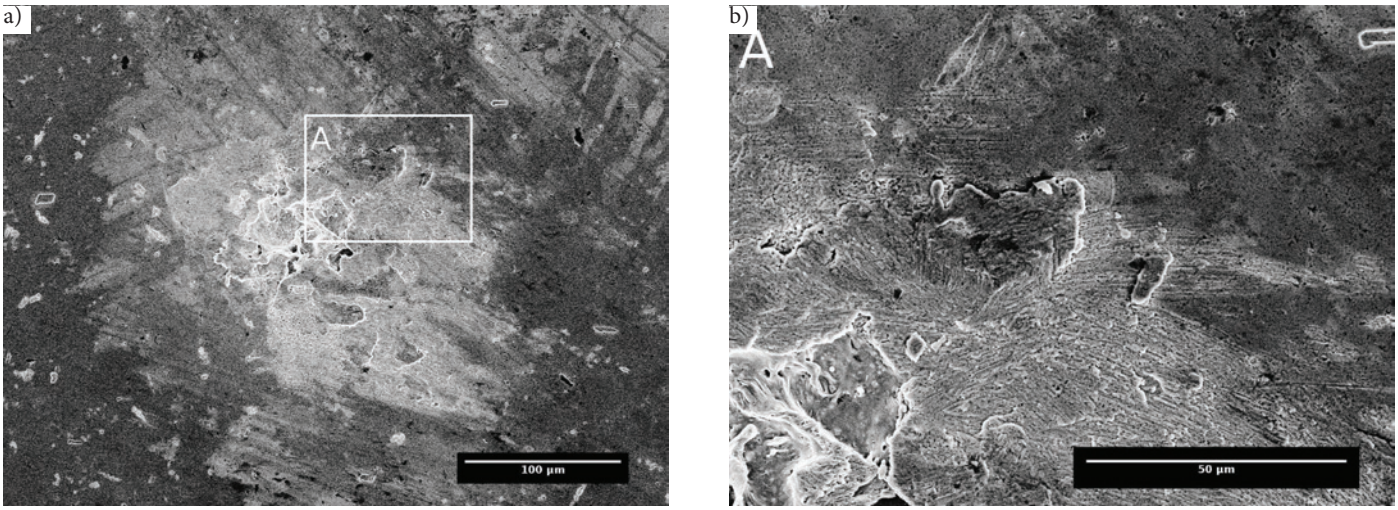


Fig. 11. View of the contact surface under the electron microscope, when the short-circuit current is switched at peak value (case B); contact material $AgSnO_2$ P

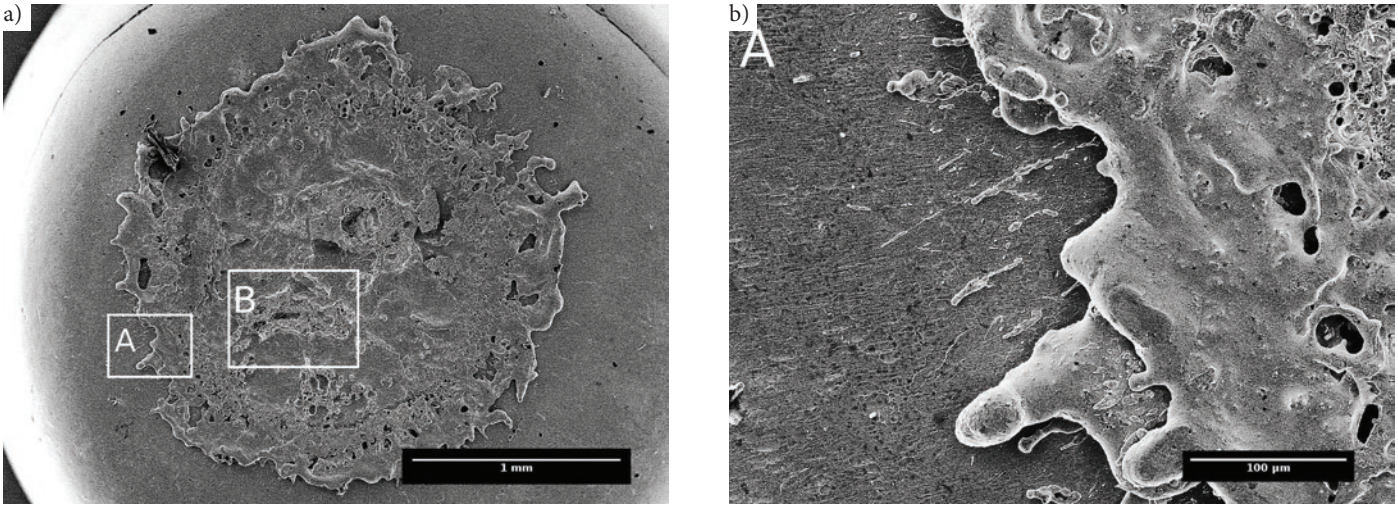


Fig. 12. View of the contact surface under the electron microscope, when the short-circuit current is switched at peak value (case B); contact material AgCdO

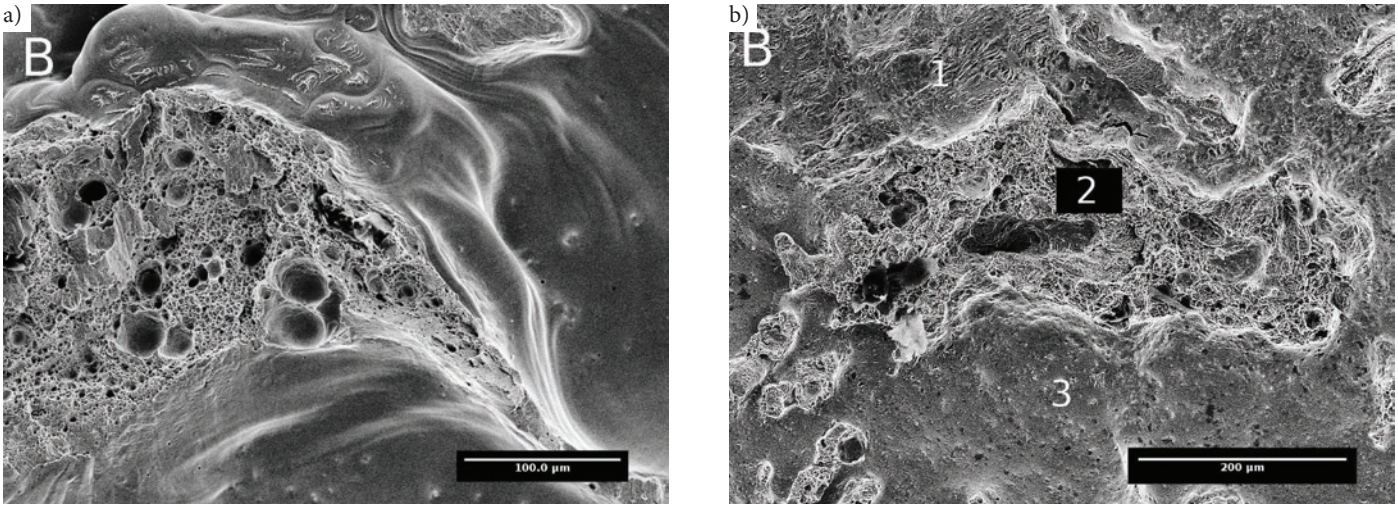


Fig. 13. Photo of the contact surface at the point where the contacts are welded: a) AgNi, b) AgCdO, test current I =320 A

have length from 500 µm and width 200 µm, and their area is approximately 0.107 mm^2 and 0.09 mm^2. For the AgNi material it is seen that the melted material flows from the area where the weld has occurred. For AgCdO, there are three levels of the contact area, where the 1 area is higher and 3 lower than the area 2. The boundary of material disruption and their breakthrough is plastic-brittle.

5. Conclusions

Based on these results, it can be concluded that the most resistant to welding ($AgSnO_2$) [12] was characterized by the smallest deformation of the contact surface. In the next order is AgNi, and the biggest changes are observed for AgCdO. The melting point of these materials, according to the manufacturer's data, is identical (961 °C). The differences occur in the case of thermal conductivity. For $AgSnO_2$ the coefficient of thermal conductivity is not described by the manufacturer, but for AgCdO it is by 12 % smaller than for AgNi, hence the susceptibility of this material to the effects of thermal interactions can be greater. This may justify the observed increase in surface degradation at the AgCdO contacts.

In the analysed case, the level of local destruction of the contact surface will be significantly affected by the value of thermal conductivity, and not by the hardness or melting temperature (which are almost identical) of the contact material. Higher thermal conductivity means better heat transfer from its source (contact area) deeper into the material. Thus, with better conductivity, the heating in the contact zone, up to the melting point, will take longer or will require more energy (higher current). In the discussed experimental conditions, the worst conditions occur for the material with the least thermal conductivity.

Contacts bounces occurred only when the current was switched on in maximum voltage. Switching on the circuit at the moment when the voltage (and current) reaches the value of zero does not cause unambiguously negative effects on the contact surfaces of the tested relays. In the range of the applied test currents, welding occurred only for contacts made with AgNi and AgCdO and only for switching the circuit when the current reached its peak value (case B). One should also pay attention to the convergence of the occurrence of contact bounce and its welding. The appearance of the bounce almost for each trial ended with contact weld. Thus, contacts made with AgNi and AgCdO are characterized by low resistance to the welding process. On the contact surfaces, a single contact area is visible, with the shape of a circle that has eroded. This area is located almost in the middle of the contact.

There is no visible discolouration of the unevened contact surface. The material most resistant to welding ($AgSnO_2$) was characterized by the smallest erosion of the contact surface, higher erosion is visible for AgNi and AgCdO respectively. When switching on the voltage, splashes of the contact material are not noticeable. This means that the erosion of the contact area occurred under the influence of its heating without the participation of arc. For switching on the maximum voltage, a significant part of the contact material is melted and can be ejected out of the contact area. The direction of ejection of this material is radiant from the point of contact.

Acknowledgement

The research was financed from resources of the Ministry of Science and Higher Education for statutory activities No. 04/41/DSMK/4133, name of the task: Switching and electrode processes in low voltage relays – the effect of switching on nominal and short-circuit currents.

References

1. Borkowski P. Arc erosion of contacts on switching high currents. Archives of Electrical Engineering 2004; 53(209): 259–287.
2. Chen Z, Witter G. Dynamic welding of silver contacts under different mechanical bounce conditions. 45th IEEE HOLM Conference on Electrical Contacts 1999: 1–8, https://doi.org/10.1109/HOLM.1999.795920.
3. Chen Z K, Witter G J. A study of dynamic welding of electrical contacts with emphasis on the effects of oxide content for silver tin indium oxide contacts. 56th IEEE Holm Conference on Electrical Contacts 2010: 121–126, https://doi.org/10.1109/HOLM.2010.5619552.
4. Doublet L, Jemaa N B, Hauner F, Jeannot D. Make arc erosion and welding tendency under 42 VDC in automotive area. 49th IEEE Holm Conference on Electrical Contacts 2003: 158–162.
5. Kharin S, Sarsengeldin M. Influence of contact materials on phenomena in a short electrical arc. Key Engineering Materials 2012; 1: 321–329, https://doi.org/10.4028/www.scientific.net/KEM.510-511.321.
6. Kolimas S. Analysis of the making arcing time during current switching. Przegląd Elektrotechniczny 2004; 84:72–74.
7. Ksiazkiewicz A, Janiszewski J, Batura R. Influence of short-circuit ac currents on electrical contact resistance of low voltage relays. Poznan University of Technology Academic Journals. Electrical Engineering 2012; 70: 99-103.
8. Leung C, Streicher E. Material transfer in dynamic welding of ag and ag/sno2 contact material. 48th IEEE Holm Conference on Electrical Contacts 2002: 21–28.
9. McBride J. An experimental investigation of contact bounce in medium duty contacts. IEEE Transactions on Components, Hybrids, and Manufacturing Technology 1991; 14(2):319–326, https://doi.org/10.1109/33.87312.
10. McBride J W. Electrical contact bounce in medium duty contacts. 34th IEEE Holm Conference on Electrical Contacts 1988: 141–149, https://doi.org/10.1109/HOLM.1988.16110.
11. McBride J, W Sharkh S M. Electrical contact phenomena during impact. IEEE Transactions on Components, Hybrids, and Manufacturing Technology 1992; 15(2):184–192, https://doi.org/10.1109/33.142893.
12. Morin L, Jemaa N B, and Jeannot D. Make arc erosion and welding in the automotive area. IEEE Transactions on Components and Packaging Technologies 2000; 23(2): 240–246, https://doi.org/10.1109/6144.846760.
13. Neuhaus A, Rieder W, Hammerschmidt M. Influence of electrical and mechanical parameters on contact welding in low power switches. IEEE Transactions on Components and Packaging Technologies 2004; 27(1): 4–11, https://doi.org/10.1109/TCAPT.2004.825777.
14. Pons F, Cherkaoui M. An electrical arc erosion model valid for high current: Vaporization and splash erosion. 54th IEEE Holm Conference on Electrical Contacts 2008: 9–14, https://doi.org/10.1109/HOLM.2008.ECP.15.
15. Rieder W, Neuhaus A. Contact welding influenced by anode arc and cathode arc, respectively. Proceedings of the 50th IEEE Holm Conference on Electrical Contacts and the 22nd International Conference on Electrical Contacts 2004: 378–381, https://doi.org/10.1109/HOLM.2004.1353144.
16. Rieder W F, Neuhaus A R. Short arc modes determining both contact welding and material transfer. Short arc modes determining both contact welding and material transfer 2007; 30(1): 9–14.
17. Walczuk W, Borkowski P, Ksiarek S, Missol W, Rdzawski Z, Durst K. Evaluation of basic electrical parameters of silver-based contact materials of different chemical composition and manufacturing technology. 56th IEEE Holm Conference on Electrical Contacts 2010: 134–141, https://doi.org/10.1109/HOLM.2010.5619550.
18. Wang Z, Huang Z, Wang J, Shang S, Zhai G, The failure mechanism of electromagnetic relay in accelerated storage degradation testing.

IEEE Holm Conference on Electrical Contacts 2017: 164–168, https://doi.org/10.1109/HOLM.2017.8088080.
19. Wang Z, Shang S, Wang J, Huang Z, Sai F. Accelerated storage degradation testing and failure mechanisms of aerospace electromagnetic relay. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19(4): 530–541, https://doi.org/10.17531/ein.2017.4.6.
20. Zhenbiao L, Lichun C, Jiyan Z. The metallurgical research on contact surface deterioration of agni, agw, agfe, agcu contact materials. 41st IEEE Holm Conference on Electrical Contacts 1995: 346– 349, https://doi.org/10.1109/HOLM.1995.482890.

Andrzej KSIAZKIEWICZ
Jerzy JANISZEWSKI
Poznan University of Technology
5 M. Sklodowska-Curie Square
60-965 Poznan, Poland

e-mail: andrzej.ksiazkiewicz@put.poznan.pl,
jerzy.janiszewski@put.poznan.pl

Article citation info:
LI J, WANG Z, LIU C, QIU M. Accelerated degradation analysis based on a random-effect Wiener process with one-order autoregressive errors. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 246–255, http://dx.doi.org/10.17531/ein.2019.2.8.

Junxing LI
Zhihua WANG
Chengrui LIU
Ming QIU

ACCELERATED DEGRADATION ANALYSIS BASED ON A RANDOM-EFFECT WIENER PROCESS WITH ONE-ORDER AUTOREGRESSIVE ERRORS

PRZYSPIESZONA ANALIZA DEGRADACJI W OPARCIU O PROCES WIENERA Z EFEKTEM LOSOWYM Z BŁĘDAMI AUTOREGRESYJNYMI PIERWSZEGO RZĘDU

For highly reliable and long-life products, accelerated degradation test (ADT) is often an effective and attractive way to assess the reliability. To analyze the accelerated degradation data, it has been well recognized that it is necessary to incorporate three sources of variability including the temporal variability, the unit-to-unit variability and measurement errors into the ADT model. The temporal variability can be properly described by the Wiener process. However, the randomness of the initial degradation level, which is an important part of the unit-to-unit variability, has been often neglected. In addition, regarding the measurement errors, current ADT models often assumed them to follow a mutually independent normal distribution and ignored the autocorrelation that may probably exist in them. These problems lead to a poor accuracy for reliability evaluation in some situation. Thus, a random-effect Wiener process-based ADT model considering one-order autoregressive (AR(1)) errors is proposed. Then closed-form expressions for the failure time distribution (FTD) is derived based on the concept of first hitting time (FHT). A statistical inference method is adopted to estimate unknown parameters. Finally, a comprehensive simulation study and a practical application are given to demonstrate the rationality and effectiveness of the proposed model.

Keywords: *reliability evaluation, accelerated degradation modeling, Wiener process, unit-to-unit variability, measurement errors.*

W przypadku wysoce niezawodnych produktów o długim cyklu życia, przyspieszone badanie degradacji (ADT) często stanowi skuteczny i atrakcyjny sposób oceny niezawodności. Jak wiadomo, analiza danych z przyspieszonej degradacji wymaga włączenia do modelu ADT trzech źródeł zmienności, w tym zmienności czasowej, zmienności między jednostkami i błędów pomiarowych. Zmienność czasową można odpowiednio opisać za pomocą procesu Wienera. Jednak losowość początkowego poziomu degradacji, który stanowi ważną część zmienności między jednostkami, jest często w badaniach pomijana. Ponadto, w odniesieniu do błędów pomiaru, obecne modele ADT często zakładają, że mają one wzajemnie niezależne rozkłady normalne, ignorując możliwą autokorelację. Problemy te prowadzą w niektórych sytuacjach do niskiej trafności oceny niezawodności. W związku z powyższym, zaproponowano model ADT oparty na procesie Wienera z efektem losowym, w którym uwzględniono błędy autoregresyjne pierwszego rzędu (AR (1)). Następnie, w oparciu o pojęcie pierwszego czasu przejścia, wyprowadzono wyrażenia w postaci zamkniętej dla rozkładu czasu uszkodzenia (FTD). Do oszacowania nieznanych parametrów przyjęto metodę wnioskowania statystycznego. Na koniec przedstawiono kompleksowe studium symulacyjne i wskazano praktyczne zastosowanie modelu w celu wykazania jego racjonalności i skuteczności.

Słowa kluczowe: *ocena niezawodności, przyspieszone modelowanie degradacji, proces Wienera, zmienność między jednostkami, błędy pomiaru.*

1. Introduction

With increasing requirements from customers, more and more products are requested to have long life and high reliability. For products with these features, degradation data has been recognized as a valuable life information source and has been commonly adopted in reliability assessment [1, 26]. To guarantee the analysis accuracy, it is necessary to construct a reasonable degradation model. In the literature, many real applications suggest that degradation of a batch of products is usually affected by three types of variability including temporal variability (also defined as time-correlated structure), unit-to-unit variability and measurement errors. The temporal variability is referred to as the inherent variation of the degradation process over time. The unit-to-unit variability describes the heterogeneity among the degradation paths of multiple items. Measurement errors are usually created during the degradation investigation process due to imperfect measurements. For example, the imperfect measuring tool, the randomness of environmental factors and lower-skilled technicians may all result in imperfect measurements, especially when the data are obtained in an indirect way [19, 27, 28].

There is considerable interest on the part of the scientists and engineers in understanding and modeling the degradation process of products and components. Simultaneously considering the three types of variability, Peng and Tseng [19] proposed a general linear degradation model, Ye et al. [31] presented a well-adopted Wiener process degradation model and Li et al. [6] constructed a generalized nonlinear Wiener process-based degradation model. Meanwhile, Pan et al. [19] developed a reliability estimation approach based on EM algorithm for Wiener process degradation model by simultaneously

considering the variabilities. Moreover, degradation models adopted for the remaining useful life prediction that simultaneously considering the above three types of variability can be seen in [22, 23, 35].

In practical engineering, ADT has been recognized as an effective way to obtain degradation information quickly and efficiently within a reasonable time span and budget [31]. For accelerated degradation processes, temporal variability, unit-to-unit variability and measurement errors have also been recognized as three main uncertainty sources. Therefore, it is necessary to incorporate the three types of variability simultaneously for reasonable ADT modeling. Since the temporal variability of a degradation process can be properly described by a stochastic process via its the stochastic characteristics, stochastic process-based models including Wiener process, Gamma process and Inverse Gaussian Process have been in favor with many researchers in ADT analysis [11]. Among them, due to the attractive mathematical properties and physical interpretations, Wiener process and its various variants have been extensively developed and applied for accelerated degradation analysis of products [33-34]. Whitmore and Schenkelberg [30] adopted a Wiener process with one time scale transformation to model a constant stress ADT (CSADT).Liao and Tseng [10] provided an optimal step-stress ADT (SSADT) plan based on a Wiener process through a time transformation. Lim and Yum [11] developed an optimal ADT plan by assuming that the degradation characteristic follows a Wiener process.

Regarding unit-to-unit variability in ADT modeling, the degradation rate has been considered as a unit specific property. Then the drift coefficient (denoting the degradation rate) in Wiener process models is usually supposed to be random variable to describe the heterogeneity among test specimens. Tang et al. [25] proposed a nonlinear Wiener process to model ADT where the variability of an individual specimen was considered by assuming the drift coefficient as a random variable. Sun et al. [24] considered the individual variation by regarding the drift parameter as random in the Wiener process ADT model. Liu et al. [14] proposed a general Wiener process ADT model considering the unit-to-unit uncertainty. Meanwhile, measurement errors have been incorporated in Wiener process ADT modeling [4, 7].

According to the literature review, although a few Wiener process models have considered three types of variability into the ADT modeling, multiple problems have to be settled to enhance the model reasonability and the analysis accuracy. According to the best of our knowledge, it is a standard assumption for Wiener process-based ADT models that all measurement error terms are mutually independent and follow a normal distribution with zero mean and equal variance in the current literature. In practice, however, it is an oversimplification to suppose measurement errors are mutually independent. Degradation measurements (comprising the true degradation and measurement errors) are observed on a unit over time, and then it is reasonable to consider the degradation observation sequence and its measurement error dataset as time series [3, 17]. It is well known that time series datasets usually exhibit autocorrelation because of modeling errors or cyclic changes in ambient conditions (e.g., temperature)[5, 13]. Therefore, autocorrelation may probably exist in measurement errors, and it may be nonnegligible in many practical situations. It is further worth noticing that the autocorrelation may become stronger when the inspection time interval is relatively short. A Wiener process degradation model with AR(1) measurement errors for general degradation analysis was proposed in our previous work [9]. However, regarding accelerated degradation reliability analysis, the issue has not been reported in the literature. Thus, the first main objective of this paper is to extend the degradation model subject to autoregressive measurement errors in [9] to the case of accelerated degradation analysis.

On the other hand，in ADT modeling literature, the drift parameter has been usually supposed as a random variable to consider the unit-to-unit variability, while the initial degradation value has been always assumed or transformed as zero or a constant for most current Wiener process models [14, 25]. In real applications, however, the initial degradation level may probably not be a fixed value across all items, and may show unit-to-unit variabilities [4]. For example, the difference of the initial degradation level may be caused by the variant properties of material, the geometry differences of products, and so on. Therefore, it is necessary to incorporate unit-to-unit variabilities of both degradation rate and initial degradation level into the ADT modeling procedure to enhance its reasonability. That is the second objective of the current study.

In addition, it is well known that FHT of a standard Wiener process follows an inverse Gaussian distribution, and this is very useful for reliability analysis and maintenance decision-making [2]. For general degradation model and ADT model based on Wiener process, closed- form FTD expressions have been derived for situations when unit-to-unit variability regarding degradation rate is considered in the literature. As previously discussed, it is necessary for a reasonable ADT model to consider the unit specific variability caused by initial degradation level simultaneously. Considering the degradation models subject to measurement errors, most studies assume that a product is considered as been failed when its true degradation level first hits a predefined critical level [19, 35]. Therefore, the current study derives the closed-form ADT expressions along this line.

The remainder of the paper is organized as follows. Section 2 introduces an improved Wiener process ADT model which can consider the autoregressive measurement errors and the unit specific properties of both degradation rate and initial degradation level. The lifetime distribution is derived based on the FHT concept. Section 3 discusses MLEs for model parameters and an initial guesses method for optimization algorithm is given. In Section 4, the efficiency and reasonability of the established methodology is validated via a comprehensive Monte Carlo simulation study. In Section 5, the proposed approach is illustrated by a real application involving an electronic transistor ADT and comparative results are given. A summary and conclusion is given in Section 6.

2. ADT modeling

According to stress loading modes, there are mainly three ADT types including CSADT, SSADT and progressive stress ADT (PSADT). In real applications, comparing with SSADT and PSADT, CSADT has been recognized as the most commonly adopted type because of its simplicity and conveniences. Therefore, CSADT is focused in the current study.

2.1. Model formulation

A random-effect Wiener process-based ADT model considering autoregressive errors can be given by:

$$\begin{cases} Y(t) = X(t) + \varepsilon(t) \\ X(t) = X_0 + \upsilon(S,\boldsymbol{\beta})\Lambda(t,\boldsymbol{\theta}) + \sigma B(\Lambda(t,\boldsymbol{\theta})) \\ \varepsilon(t_i) = \sum_{k=1}^{p} \varphi_k \varepsilon(t_{i-k}) + e_i, \quad k \le i \end{cases} \tag{1}$$

where $Y(t)$ and $X(t)$ respectively denote the degradation inspection and the true performance degradation value at time t; $X_0 = X(0)$ is the true initial degradation level; $\upsilon(S,\beta)$ is the drift coefficient; S is the stress, and β is the unknown parameter vector in $\upsilon(S,\beta)$; $\Lambda(t,\theta)$ called transformed time scale is a positive non-decreasing function and θ is the unknown parameter vector in $\Lambda(t,\theta)$; For convenience, let $\Lambda = \Lambda(t,\theta)$; $B(\cdot)$ is a standard Wiener process and

$\sigma > 0$ denotes the diffusion coefficient; $\varepsilon(t)$ is the measurement error term at time t; p denotes the order of the autoregressive process; φ_k is the autocorrelation coefficient; $e_i \overset{iid}{\sim} N(0,\sigma_e^2)$ is the normally distributed and mutually independent. The improved ADT model expressed by Eq.(1) shows the following main properties:

1. The true performance degradation $X(t)$ under stress level S is assumed to be characterized by a Wiener process $B(\Lambda(t,\theta))$ and a drift coefficient $\upsilon(S,\beta)$ along with the diffusion coefficient σ and the true initial degradation level X_0. $\upsilon(S,\beta)$ is adopted to describe the accelerated relationship between the degradation rate and the stress S.
2. The accelerated model $\upsilon(S,\beta)$ can be constructed as $\upsilon(S,\beta)=\beta_0\lambda(S,\beta_1)$, where $\beta=(\beta_0,\beta_1)$ is unknown parameter vector; $\lambda(S,\beta_1)$ is a function that depends on stress S and can be obtained based on the acceleration relationship. Currently, commonly adopted acceleration models include the power rule model, Arrhenius model and Eyring model. For example, when Arrhenius model is considered as the acceleration model, the acceleration relationship can be supposed as $\lambda(S,\beta_1)=\exp(-\beta_1/S)$; While for the power rule model, one has $\lambda(S,\beta_1)=S^{\beta_1}$. In ADT analysis, to analyze the reliability at a use condition, the above accelerated models are usually utilized to describe the relation between the degradation rate and the stress [4, 7, 10-12, 14, 25].
3. The measurement error term $\varepsilon(t)$ in the accelerated degradation model is considered as a p-order autoregressive time series process $\varepsilon(t_i)=\sum_{k=1}^{p}\varphi_k\varepsilon(t_{i-k})+e_i$, where $|\varphi_k|<1$; i.e., AR(p). In practical engineering, it has been recognized that for longitudinal data or degradation data, an one-order autoregressive model can usually effectively describe the autocorrelation in the within-individual measurement errors [3, 9, 13]. Meanwhile, for the ADT model with p-order autoregressive error, the complex model increases the difficulties of the solving process and it is difficult to acquire the estimation of the unknown parameters. Thus, in this paper, the situation that the measurement error term $\varepsilon(t)$ is a one-order autoregressive time series process is focused, that is $\varepsilon(t_i)=\varphi\varepsilon(t_{i-1})+e_i$.
4. To consider the unit specific variant properties cause by the product-to-product differences, both the parameter β_0 in the drift coefficient $\upsilon(S,\beta)$ and the initial degradation level X_0 are assumed to be normally distributed random variables; i.e. $X_0 \sim N(\mu_0,\sigma_0^2)$, $\beta_0 \sim N(\mu_b,\sigma_b^2)$. Moreover, X_0, β_0, $B(\Lambda(t,\theta))$ and $\varepsilon(t)$ are assumed to be mutually independent of each other.

As described above, the proposed ADT model in Eq. (1) can depict the uncertainties from the temporal variability, the unit-to-unit variability ($\sigma_0 \neq 0$ or $\sigma_b \neq 0$) and measurement errors ($\sigma_e \neq 0$) incorporated in accelerated degradation processes, and is applicable for linear and nonlinear degradation processes. In addition, the proposed ADT model can cover several commonly Wiener process-based ADT models as its limiting cases, for example, if $\sigma_0 = 0$, $\sigma_e = 0$ and $\varphi_k = 0$, $k = 1,2,\ldots,p$, the proposed ADT model can be simplified to the existing widely used Wiener process-based ADT model [24, 25].

2.2. Derivation of lifetime distribution

To assess the product reliability at a use condition, it is necessary to derive FTD based on the proposed CSADT model. Without loss of generality, we first assume the degradation to be an increasing procedure over time, and a product is deemed to be failed when its true degradation first exceeds a predefined failure threshold. As discussed above, in some cases, it is necessary to incorporate the randomness of the initial degradation level X_0 into the ADT modeling procedure.

To derive the lifetime distribution, first let D_f^* denote the failure threshold when the initial true degradation level X_0 is considered as 0; i.e., $X_0 = 0$. According to the FHT concept, it is natural to define life T based on the true degradation path $X(t)$ as:

$$T_{|X_0=0} = \inf\{t : X(t) \geq D_f^* \mid X_0 = 0\} \tag{2}$$

In this situation, life T follows an inverse Gaussian distribution under the concept of FHT for a Wiener process according to the literature [2]. When the initial degradation level and the drift coefficient are supposed as $X_0 = 0$ and $\upsilon(S,\boldsymbol{\beta}) \sim N(\mu_\upsilon,\sigma_\upsilon^2)$, the probability distribution function (PDF) and the cumulative distribution function (CDF) of life T for the Wiener process given in Eq. (1) can be expressed as [25]:

$$f_T(t|X_0=0) = \frac{D_f^*}{\sqrt{2\pi\Lambda^3(\sigma_\upsilon^2\Lambda+\sigma^2)}}\exp\left\{-\frac{(D_f^*-\mu_\upsilon\Lambda)^2}{2\Lambda(\sigma_\upsilon^2\Lambda+\sigma^2)}\right\}\frac{d\Lambda}{dt} \tag{3}$$

$$F_T(t|X_0=0) = \Phi\left(\frac{\mu_\upsilon\Lambda-D_f^*}{\sqrt{\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda}}\right)+\exp\left(\frac{2\mu_\upsilon D_f^*}{\sigma^2}+\frac{2\sigma_\upsilon^2(D_f^*)^2}{\sigma^4}\right)\Phi\left(-\frac{\sigma^2\mu_\upsilon\Lambda+(2\sigma_\upsilon^2\Lambda+\sigma^2)D_f^*}{\sigma^2\sqrt{\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda}}\right) \tag{4}$$

where $\Phi(\cdot)$ is the distribution function of a standard normal distribution. For the proposed CSADT model, $\mu_\upsilon = \mu_b\lambda(S,\beta_1)$ and $\sigma_\upsilon = \sigma_b\lambda(S,\beta_1)$.

Then, further considering the randomness of the initial degradation level X_0, let D_f denote the failure threshold, life T regarding the true degradation path $X(t)$ can be defined as:

$$T = \inf\{t : X(t) \geq D_f \mid X_0 \leq D_f\} \tag{5}$$

where the failure threshold $D_f = X_0 + D_f^*$. Then $D_f^* = D_f - X_0$ can be considered as a random variable with a normal distribution $D_f^* \sim N(D_f-\mu_0,\sigma_0^2)$. To this end, a proposition, which can significantly simplify the CDF and PDF derivation procedure, is given as follows:

Proposition 1: Let $X \sim N(\mu,\sigma^2)$ and a, b, c, $d \in \mathbf{R}$, then the following constructions hold:

$$E_X[\Phi(a+bX)] = \Phi\left((a+b\mu)/\sqrt{1+b^2\sigma^2}\right) \tag{6}$$

$$E_X\left[\exp\left(aX+bX^2\right)\Phi\left(c+dX\right)\right]=\frac{1}{\sqrt{1-2b\sigma^2}}\exp\left[\frac{2\mu\left(a+b\mu\right)+a^2\sigma^2}{2\left(1-2b\sigma^2\right)}\right]\times\Phi\left(\frac{c+d\mu+\left(ad-2bc\right)\sigma^2}{\sqrt{\left(1-2b\sigma^2\right)\left(1-2b\sigma^2+d^2\sigma^2\right)}}\right) \tag{7}$$

The proof is given in the **Appendix A**. Then based on Proposition 1, Eq. (6) and Eq. (7), CDF and PDF of life T for the proposed model can be obtained via the total probability law. It is shown in the following **Proposition 2**.

Proposition 2: When the initial degradation level is considered as a normal random variable $X_0 \sim N\left(\mu_0,\sigma_0^2\right)$, PDF and the CDF of life T for the proposed model can be expressed as:

$$f_T\left(t\right)=\frac{\left(\sigma_\upsilon^2\Lambda+\sigma^2\right)\left(D_f-\mu_0\right)+\mu_\upsilon\sigma_0^2}{\sqrt{2\pi\left(\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda+\sigma_0^2\right)^3}}\exp\left\{-\frac{\left(D_f-\mu_0-\mu_\upsilon\Lambda\right)^2}{2\left(\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda+\sigma_0^2\right)}\right\}\frac{d\Lambda}{dt} \tag{8}$$

$$F_T\left(t\right)=\Phi\left(\frac{\mu_0+\mu_\upsilon\Lambda-D_f}{\sqrt{\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda+\sigma_0^2}}\right)+\frac{\sigma^2}{\sqrt{\sigma^4-4\sigma_\upsilon^2\sigma_0^2}}\exp\left\{\frac{\left(D_f-\mu_0\right)\left[2\mu_\upsilon\sigma^2+2\sigma_\upsilon^2\left(D_f-\mu_0\right)\right]+2\mu_\upsilon^2\sigma_0^2}{\sigma^4-4\sigma_\upsilon^2\sigma_0^2}\right\}\times\Phi\left[-\frac{\mu_\upsilon\left(\sigma^2\Lambda+2\sigma_0^2\right)+\left(2\sigma_\upsilon^2\Lambda+\sigma^2\right)\left(D_f-\mu_0\right)}{\sqrt{\left(\sigma^4-4\sigma_\upsilon^2\sigma_0^2\right)\left(\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda+\sigma_0^2\right)}}\right] \tag{9}$$

The proof is given in the **Appendix B**. Then, the mean time to failure (MTTF) t_{MTTF} can be approximately obtained by:

$$t_{MTTF}=\Lambda^{-1}\left(\frac{D_f-\mu_0}{\mu_\upsilon}\right) \tag{10}$$

3. Parameter estimation

In a CSADT, let S_0 be a use stress level and $S_1<S_2<,\ldots,<S_L$ denote L higher stress levels. Suppose m_l units are tested under stress level S_l, and the corresponding performance degradation of the ith unit is measured at n_{li} test time points $t_{li1}<t_{li2}<,\ldots,<t_{lin_{li}}$, $l=1,2,\ldots,L$, $i=1,2,\ldots,m_l$.

3.1. MLE for unknown parameters

For simplicity, let $y_{lij}=Y\left(t_{lij}\right)$ denote the degradation inspection for unit i at time t_{lij} under stress level S_l, and suppose $\Lambda_{lij}=\Lambda\left(t_{lij}\right)$ and $\lambda_l=\lambda\left(S_l,\beta_1\right)$, $l=1,2,\ldots,L$, $i=1,2,\ldots,m_l$, $j=1,2,\ldots,n_{li}$. Meanwhile, further define $\Lambda_{li}=\left(\Lambda_{li1},\Lambda_{li2},\ldots,\Lambda_{lin_{li}}\right)'$, $\boldsymbol{y}_{li}=\left(y_{li1},y_{li2},\ldots,y_{lin_{li}}\right)'$, $\boldsymbol{y}_l=\left(\boldsymbol{y}_{l1},\boldsymbol{y}_{l2},\ldots,\boldsymbol{y}_{lm_l}\right)$ and $\boldsymbol{y}=\left(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_L\right)$. Then, $\boldsymbol{y}_{li}$ can be concluded to follow a multivariate normal distribution with mean $\mu_0 1_{li}+\mu_b\lambda_l\Lambda_{li}$ and covariance $\Sigma_{li}=\sigma_0^2 1_{li}1_{li}'+\sigma_b^2\lambda_l^2\Lambda_{li}\Lambda_{li}'+\sigma^2\boldsymbol{Q}_{li}+\gamma_0\Omega_{li}$, where $\gamma_0=\sigma_e^2/\left(1-\varphi^2\right)$, $1_{li}=\left(1,1,\ldots,1\right)'$ is an n_{li} dimensional column vector, and:

$$\boldsymbol{Q}_{li}=\begin{bmatrix}\Lambda_{li1} & \Lambda_{li1} & \cdots & \Lambda_{li1}\\ \Lambda_{li1} & \Lambda_{li2} & \cdots & \Lambda_{li2}\\ \vdots & \vdots & \ddots & \vdots\\ \Lambda_{li1} & \Lambda_{li2} & \cdots & \Lambda_{lin_{li}}\end{bmatrix}_{n_{li}\times n_{li}}$$

and

$$\Omega_{li}=\begin{bmatrix}1 & \varphi & \varphi^2 & \cdots & \varphi^{n_{li}-1}\\ \varphi & 1 & \varphi & \cdots & \varphi^{n_{li}-2}\\ \varphi^2 & \varphi & 1 & \cdots & \varphi^{n_{li}-3}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \varphi^{n_{li}-1} & \varphi^{n_{li}-2} & \varphi^{n_{li}-3} & \cdots & 1\end{bmatrix}_{n_{li}\times n_{li}}.$$

Moreover, to facilitate the estimation and inference, the parameters are re-parameterize as $\tilde{\sigma}_b^2=\sigma_b^2/\sigma_0^2$, $\tilde{\sigma}^2=\sigma^2/\sigma_0^2$, $\tilde{\gamma}_0=\gamma_0/\sigma_0^2$ and $\tilde{\Sigma}_{li}=\Sigma_{li}/\sigma_0^2$. $\Theta=\left(\mu_0,\sigma_0^2,\mu_b,\tilde{\sigma}_b^2,\beta_1,\theta,\tilde{\sigma}^2,\tilde{\gamma}_0,\varphi\right)$ is further defined as a vector involving all unknown parameters in the proposed model. Then the log-likelihood function (Log- LF) of Θ can be expressed as:

$$\ell\left(\Theta|\boldsymbol{y}\right)=-\frac{N}{2}\left[\ln\left(2\pi\right)+\ln\sigma_0^2\right]-\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{m_l}\ln\left|\tilde{\Sigma}_{li}\right|-\frac{1}{2\sigma_0^2}\sum_{l=1}^{L}\sum_{i=1}^{m_l}\left(\boldsymbol{y}_{li}-\mu_0 1_{li}-\mu_b\lambda_l\Lambda_{li}\right)'\tilde{\Sigma}_{li}^{-1}\left(\boldsymbol{y}_{li}-\mu_0 1_{li}-\mu_b\lambda_l\Lambda_{li}\right) \tag{11}$$

where $N=\sum_{l=1}^{L}\sum_{i=1}^{m_l}n_{li}$.

Taking the first derivative of $\ell\left(\Theta|\boldsymbol{y}\right)$ in Eq. (11) with respective to μ_0, μ_b and σ_0^2, one can obtain:

$$\frac{\partial\ell\left(\Theta|\boldsymbol{y}\right)}{\partial\mu_0}=\frac{1}{\sigma_0^2}\sum_{l=1}^{L}\sum_{i=1}^{m_l}1_{li}'\tilde{\Sigma}_{li}^{-1}\left(\boldsymbol{y}_{li}-\mu_0 1_{li}-\mu_b\lambda_l\Lambda_{li}\right) \tag{12}$$

$$\frac{\partial\ell\left(\Theta|\boldsymbol{y}\right)}{\partial\mu_b}=\frac{1}{\sigma_0^2}\sum_{l=1}^{L}\sum_{i=1}^{m_l}\lambda_l\Lambda_{li}'\tilde{\Sigma}_{li}^{-1}\left(\boldsymbol{y}_{li}-\mu_0 1_{li}-\mu_b\lambda_l\Lambda_{li}\right) \tag{13}$$

$$\frac{\partial\ell\left(\Theta|\boldsymbol{y}\right)}{\partial\sigma_0^2}=-\frac{N}{2\sigma_0^2}+\frac{1}{2\sigma_0^4}\sum_{l=1}^{L}\sum_{i=1}^{m_l}\left(\boldsymbol{y}_{li}-\mu_0 1_{li}-\mu_b\lambda_l\Lambda_{li}\right)'\tilde{\Sigma}_{li}^{-1}\left(\boldsymbol{y}_{li}-\mu_0 1_{li}-\mu_b\lambda_l\Lambda_{li}\right) \tag{14}$$

Then, by equating Eq. (12), Eq. (13) and Eq. (14) to zero respectively, the MLE of μ_0, μ_b and σ_0^2 can be obtained as:

$$\hat{\mu}_0=\left(A_3B_2-A_2B_3\right)/\left(A_1B_2-A_2B_1\right) \tag{15}$$

$$\hat{\mu}_b=\left(A_1B_3-A_3B_1\right)/\left(A_1B_2-A_2B_1\right) \tag{16}$$

$$\hat{\sigma}_0^2=\frac{1}{N}\sum_{l=1}^{L}\sum_{i=1}^{m_l}\left(\boldsymbol{y}_{li}-\hat{\mu}_0 1_{li}-\hat{\mu}_b\lambda_l\Lambda_{li}\right)'\tilde{\Sigma}_{li}^{-1}\left(\boldsymbol{y}_{li}-\hat{\mu}_0 1_{li}-\hat{\mu}_b\lambda_l\Lambda_{li}\right) \tag{17}$$

where $A_1=\sum_{l=1}^{L}\sum_{i=1}^{m_l}1_{li}^{'}\tilde{\Sigma}_{li}^{-1}1_{li}$, $A_2=\sum_{l=1}^{L}\sum_{i=1}^{m_l}\lambda_l 1_{li}^{'}\tilde{\Sigma}_{li}^{-1}\Lambda_{li}$,

$A_3=\sum_{l=1}^{L}\sum_{i=1}^{m_l}1_{li}^{'}\tilde{\Sigma}_{li}^{-1}\boldsymbol{y}_{li}$, $B_1=\sum_{l=1}^{L}\sum_{i=1}^{m_l}\lambda_l\Lambda_{li}^{'}\tilde{\Sigma}_{li}^{-1}1_{li}$, $B_2=\sum_{l=1}^{L}\sum_{i=1}^{m_l}\lambda_l^2\Lambda_{li}^{'}\tilde{\Sigma}_{li}^{-1}\Lambda_{li}$

and $B_3=\sum_{l=1}^{L}\sum_{i=1}^{m_l}\lambda_l\Lambda_{li}^{'}\tilde{\Sigma}_{li}^{-1}\boldsymbol{y}_{li}$.

Next, substituting Eq. (16) and Eq. (17) into Eq. (11), a profile log-likelihood function of unknown parameter $\tilde{\Theta}=\left(\tilde{\sigma}_b^2,\beta_1,\theta,\tilde{\sigma}^2,\tilde{\gamma}_0,\varphi\right)$ can be obtained as:

$$\ell\left(\tilde{\Theta}|\boldsymbol{y}\right)=C-\frac{N}{2}\ln\hat{\sigma}_0^2-\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{m_l}\ln\left|\tilde{\Sigma}_{li}\right| \tag{18}$$

where C is a constant. Based on a multiple-dimensional search optimization algorithm, MLE $\hat{\tilde{\Theta}}=\left(\hat{\tilde{\sigma}}_b^2,\hat{\beta}_1,\hat{\theta},\hat{\tilde{\sigma}}^2,\hat{\tilde{\gamma}}_0,\hat{\varphi}\right)$ of unknown parameters can be obtained by maximizing the profile Log-LF. Then MLE of μ_0, μ_b and σ_0^2 can be calculated by substituting $\hat{\tilde{\Theta}}=\left(\hat{\tilde{\sigma}}_b^2,\hat{\beta}_1,\hat{\theta},\hat{\tilde{\sigma}}^2,\hat{\tilde{\gamma}}_0,\hat{\varphi}\right)$ into Eq. (16) and Eq. (17). The other unknown parameters can further be determined via $\hat{\sigma}_b^2=\hat{\sigma}_0^2\hat{\tilde{\sigma}}_b^2$, $\hat{\sigma}^2=\hat{\sigma}_0^2\hat{\tilde{\sigma}}^2$, $\hat{\gamma}_0=\hat{\sigma}_0^2\hat{\tilde{\gamma}}_0$ and $\hat{\sigma}_e^2=\hat{\gamma}_0\left(1-\hat{\varphi}^2\right)$.

3.2. Initial guesses

When a multiple-dimensional search optimization algorithm is applied to numerically maximize the log-likelihood function, a reasonable initial guess is necessary for numerical calculation. Thus, a simple method is given to obtain an educated guess for the initial interval. The detailed procedure is as follows:

1. Let X_{0li} and υ_{li} denote the initial degradation level and the drift coefficient parameter $\upsilon\left(S,\beta\right)$ for the ith unit under stress level S_l respectively, $l=1,2,\ldots,L$, $i=1,2,\ldots,m_l$. Based on the least square method, rough estimates of $X_{0l1},X_{0l2},\ldots,X_{0lm_l}$, $\upsilon_{l1},\upsilon_{l2},\ldots,\upsilon_{lm_l}$ and θ can be obtained by minimizing the mean squared error (MSE):

$$\mathrm{MSE}_1=\sum_{l=1}^{L}\sum_{i=1}^{m_l}\left(\boldsymbol{y}_{li}-X_{0li}1_{li}-\upsilon_{li}\Lambda_{li}\right)^{'}\left(\boldsymbol{y}_{li}-X_{0li}1_{li}-\upsilon_{li}\Lambda_{li}\right) \tag{19}$$

2. The rough estimates of μ_0 and σ_0^2 can be calculated by fitting the estimations $X_{0l1},X_{0l2},\ldots,X_{0lm_l}$, $l=1,2,\ldots,L$.
3. Based on the least square method, the rough estimates of $\beta_{0l1},\beta_{0l2},\ldots,\beta_{0lm_l}$, $l=1,2,\ldots,L$ and β_1 can be obtained by minimizing the following mean squared error (MSE):

$$\mathrm{MSE}_2=\sum_{l=1}^{L}\sum_{i=1}^{m_l}\left(\beta_{li}-\beta_{0li}\lambda\left(S_l,\beta_1\right)\right)^{'}\left(\beta_{li}-\beta_{0li}\lambda\left(S_l,\beta_1\right)\right) \tag{20}$$

4. The rough estimates of μ_b and σ_b^2 can be calculated by fitting the estimations $\beta_{0l1},\beta_{0l2},\ldots,\beta_{0lm_l}$, $l=1,2,\ldots,L$.
5. Based on the estimations of μ_0, σ_0^2, μ_b, σ_b^2, β_1 and θ, the rough estimates of σ^2, σ_e^2 and φ can be obtained by maximizing the profile log-likelihood function in Eq. (11).

Therefore, the starting intervals of the unknown parameters for maximizing the log likelihood function via a multiple-dimensional search optimization algorithm are determined.

4. Simulation study

In this section, to test the efficiency of the proposed method, a comprehensive simulation study has been conducted considering CSADT. For comparison, let M_0 denote the proposed method. To demonstrate the necessity of considering autoregressive measurement errors in the accelerated degradation modeling procedure, a Wiener process-based ADT model M_1, as a special case of model M_0 by setting $\varphi=0$, is considered as a reference method. Meanwhile, another reference model M_2 that setting $\sigma_0=0$ in the model M_0 is also adopted to show the benefits of incorporating the random effects of the initial degradation level. Therefore, both the model comparison and sensitivity analysis of the standard deviation $\sigma_0=0$ of initial degradation and the autocorrelation coefficient φ are conducted to test the efficiency and necessary of the proposed method.

4.1. Model Comparison

Without loss of generality, temperature is considered as the accelerated stress and the transformed time scale function is defined as $\Lambda=t^{\theta}$. The normal stress level is S_0=303.5 K (30 °C) and three accelerated levels are supposed to be S_1 =333.5 K (60 °C), S_2 = 343.5 K (70 °C) and S_3 =353.5 K (80 °C). And an Arrhenius model $\upsilon\left(S,\beta\right)=\beta_0\exp\left\{\beta_1/S\right\}$ is adopted to describe the accelerated relationship. The parameters in simulation model are preset as μ_0 =3, σ_0 =0.9, μ_b =7, σ_b =0.7, β_1 =1200, θ =1.5, σ =0.3, σ_e =0.5 and φ =0.8. A failure threshold value is predefined as D_f = 12. For convenience, suppose m units are tested under each stress level and all items are inspected at n time points with $t=i$, $i=1,2,\ldots,n$.

In order to examine the influence of the sample size on the analytical precision, different combinations of (m,n), which are chosen to be (5, 10), (10, 10), and (20, 10) are considered for the simulation study sequentially. For each combination of (m,n), the mean absolute relative errors (MREs) and the mean square errors (MSEs) of the medium life $t_{0.5}$ and the FTD percentile are calculated by Eq. (21) and Eq. (22) based on Monte Carlo replications.

$$MSE=\frac{1}{K}\sum_{k=1}^{K}\left(\hat{t}_{pk}-t_p\right)^2 \tag{21}$$

$$MRE=\frac{1}{K}\sum_{k=1}^{K}\frac{\left|\hat{t}_{pk}-t_p\right|}{t_p} \tag{22}$$

where K is the number of Monte Carlo replications, t_p denotes the true $100p$th FTD percentile value and $\hat{t}_{pk}$ is the corresponding estimated result under the kth simulation, $k=1,2,\ldots,K$. Comparative results are given in Table 1 and Table 2 based on K=5000 Monte Carlo replications.

From Table 1 and Table 2, it can be observed that for each combinations of (m,n), MSEs and MREs of $t_{0.5}$ and $t_{0.1}$ based on the proposed model M_0 are smaller than that from reference models M_1 and M_2. When sample size is small, MSEs and MREs from the proposed model M_0 are significantly lower than the results given by reference models M_1 and M_2. Meanwhile, when the sample size increases, although MSEs and MREs of $t_{0.5}$ and $t_{0.1}$ based on reference models

Table 1. MSEs and MREs of $t_{0.5}$ from different degradation models

Model	(5,10)		(10,10)		(20,10)	
	MSE	MRE	MSE	MRE	MSE	MRE
M_0	2.6371	0.0785	1.6664	0.0609	1.2013	0.0547
M_1	24.8524	0.1960	14.6416	0.1726	6.6128	0.1209
M_2	17.5616	0.1894	13.1900	0.1637	5.2263	0.1076

Table 2. MSEs and MREs of $t_{0.1}$ from different degradation models

Model	(5,10)		(10,10)		(20,10)	
	MSE	MRE	MSE	MRE	MSE	MRE
M_0	1.4047	0.0737	0.6919	0.0576	0.4719	0.0485
M_1	4.1695	0.1250	2.9078	0.1205	1.7061	0.0939
M_2	7.1508	0.1607	3.6679	0.1272	1.6009	0.0867

Table 3. Comparison of Log_LF and AIC results

Model	(5,10)		(10,10)		(20,10)	
	Log-LF	AIC	Log-LF	AIC	Log-LF	AIC
M_0	-166.14	350.29	-350.26	718.52	-701.94	1421.88
M_1	-171.56	359.11	-357.43	730.85	-715.94	1447.89
M_2	-169.91	355.82	-355.76	727.53	-712.75	1441.50

M_1 and M_2 becomes closer to those of the constructed model M_0, the proposed model M_0 still can yield smaller MSEs and MREs. In addition, one can see that MSEs and MREs of $t_{0.5}$ are larger than those of $t_{0.1}$. This is because that the true value $t_{0.1}$ is smaller comparing with the true value of $t_{0.5}$, which will consequently result in a smaller error.

To further illustrate the efficiency of model M_0, the log-likelihood function value (Log-LF) and the corresponding Akaike information criterion (AIC) value are calculated to compare the modeling reasonableness for each combination of (m,n), where AIC is defined as:

$$AIC = -2 \times \left\{ \max\left[\log\left(likelihood \right) \right] \right\} + 2q \tag{23}$$

where q is the number of the unknown parameters in the adopted accelerated degradation model.

Table 3 lists the average results of Log-LF value and AIC value under different combinations of (m,n) based on K=5000 Monte Carlo replications. It is obvious that in both terms of Log-LF and AIC, the proposed model M_0 gives a better fitting and modeling property. Therefore, although the constructed model M_0 involves one more parameter than reference models M_1 and M_2, it is necessary and worthy to construct the complicated analysis procedure for reasonable and accurate analysis.

4.2. Sensitivity Analysis

In this section, to furthermore test the necessary of considering the autocorrelation among measurement errors and the randomness of initial degradation level into the unit-to-unit variability, the sensitivity of reliability estimation is analyzed by setting different values of the autocorrelation coefficient φ and the standard deviation σ_0 of initial degradation level for the simulation example.

To this end, we first set the autocorrelation coefficient φ =0.1 (0.1) 0.9, and keep all other parameters unchanged. After that, the absolute error and the relative error of reliability evaluation results at normal stress levels are calculated in such a case by comparing model M_0 with M_1. Then, we set the standard deviation σ_0 =0.25 (0.25) 2 and the errors are calculated by comparing model M_0 with M_2. If the autocorrelation coefficient φ and the standard deviation σ_0 are sensitivity, the errors of reliability evaluation results at normal stress levels should increase with the values of φ and σ_0.

Herein, we repeated the simulation procedure of CSADT data for N_s = 100 times under the situation in Section 4.1. Then, the mean absolute error (MAE) and the mean absolute relative error (MARE) of reliability evaluation results for the autocorrelation coefficient φ and the standard deviation σ_0 can be given by:

$$\begin{cases} \mathrm{MAE}_i = \dfrac{1}{N_s}\displaystyle\sum_{k=1}^{N_s}\frac{1}{n}\sum_{j=1}^{n}\left| F_T^k\left(t_j|\mathrm{M}_0\right) - F_T^k\left(t_j|\mathrm{M}_i\right)\right| \\ \mathrm{MARE}_i = \dfrac{1}{N_s}\displaystyle\sum_{k=1}^{N_s}\frac{1}{n}\sum_{j=1}^{n}\frac{\left| F_T^k\left(t_j|\mathrm{M}_0\right) - F_T^k\left(t_j|\mathrm{M}_i\right)\right|}{F_T^k\left(t_j|\mathrm{M}_0\right)} \end{cases} \quad i = 1,2 \tag{24}$$

where $F_T^k\left(t_j|\mathrm{M}_i\right)$ is the CDF at time t_j under the normal stress level for the k th simulation under model M_i, $i = 0,1,2$, $k = 1,2,\ldots,N_s$.

The results of the sensitivity analysis of autocorrelation coefficient φ and standard deviation σ_0 are shown in Fig.1 and Fig.2 respectively. From Fig.1 and Fig.2, it can be observed that MAE and MARE of reliability evaluation results will increase with the autocorrelation coefficient φ and the standard deviation σ_0. Thus, it is clear that the effect of ignoring the autocorrelation among measurement errors and the randomness of initial degradation level on the reliability

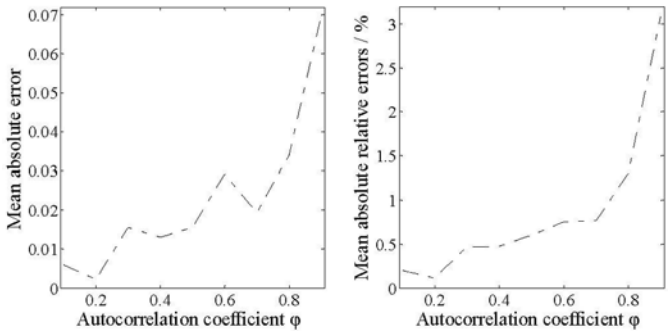


Fig. 1. Sensitivity analysis of autocorrelation coefficient φ

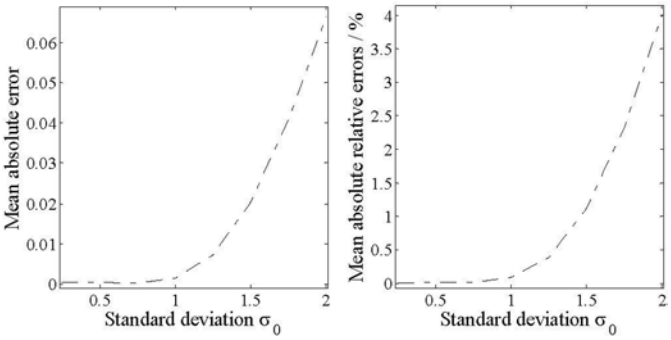


Fig. 2. Sensitivity analysis of standard deviation σ_0

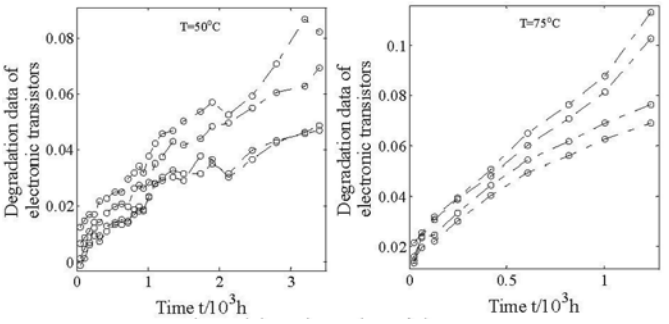


Fig. 3. Accelerated degradation data of electronic transistors

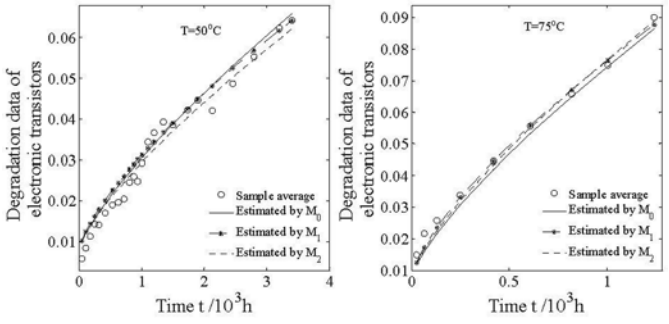


Fig. 4. Estimated mean degradation path based on three models

analysis is not critical under a small autocorrelation coefficient and the standard deviation σ_0 of initial degradation level situation. However, when the autocorrelation coefficient the standard deviation σ_0 are large, the effect is quite serious.

From the above analysis and comparison incorporating the reference models M_1 and M_2, one can conclude that it is necessary to consider the autoregressive measurement errors and the random effects of the initial degradation level into the Wiener process ADT modeling procedure, especially for limited sample size situations. In addition, if degradation model is mis-specified, unreliable results may probably be derived.

5. Illustrative example

In this section, a real application regarding a CSADT of electronic transistors is involved to further illustrate the necessary and validity of the proposed model M_0 for accelerated degradation analysis. An electronic transistor degrades over time and finally fails when its gain, a key performance of transistor, falls to a preset threshold level that makes it nonfunctional in the device where it is placed. To assess the electronic transistor reliability under a use stress level S_0=25ºC, a CSADT was conducted under two higher stress levels S_1=50ºC and S_2=75ºC. For each accelerated stress level, 4 electronic transistors are randomly selected for the degradation test. The original accelerated degradation data are given in [15], and is shown in Fig.3. The failure threshold is preset as D_f=0.15. Meanwhile, models M_1 and M_2 are also considered as reference for comparison.

According to [1, 18], Arrhenius acceleration model is a most common model and has been widely applied when the accelerated variable is temperature. Thus, without loss of generality, Arrhenius acceleration model is utilized to describe the relationship between the degradation rate of electronic transistors and the stress; i.e., $\upsilon(S,\beta)=\beta_0\lambda(S,\beta_1)$ and $\lambda(S,\beta_1)=\exp(-\beta_1/S)$. In addition, empirical studies have shown that $\Lambda(t,\theta)=t^{\theta}$ can be considered as a reasonable transformed time scale form [9, 15]. Consequently, this form is adopted in the current study.

To test the fitting goodness, models M_0, M_1 and M_2 are adopted to fit the accelerated degradation dataset. Unknown parameters of different models are estimated according to the inference procedure given in Section 4. Meanwhile, both Log-LF and AIC values are calculated. Results are summarized in Table 4. From Table 4, on can see that compared with models M_1 and M_2, model M_0 displays a best fitting with a largest Log-LF value and a smallest AIC value.

The one-order autocorrelation coefficient values φ estimated by models M_0 and M_2 are larger than or equal to 0.9, which indicates that the autocorrelation among measurement errors is non-ignorable for this practical problem. Additionally, it is obvious that the estimated value of σ_0 in models M_0 and M_1 are also relatively larger, and one can conclude that the random effect of the initial degradation level is necessary to be considered in ADT modeling.

Based on the results in Table 4, estimated mean degradation paths for the above three models are obtained and shown in Fig.4. From Fig. 4, it shows that for both stress levels, the estimated mean degradation paths based on the three models are all in conjunction with the sample average.

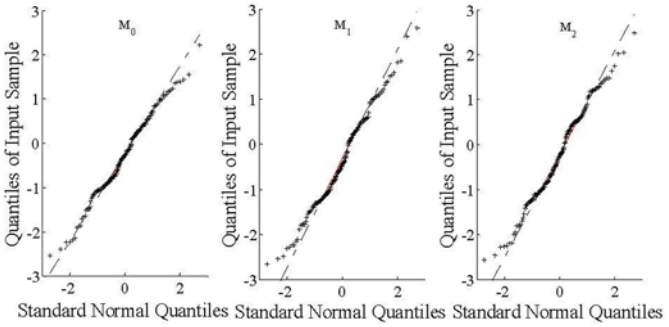


Fig. 5. Comparison of normal probability plots

For further illustration, the normal probability plot is adopted to assess the fitting goodness of models M_0, M_1 and M_2. Fig.5 gives the normal probability plots of the three adopted models. It is well known that if a normal probability plot approximates a straight line, a good fitting can be concluded. Otherwise, more proper degradation models should be considered when a poor fitting is derived. From Fig.5, it is

Table 4. Comparison fitting goodness for different degradation models

Model	Estimated parameters									Log -LF	AIC
	μ_0	σ_0	$\mu_b/10^4$	σ_b	β_1	θ	σ	σ_e	φ		
M_0	0.0077	0.0012	7.6756	0.0117	4.86	0.7707	0.0097	0.0002	0.9995	566.7	-1115.4
M_1	0.0074	0.0028	6.5938	0.0276	4.79	0.7129	0.0087	0.0014	---	563.4	-1110.8
M_2	0.0079	---	24.4991	0.0021	5.25	0.7592	0.0082	0.0018	0.9000	564.3	-1112.5

Table 5. Comparative results of different predicting life

Model	Estimated percentiles of the FTD	
	$t_{0.5}/10^3$h	$t_{0.1}/10^3$h
M_0	55.91	32.62
M_1	70.97	43.00
M_2	73.02	44.40

clear that the proposed model M_0 shows a best fitting compared with the reference models.

In addition, the medium life $t_{0.5}$ and the 10th FTD percentile $t_{0.1}$ for the three models are given in Table 5. It can be observed that results of $t_{0.5}$ and $t_{0.1}$ by reference models M_1 and M_2 are significantly larger than those from the constructed model M_0. It is well known that q-percentile life is commonly considered as an important evidence for making effective maintenance schedule. In practical engineering, a conservative q-percentile life estimation may lead to hysteretic maintenance and increase the failure risk at an early time.

From the above analysis, one can conclude that it is necessary to consider the autocorrelation among measurement errors for reasonable results when modeling accelerated degradation processes. Meanwhile, it is also necessary to incorporate the random effect of the initial degradation level into Wiener process ADT modeling. Although the proposed method may illustrate a more complicated modeling procedure because of the one more parameter, reasonable and reliable results can be governed.

6. Conclusions

Motivated by real applications, this paper proposed a Wiener process accelerated degradation model, which simultaneously considers the temporal variability, the unit-to-unit variability, and measurement errors. In the ADT modeling process, a one-order autoregressive model is utilized to reasonably describe the autocorrelation that may exist among measurement errors. Moreover, the random effects of both the initial degradation level and the degradation rate are incorporated regarding unit specific properties. Then, explicit form of lifetime distribution is derived based on the FHT concept, and a statistical inference method is given for unknown parameter estimation.

A comprehensive simulation study has demonstrated the necessity and efficiency of the proposed model with respect to CSADT analysis via an enhanced accuracy. Finally, a real application about CSADT of electronic transistors has verified the effectiveness and superiority of the constructed method comparing with the commonly used Wiener process models.

In this paper, accelerated degradation analysis for CSADT in focused. However, in practical engineering, SSADT is another effective way to evaluate the reliability of highly reliable products. Thus, the future research will focus on SSADT modeling. In addition, it may be of interest to predict the remaining useful life based on the proposed model. Meawhile, the failure threshold of many degradation process may be unknown and has uncertainty, which can be worth studying. We will work on these problems and hope to have useful findings.

Acknowledgement: The authors are grateful to the anonymous reviewers, and the editor, for their critical and constructive review of the manuscript. This study was co-supported by the National Natural Science Foundation of China (Grant No. 11872085), and Key scientific research projects of Universities of Henan of China (Grant No. 19A460002).

Appendix A

Proof of Proposition 1:

Let $Z \sim N(0,1)$ and $\gamma, \kappa \in \mathbf{R}$, one can derive [22]:

$$E_Z\left[\Phi(\gamma+\kappa Z)\right]=\Phi\left(\gamma\Big/\sqrt{1+\kappa^2}\right) \tag{25}$$

Then, let $X=\mu+\sigma Z$. It can further be obtained that:

$$E_X\left[\Phi(a+bX)\right]=E_Z\left[\Phi(a+b\mu+b\sigma Z)\right]=\Phi\left((a+b\mu)\Big/\sqrt{1+b^2\sigma^2}\right) \tag{26}$$

$$\begin{aligned}E_X\left[\exp\left(aX+bX^2\right)\Phi(c+dX)\right]&=\frac{1}{\sigma}\int_{\mathbf{R}}\exp\left(aX+bX^2\right)\Phi(c+dX)\phi\left(\frac{X-\mu}{\sigma}\right)dX\\&=\exp\left[\mu(a+b\mu)\right]\int_{\boldsymbol{R}}\exp\left[(a+2b\mu)\sigma Z+b\sigma^2Z^2\right]\Phi(c+d\mu+d\sigma Z)\phi(Z)dZ\\&=\frac{1}{\sqrt{1-2b\sigma^2}}\exp\left[\frac{2\mu(a+b\mu)+a^2\sigma^2}{2\left(1-2b\sigma^2\right)}\right]E_Z\left[\Phi\left(c+d\mu+\frac{(a+2b\mu)d\sigma^2}{\left(1-2b\sigma^2\right)}+\frac{d\sigma}{\sqrt{1-2b\sigma^2}}Z\right)\right]\end{aligned} \tag{27}$$

According to Eq. (26), we can derive:

$$E_Z\left[\Phi\left(c+d\mu+\frac{(a+2b\mu)d\sigma^2}{\left(1-2b\sigma^2\right)}+\frac{d\sigma}{\sqrt{1-2b\sigma^2}}Z\right)\right]=\Phi\left(\frac{c+d\mu+(ad-2bc)\sigma^2}{\sqrt{\left(1-2b\sigma^2\right)\left(1-2b\sigma^2+d^2\sigma^2\right)}}\right) \tag{28}$$

Thus:

$$E_X\left[\exp\left(aX+bX^2\right)\Phi(c+dX)\right]=\frac{1}{\sqrt{1-2b\sigma^2}}\exp\left[\frac{2\mu(a+b\mu)+a^2\sigma^2}{2\left(1-2b\sigma^2\right)}\right]$$

$$\times\Phi\left(\frac{c+d\mu+(ad-2bc)\sigma^2}{\sqrt{\left(1-2b\sigma^2\right)\left(1-2b\sigma^2+d^2\sigma^2\right)}}\right) \quad (29)$$

This completes the proof of **Proposition 1**.

Appendix B

Proof of Proposition2:

Considering the randomness of the initial degradation level X_0, the failure threshold $D_f^* = D_f - X_0$ can be considered as a random variable with the normal distribution; i.e., $D_f^* \sim N\left(D_f-\mu_0,\sigma_0^2\right)$. Thus, according to Eq. (4) and the law of total probability, CDF of life T for the proposed model can be expressed as:

$$F_T(t)=E_{D_f}\left\{\Phi\left(\frac{\mu_\upsilon\Lambda-D_f}{\sqrt{\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda}}\right)\right\}$$
$$+E_{D_f}\left\{\exp\left(\frac{2\mu_\upsilon D_f}{\sigma^2}+\frac{2\sigma_\upsilon^2 D_f^2}{\sigma^4}\right)\Phi\left(-\frac{\sigma^2\mu_\upsilon\Lambda+\left(2\sigma_\upsilon^2\Lambda+\sigma^2\right)D_f}{\sigma^2\sqrt{\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda}}\right)\right\} \quad (30)$$

For the first term in Eq. (30), let:

$$a=\frac{\mu_\upsilon\Lambda}{\sqrt{\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda}},\quad b=-\frac{1}{\sqrt{\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda}} \quad (31)$$

Meanwhile, for the second term in Eq. (30), let:

$$a=\frac{2\mu_\upsilon}{\sigma^2},\ b=\frac{2\sigma_\upsilon^2}{\sigma^4},\ c=-\frac{\sigma^2\mu_\upsilon\Lambda}{\sigma^2\sqrt{\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda}},\ d=-\frac{\left(2\sigma_\upsilon^2\Lambda+\sigma^2\right)}{\sigma^2\sqrt{\sigma_\upsilon^2\Lambda^2+\sigma^2\Lambda}} \quad (32)$$

Then, according to **Proposition 1**, CDF of life T can be derived for the proposed model. Furthermore, PDF of life can be accordingly obtained by taking the derivative of CDF with respect to t.

This completes the proof of **Proposition 2**.

References

1. Chen D G, Lio Y, Ng H K T, Tsai T R. Statistical Modeling for Degradation Data. Springer, 2017, https://doi.org/10.1007/978-981-10-5194-4.
2. Chhikara R S, Folks J L. The inverse Gaussian distribution: theory, methodology, and applications, Marcel Dekker, Inc.1989.
3. Chi E M, Reinsel G C. Models for longitudinal data with random effects and AR (1) errors, Publications of the American Statistical Association 1989; 84: 452-459, https://doi.org/10.1080/01621459.1989.10478790.
4. Hao S, Yang J, Berenguer C. Nonlinear step-stress accelerated degradation modelling considering three sources of variability, Reliability Engineering & System Safety 2017; 172.
5. Josephlu C, Meeker W. Using degradation measures to estimate a Time-to-Failure distribution. Technometrics 1993: 161-174.
6. Lei Y, Li N, Jia F, Lin J, Xing S. A nonlinear degradation model based method for remaining useful life prediction of rolling element bearings, Prognostics and System Health Management Conference 2016: 1-8.
7. Li J, Wang Z, Liu X, Zhang Y, Fu H, Liu C. A Wiener process model for accelerated degradation analysis considering measurement errors, Microelectronics Reliability 2016; 65: 8-15,https://doi.org/10.1016/j.microrel.2016.08.004.
8. Li J, Wang Z, Zhang Y, Fu H, Liu C, Krishnaswamy S. Degradation data analysis based on a generalized Wiener process subject to measurement error, Mechanical Systems & Signal Processing 2017; 94: 57-72, https://doi.org/10.1016/j.ymssp.2017.02.031.
9. Li J, Wang Z, Zhang Y, Liu C, Fu H. A nonlinear Wiener process degradation model with autoregressive errors, Reliability Engineering & System Safety 2018: 173, https://doi.org/10.1016/j.ress.2017.11.003.
10. Liao C M, Tseng S T. Optimal design for step-stress accelerated degradation tests, IEEE Transactions on Reliability 2006; 55(1):59-66, https://doi.org/10.1109/TR.2005.863811.
11. Lim H, Yum B J. Optimal design of accelerated degradation tests based on Wiener process models, Journal of Applied Statistics 2011; 38(2):17-27, https://doi.org/10.1080/02664760903406488.
12. Limon S, Yadav O P, Liao H. A literature review on planning and analysis of accelerated testing for reliability assessment. Quality and Reliability Engineering International 2017; 33: 2361-2383, https://doi.org/10.1002/qre.2195.
13. Lin J, Wei B. Testing for heteroscedasticity and/or autocorrelation in longitudinal mixed effect nonlinear models with AR(1) errors. Commun Stat –Theory Methods 2007; 36: 67–86, https://doi.org/10.1080/03610920601001816.
14. Liu L, Li X, Sun F, Wang N. A General Accelerated Degradation Model Based on the Wiener Process, Materials 2016; 9, https://doi.org/10.3390/ma9120981.
15. Lu J. Degradation processes and related reliability models, Ph.D. thesis. Montreal, Canada: McGill University,1995.
16. Lu C J, Meeker W Q, Escobar L A. A comparison of degradation and failure-time analysis methods for estimating a time-to-failure distribution. Stat. Sin. 1996; 6: 531–546.
17. Madsen H, Time Series Analysis, Chapman & Hall/crc Boca Raton, 2007.
18. Meeker W Q, Escobar L A. Statistical methods for reliability data, John Wiley & Sons, 2014.
19. Pan D, Wei Y, Fang H, Yang W. A reliability estimation approach via Wiener degradation model with measurement errors. Applied Mathematics & Computation, 2018; 320,https://doi.org/10.1016/j.amc.2017.09.020.
20. Peng C Y, Tseng S T. Mis-Specification Analysis of Linear Degradation Models, IEEE Transactions on Reliability 2009; 58: 444-455, https://

doi.org/10.1109/TR.2009.2026784.
21. Shahraki A F, Yadav O P, Liao H. A Review on Degradation Modelling and Its Engineering Applications, International Journal of Performability Engineering 2017;13: 299-314, https://doi.org/10.23940/ijpe.17.03.p6.299314.
22. Si X S, Wang W, Hu C H, Chen M Y, Zhou D H. A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation, Mechanical Systems & Signal Processing 2013; 35: 219-237, https://doi.org/10.1016/j.ymssp.2012.08.016.
23. Si X S, Wang W, Hu C H, Zhou D H. Estimating remaining useful life with three-source variability in degradation modeling. IEEE Transactions on Reliability 2014; 63:167-190, https://doi.org/10.1109/TR.2014.2299151.
24. Sun L, Gu X, Song P. Accelerated Degradation Process Analysis Based on the Nonlinear Wiener Process with Covariates and Random Effects, Mathematical Problems in Engineering 2016; 1-13, https://doi.org/10.1155/2016/5246108.
25. Tang S, Guo X, Yu C, Xue H, Zhou Z. Accelerated Degradation Tests Modeling Based on the Nonlinear Wiener Process with Random Effects, Mathematical Problems in Engineering 2014: 1-11,https://doi.org/10.1155/2014/560726.
26. Vališ D, Žák L, Pokora O. Perspective approach in using anti-oxidation and anti-wear particles from oil to estimate residual technical life of a system, Tribology International 2018; 118: 46-59, https://doi.org/10.1016/j.triboint.2017.09.017.
27. Vališ D, Žák L, Pokora O, Lánský P. Perspective analysis outcomes of selected tribodiagnostic data used as input for condition based maintenance, Reliability Engineering & System Safety 2016; 145: 231-242,https://doi.org/10.1016/j.ress.2015.07.026.
28. Vališ D, Nováček O, Hasilová K, Leuchter J. Modelling of degradation and a soft failure moment during the operation of a supercapacitor applying selected diffusion processes, Engineering Failure Analysis 2017; 82: 566-582, https://doi.org/10.1016/j.engfailanal.2017.04.019.
29. Whitmore G A. Estimating degradation by a wiener diffusion process subject to measurement error, Lifetime Data Analysis 1995; 1: 307-319, https://doi.org/10.1007/BF00985762.
30. Whitmore G A, Schenkelberg F. Modelling accelerated degradation data using Wiener diffusion with a time scale transformation, Lifetime Data Analysis 1997; 3: 27-45, https://doi.org/10.1023/A:1009664101413.
31. Yang G. Life cycle reliability engineering. Wiley, 2008.
32. Ye Z S, Wang Y, Tsui K L, Pecht M. Degradation Data Analysis Using Wiener Processes With Measurement Errors, IEEE Transactions on Reliability 2013; 62: 772-780, https://doi.org/10.1109/TR.2013.2284733.
33. Ye Z S, Xie M. Stochastic modelling and analysis of degradation for highly reliable products, Applied Stochastic Models in Business & Industry 2015; 31: 16-32, https://doi.org/10.1002/asmb.2063.
34. Zhang Z, Si X, Hu C, Lei Y. Degradation Data Analysis and Remaining Useful Life Estimation: A Review on Wiener-Process-Based Methods, European Journal of Operational Research 2018,https://doi.org/10.1016/j.ejor.2018.02.033.
35. Zheng J F, Si X S, Hu C H, Zhang Z X, Jiang W. A Nonlinear Prognostic Model for Degrading Systems With Three-Source Variability, IEEE Transactions on Reliability 2016; 65: 736-750, https://doi.org/10.1109/TR.2015.2513044.

Junxing LI
School of Mechatronical Engineering
Henan University of Science and Technology
Luoyang 471003, China

Zhihua WANG
School of Aeronautic Science and Engineering
Beihang University
Beijing 100191, China

Chengrui LIU
Beijing Institute of Control Engineering
Beijing 100080, China

Ming QIU
School of Mechatronical Engineering
Henan University of Science and Technology
Luoyang 471003, China

E-mails: lijun-xing2008@163.com, wangzhihua@buaa.edu.cn, liuchengrui_502@163.com, qiuming69@126.com

Article citation info:
KNOPIK L, MIGAWA K. Semi-Markov system model for minimal repair maintenance. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 256–260, http://dx.doi.org/10.17531/ein.2019.2.9.

Leszek KNOPIK
Klaudiusz MIGAWA

SEMI-MARKOV SYSTEM MODEL FOR MINIMAL REPAIR MAINTENANCE

SEMI-MARKOWSKI MODEL SYSTEMU OBSŁUGI Z MINIMALNĄ NAPRAWĄ*

This paper analyzes the semi-Markov model of technical objects age-replacements. The model includes two types of repairs: perfect repairs and minimal repairs. Minimal repairs in semi-Markov models have been studied in literature only to an extent. In this paper, the asymptotic availability coefficient and profit per time unit are considered as criteria for the quality of the system operation. The paper formulates various conditions for the occurrence of the maximum of criteria functions. The two numerical examples given at the end of the paper illustrate the results obtained in the paper.

Keywords: *age-replacement, minimal repair, perfect repair, profit per time unit, availability, semi-Markov processes, preventive maintenance, corrective maintenance.*

W pracy bada się semimarkowski model wymian według wieku obiektów technicznych. W modelu uwzględnia się dwa rodzaje napraw: naprawy dokładne i naprawy minimalne. Naprawy minimalne w modelach semimarkowskich były badane w literaturze w niewielkim stopniu. Jako kryteria jakości pracy systemu rozważa się asymptotyczny współczynnik gotowości i zysk przypadający na jednostkę czasu. W pracy sformułowano różne warunki istnienia maksimum funkcji kryterialnych. Podane na końcu pracy dwa przykłady numeryczne ilustrują wyniki uzyskane w pracy.

Słowa kluczowe: *wymiana według wieku, naprawa minimalna, naprawa dokładna, zysk na jednostkę czasu, gotowość, procesy semi-Markowa, obsługa prewencyjna, obsługa korekcyjna.*

1. Introduction

During a long period of use, technical systems are prone to degradation processes. The resulting failure has a negative impact on the security and income of the system. Failure can, in turn, cause further failure to the system. In order to reduce the amount of failure to technical objects, various strategies of preventive actions are introduced into system management. The problem of reduction of the costs of system maintenance arises. This requires the development of effective repair and replacement strategies. Managing exchanges and repairs in industrial systems requires introducing various activities related to maintenance as well as appropriate level of reliability and availability into the system. These activities are divided into two types: preventive maintenance (PM) and repair, or corrective maintenance (CM). Corrective maintenance in practice is carried out in two variants: after repair the system is "good as new" (perfect repair) or "bad as old" (minimal repair). Minimal repair restores the system to its reliability condition just before failure. In practice, it restores the system to an intermediate state between the two possible extreme cases. The condition resulting from this activity is referred to as imperfect maintenance. Various models of imperfect maintenance are presented in detail in the review papers [12, 13].

Reduction of system maintenance costs is achieved by implementing various effective prevention strategies and repairs. These activities include the replacement of important system components and determining the frequency of inspections. The schedule of these activities is often set by the system designer or manufacturer. The maintenance department also decides about the replacement of worn components. CM always requires prior diagnosis and identification of failure, therefore it is expensive and can be done by highly trained personnel only. CM repair costs are generally higher than the costs of preventive maintenance (PM). Similarly, average repair times are higher than average times of preventive maintenance. For some industrial systems, it is also possible to repair a failed component without replacing it. This type of repair can be considered as a minimal repair (MR). Minimal repair restores the failed object to the state before the failure. From this point of view, some replacements can be considered minimal repairs. Based on this argument, many practical models of exchanges with minimal repair have been suggested in literature. As a result, developing different prevention strategies suggested by optimal decision-making models to reduce system maintenance costs and reduce the risk of undesired events is an important research topic in reliability engineering. In the last four decades, preventive maintenance models have generated growing interest in system reliability research.

The concept of minimal repair was introduced by Brown and Prochan in paper [2]. The minimum repair model assumes that once the failure occurs, perfect repair is carried out with p probability and minimal repairs are carried out with 1-p probability. Perfect repair restores the technical object to the "good as new" condition. If $p = 0$, the repair is always minimal, while if $p = 1$, the repair is always perfect. Pham and Wang in paper [13] called such a mechanism of repairing an imperfect maintenance model with the rule (p, q). In paper [2] it is assumed that the probability of perfect repair depends on the age of the technical object at the time of failure. In literature, the construction of the minimal repairs model is carried out using various mathematical methods. The review of the methods of constructing criterion functions in the models of minimal repairs with preventative maintenance by age is found in papers [12, 13]. However, only one work cited there [3] uses semi-Markov processes. A more recent review of papers on minimal repairs is featured in book [15]. Recently, in papers [4, 5, 16, 17] new results have been obtained regarding minimal repairs. The problem of minimal repairs is considered from the economic point of view in paper [6]. In addition, the article contains an up-to-date and

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

extensive literature review on minimal preventive repairs and replacements.

This paper analyzes the strategy of maintaining the system using the (p, q) rules of age replacement. The possibility of using semi-Markov processes to build a preventive replacement model in systems with minimal repair is discussed. The basis for building a criterion function is a certain border theorem for semi-Markov processes [7, 8]. This approach to the construction of the criterion function was used in the works [11, 12]. The results of paper [3] are a special case of results obtained for the 3-state model in paper [11] and this article. In this paper, unlike in most articles on maintenance, repair times are not negligible. In the article, profit per unit of time and system availability rate are tested as criterion function. The conditions for the occurrence of an exactly maximum of both criterion functions have been formulated. Chapter 2 defines a 4-state model of replacements with minimal repairs and specifies a criterion function as profit per unit of time. Chapter 3 contains sufficient conditions for the occurrence of maximum profit per unit of time and maximum of the availability rate. In Chapter 4, two numerical examples are analyzed showing the results obtained in the paper. In the first example, the availability rate is maximized, while in the second the profit per time unit is maximized. In both examples it was assumed that the time before failure has Weibull distribution.

2. Criterion function

The paper examines the system in which the technical object may belong to one of the four states: S_1 – failure free operation state, S_2 – minimal repair state, S_3 – perfect repair state, S_4 – preventive replacement state. Possible changes of states are shown in the graph in Fig. 1.

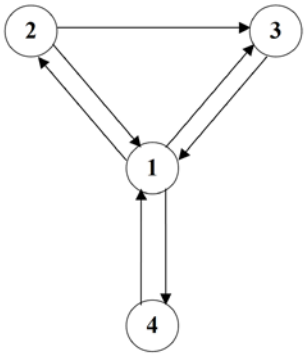


Fig. 1. Directed graph for changes of states S = { S_1, S_2, S_3, S_4 }

In cases when we know the probabilities of transition between states, we have a given Markov chain. The matrix of Markov chain transition has the following form:

$$P = \begin{bmatrix} 0 & p_{12} & p_{13} & p_{14} \\ p_{21} & 0 & p_{23} & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

By solving the appropriate system of linear equations, limit probabilities for the Markov chain are obtained. The analyzed chain has the following limit probabilities:

$$\begin{gathered} p_1^* = 1 / M, \\ p_2^* = p_{12} / M, \\ p_3^* = (p_{13} + p_{12}\, p_{23}) / M, \\ p_4^* = p_{14} / M, \end{gathered} \tag{1}$$

where $M = 2 + p_{12}\, p_{23}$.

The article analyzes the semi-Markov model of preventive repairs and age replacements. The 4-state semi-Markov X(t) process is discussed with the state space S = {1, 2, 3, 4}. If X(t) = i, then the analyzed technical object at moment t is at state i. The profit per time unit for state i is determined by z_i, i = 1, 2, 3, 4. In the paper it is assumed that $z_1 > 0$, $z_i \le 0$ for $2 \le i \le 4$. If the technical object is at state 1, it brings profit, whereas if the technical object is in the state i, where $2 \le i \le 4$, then the technical object generates loss.

The unit is replaced at age T or when it is failed, whichever comes first. Replacement or failure free time is defined by $T_1(x)$. The variable $T_1(x)$ can be written as:

$$T_1(x) = \begin{cases} T_1, & if \quad T_1 < x, \\ x, & if \quad T_1 \ge x. \end{cases} \tag{2}$$

In paper [11] it was proven that profit per time unit is expressed through the formula:

$$L = \frac{\sum_{i=1}^{4} z_i p_i^* ET_i}{\sum_{i=1}^{4} p_i^* ET_i}, \tag{3}$$

where ET_i, i = 1, 2, 3, 4 is average time of the technical object remaining at state S_i.

It is assumed that after the time x, if the object has not been failed, it goes into the prevention (replacement) state. The process of changes of states s_i, i = 1, 2, 3, 4, taking into account the preventive replacement after time x, is a new semi-Markov process with the matrix P(x) of probabilities of transition of the embedded Markov chain. In relation to the matrix P described above, only the first line of matrix P changes. In particular, based on paper [11], you can write:

$$\begin{gathered} p_{12}(x) = p_{12}\, F_{12}(x), \\ p_{13}(x) = p_{13}\, F_{13}(x), \\ p_{14}(x) = p_{14}\, F_{14}(x) + R_1(x). \end{gathered}$$

where:
$F_{1i}(x)$, i = 1, 2, 3, 4 are conditional distribution functions of time spent remaining at state i, before transition to state j, defined as follows:

$$F_{ij}(t) = P\left\{\tau_{k+1} - \tau_k < t \,\middle|\, X(\tau_{k+1}) = j, \;\; X(\tau_k) = i\right\}, \;\; for\; i, j = 1,2,3,4,$$

$R_1(x) = 1 - F_1(x)$ is a function of reliability T_1.

In order to simplify calculations, it is assumed that the following equations are true:

$$F_{12}(x) = F_{13}(x) = F_{14}(x) = F_1(x).$$

On the basis of paper [11], criterion function has the following form:

$$g(x) = \frac{z_1 ET_1(x) p_1^*(x) + z_2 ET_2 p_2^* + z_3 ET_3 p_3^* + z_4 ET_4 p_4^*}{ET_1(x) p_1^*(x) + ET_2 p_2^* + ET_3 p_3^* + ET_4 p_4^*}. \tag{4}$$

Mean value $ET_1(x)$ is calculate from the formula:

$$ET_1(x) = \int_0^x dF_1(t) + xP\{T_1 \ge x\}.$$

By integrating through parts, we get:

$$ET_1(x) = \int_0^x R_1(t)dt \quad .$$

Limit probabilities $p_1^*(x)$, $p_2^*(x)$, $p_3^*(x)$ are probabilities for Markov chain:

$$P(x) = \begin{bmatrix} 0 & p_{12}(x) & p_{13}(x) & p_{14}(x) \\ p_{21} & 0 & p_{23} & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad .$$

On the basis of the formulas (1) may be written as:

$$\begin{gathered} p_1^*(x) = 1 / M(x), \\ p_2^*(x) = p_{12}(x) / M(x), \\ p_3^*(x) = (p_{13}(x) + p_{12}(x)\, p_{23}) / M(x), \\ p_4^*(x) = p_{14}(x) / M(x), \end{gathered} \qquad (5)$$

where $M(x) = 2 + p_{12}(x)\, p_{23}$.

ET_2, ET_3 i ET_4 are mean values for times of the object remaining at states S_2, S_3 and S_4 of the system.

On the basis of (5), criterion function (4) is expressed through the formula:

$$g(x) = \frac{z_1 ET_1(x) + p_{12}F_1(x)z_2ET_2 + \left[p_{13}F_1(x) + p_{12}p_{23}F_1(x)\right]z_3ET_3 + \left[1 - p_{12}F_1(x) - p_{13}F_1(x)\right]z_4ET_4}{ET_1(x) + p_{12}F_1(x)ET_2 + \left[p_{13}F_1(x) + p_{12}p_{23}F_1(x)\right]ET_3 + \left[1 - p_{12}F_1(x) - p_{13}F_1(x)\right]ET_4} .$$

After rearrangement, one can write:

$$g(x) = \frac{z_1 ET_1(x) + F_1(x)\left[p_{12}z_2ET_2 + p_{12}p_{23}z_3ET_3 + p_{13}z_3ET_3 - p_{12}z_4ET_4 - p_{13}z_4ET_4\right] + z_4ET_4}{ET_1(x) + F_1(x)\left[p_{12}ET_2 + p_{12}p_{23}ET_3 + p_{13}ET_3 - p_{12}ET_4 - p_{13}ET_4\right] + ET_4} .$$

Now the numerator and denominator of the criterion function can be represented as:

$$\begin{gathered} L(x) = z_1\, ET(x) + F_1(x)\, B_1 + C_1, \\ M(x) = ET(x) + F_1(x)\, B + C. \end{gathered}$$

Analogically:

$$g(x) = \frac{z_1 ET_1(x) + F_1(x)B_1 + C_1}{ET_1(x) + F_1(x)B + C} ,$$

where:

$$\begin{gathered} B_1 = p_{12}\, z_2\, ET_2 + p_{12}\, p_{23}\, z_3\, ET_3 - p_{12}\, z_4\, ET_4 + p_{13}\, ET_3\, z_3 - p_{13}\, z_4\, ET_4, \\ C_1 = z_4\, ET_4, \\ B = p_{12}\, ET_2 + p_{12}\, p_{23}\, ET_3 - p_{12}\, ET_4 + p_{13}\, ET_3 - p_{13}\, ET_4, \\ C = ET_4. \end{gathered}$$

After simple transformations we get:

$$\begin{gathered} B_1 = p_{12}\, z_2 ET_2 + z_3 ET_3\, (p_{12}\, p_{23} + p_{13}) - z_4\, ET_4 (p_{12} + p_{13}), \\ B = p_{12}\, ET_2 + ET_3\, (p_{12}\, p_{23} + p_{13}) - ET_4\, (p_{12} + p_{13}). \end{gathered}$$

The following symbols are introduced:

$$\begin{gathered} \alpha = -B\, z_1 + B_1, \\ \beta = C\, z_1 - C_1, \\ \gamma = CB_1 - C_1B. \end{gathered}$$

The α, β and γ rates play an important role in formulating conditions sufficient for the occurrence of extremes of criterion functions.

3. Conditions for occurrence criterion function maximums

The assumptions for the parameters of the tested system are formulated below. These assumptions must reflect the actual relationship between perfect repairs, minimal repairs and preventive replacements:

Z1. $z_1 > 0$, $z_2 < 0$, $z_3 < 0$, $z_4 < 0$. The last element means that the technical object brings profit only at the S_1 state, while the remaining states require expenses.

Z2. $ET_3 > ET_4$, mean time of replacement (prevention) is shorter than mean time of perfect repair.

Z3. $z_3 < z_4$, unit cost $(-z_4)$ of replacement (prevention) is higher than unit cost $(-z_3)$ of perfect repair.

Z4. $ET_3 > ET_2$, mean time of minimal repair is shorter than mean time of perfect repair.

Z5. $z_3 < z_2$, unit cost $(-z_2)$ of minimal repair is higher than unit cost $(-z_3)$ of perfect repair.

The above assumptions do not include the relationship between the state of minimal repair and the state of preventive replacement. In practice, it is not known what is the relationship between the average values of ET_2 and ET_4 or z_2 and z_4. However, if $z_2 - z_4 \leq 0$, then on the basis of the assumption Z3 is $\gamma < 0$.

Below, sufficient conditions are formulated for the following inequalities to be true $\alpha < 0$, $\beta > 0$, $\gamma < 0$. The above conditions are formulated depending on mean times ET_i, costs z_i, i = 1, 2, 3, 4 as well as elements of matrix $P = [p_{ij}]$, i, j = 1, 2, 3, 4. It is relatively easy to calculate that $\beta = ET_4\, (z_1 - z_4)$. The assumption Z1 results in $\beta > 0$.

Rate α is expressed through formula:

$$\alpha = p_{12}\, ET_2 (z_2 - z_1) + (p_{12}\, p_{23} + p_{13})\, ET_3\, (z_3 - z_1) + ET_4\, (p_{12} + p_{13})\, (z_1 - z_4). \quad (6)$$

Inequality $\alpha < 0$ is equivalent to the inequality

$$(p_{12}\, p_{23} + p_{13})\, ET_3 > p_{12}\, ET_2 (z_2 - z_1) / (z_1 - z_3) + ET_4\, (p_{12} + p_{13})\, (z_1 - z_4) / (z_1 - z_3). \quad (7)$$

Rate γ is similarly determined

$$\gamma = ET_2[ET_4\, p_{14}\, (z_2 - z_4) + ET_3\, (p_{12}\, p_{23} + p_{13})\, (z_3 - z_4)]. \quad (8)$$

Inequality $\gamma < 0$ is equivalent to the inequality

$$(p_{12}\, p_{23} + p_{13})\, ET_3 > ET_4\, p_{14}\, (z_2 - z_4) / (z_4 - z_3). \quad (9)$$

Let us mark the right side of the inequality (7) and (9) with δ_1 i δ_2 respectively. Let $\delta = \max\{\delta_1, \delta_2\}$. The condition $(p_{12}\, p_{23} + p_{13})\, ET_3 > \delta$ and (6), (7), (8) and (9) is implicated by inequalities $\alpha < 0$, $\gamma < 0$. Now, the following conclusion may be drawn:

Conclusion 1. If $p_{23} > (\delta / ET_3 - p_{13}) / p_{12}$, then the inequalities $\alpha < 0$, $\gamma < 0$ are true.

In the literature on minimal repairs [2, 13, 14], it is assumed that if a technical object passes from the state of operation to the state of failure, the state of minimal repair is achieved with probability equal to 1 – p, and the state of exact repair with probability p. On the basis of elementary properties of conditional probability, the following equality is true:

$$p_{12} / p_{13} = (1 - p) / p. \quad (10)$$

Conclusion 2. If Tϵ IFR, $\lambda(t)$ is differentiable, $\alpha < 0$, $\gamma < 0$, $\beta > 0$, $\beta + \gamma f(0^+) > 0$, $\lambda(\infty)\,\alpha\,ET + \beta - \alpha < 0$, then the criterion function g(x) reaches maximum value.
Proof.

Function derivative g'(x) has the following form:

$$g'(x)= \{\alpha[f(x)\,ET(x) - R(x)\,F(x)] + \beta\,R(x) + \gamma\,f(x)\} / M^2(x),$$

where M(x) is the denominator of criterion function g(x).

It is known that if time before failure T belongs to the class of MTFR distributions, then the equality $H(x) = \lambda(x)\,ET(x) - F(x)] \geq 0$ for $x \geq 0$. The class of MTFR distributions was tested in papers [9, 10]. The MTFR class includes some lifetimes with single-mode function of failure intensity [9, 10]. From the fact that derivative H'(x) has the form $H'(x) = \lambda'(x)\,ET(x)$ it follows that if the function of intensity of failure $\lambda(t)$ increases, the function H(x) also increases. The class of distributions with the increasing failure rate function (IFR) is included in the MTFR class. The symbol of the derivative is the same as the symbol of the function:

$$h(x) = \alpha[\lambda(x)\,ET(x) - F(x)] + \beta + \gamma\,\lambda(x).$$

It is known that $H(0^+) = 0$, hence $h(0^+) = \beta + \gamma f(0^+) > 0$. From the fact that $\alpha < 0$, $\beta > 0$, $\gamma < 0$ and function H(x) increases it follows that the function h(x) decreases from $(0^+) = \beta + \gamma f(0^+) > 0$ to $h(\infty) = \lambda(\infty)\alpha\,ET + \beta - \alpha < 0$. It follows that derivative function g'(x) changes from a „+" to a „– " exactly once. Hence it is concluded the criterion function g(x) reaches precisely one maximum. □

If $\lambda(\infty) = \infty$, then for the occurrence of criterion function g(x) the following conditions are sufficient: Tϵ IFR, differentiability $\lambda(t)$, $\alpha < 0$, $\gamma < 0$, $\beta > 0$, $\beta + \gamma f(0^+) > 0$. One example is Weibull distribution with increasing failure intensity function.

Conclusions 1 and 2 result in the following sufficient condition for the occurrence of maximum criterion function:

Conclusion 3. If Tϵ IFR, $\lambda(t)$ is differentiable, $\beta + \gamma\,\lambda(0^+) > 0$, $p_{23} > (\delta / ET_3 - p_{13}) / p_{12}$, $\lambda(\infty)\,\alpha\,ET + \beta - \alpha < 0$, then criterion function g(x) reaches maximum value.

Sufficient condition for occurrence of an asymptotic maximum availability rate is formulated below. In order to obtain availability rate from criterion function g(x), it is sufficient to assume the following conditions: $z_1 = 1$, $z_2 = z_3 = z_4 = 0$. After taking into account these conditions, in formula (4) we get $B_1 = 0$, $C_1 = 0$. Hence, on the basis of (2), (3) and (5) for α, β, γ the following can be calculated:

$$\alpha = - B = - p_{12}\,ET_2 - (p_{12}\,p_{23} + p_{13})\,ET_3 + (p_{12} + p_{13})\,ET_4,$$
$$\beta = ET_4,$$
$$\gamma = 0.$$

Inequality $\alpha < 0$ is equivalent to inequality

$$p_{23} > \{[ET_4(1 + p_{13} / p_{12}) - ET_2] / ET_3\} - p_{13} / p_{12}.$$

The last inequality, taking into account (7), may be written in the following form:

$$p_{23} > \{[ET_4 / (1 - p) - ET_2] / ET_3\} - p / (1 - p).$$

Taking into account the fact that $\beta > 0$ i $\gamma = 0$, we can now formulate the sufficient condition for the occurrence of maximum availability rate.

Conclusion 4. If Tϵ IFR, $\lambda(t)$ is differentiable, $\lambda(\infty)\,\alpha\,ET + \beta - \alpha < 0$, $p_{23} > \{[ET_4 / (1 - p) - ET_2] / ET_3\} - p / (1 - p)$, then the availability rate reaches precisely one maximum value.
Proof.

Function derivative g'(x) has the following form: $g'(x)= \{\alpha[f(x)\,ET(x) - R(x)\,F(x)] + \beta\,R(x)\} / M^2(x)$, where M(x) is the denominator for criterion function g(x).

If failure intensity function $\lambda(t)$ increases, then function H(x) increases. Symbol of the derivative is identical with symbol of function $h(x) = \alpha[\lambda(x)\,ET(x) - F(x)] + \beta$. It is known that $H(0^+) = 0$, hence $h(0^+) = \beta > 0$.

From the fact that $p_{23} > \{[ET_4 / (1 - p) - ET_2] / ET_3\} - p / (1 - p)$, it follows that $\alpha < 0$ and function h(x) decreases from value $h(0^+) = \beta > 0$ to values $h(\infty)$. If $h(\infty) = \lambda(\infty)\alpha\,ET + \beta - \alpha < 0$, then derivative function g'(x) changes from a „+" to a „–" exactly once. Hence it is concluded the criterion function g(x) reaches precisely one maximum. □

If $\lambda(\infty) = \infty$, then for the occurrence of maximum availability rate the following conditions are sufficient Tϵ IFR, $p_{23} > \{[ET_4 / (1 - p) - ET_2] / ET_3\} - p / (1 - p)$.

4. Numerical examples

Example 1. In this example the value of the function g(x) is determined, when g(x) is the availability rate. The following data were used in the calculations: mean values of times of the technical object remaining at states $ET_2 = 0.2$, $ET_3 = 0.5$, $ET_4 = 0.1$, assumed for Weibull distribution of time before failure T with the scale parameter b = 6. Non-zero elements were assumed for matrix P as $p_{12} = 0.2$, $p_{13} = 0.6$, $p_{14} = 0.2$, $p_{21} = 0.2$, $p_{23} = 0.7$. The values of the c form (shape) parameter for Weibull distribution c ε {5, 6, 7} were assumed. In each of the three analyzed cases, there is an optimal value for the replacement time.

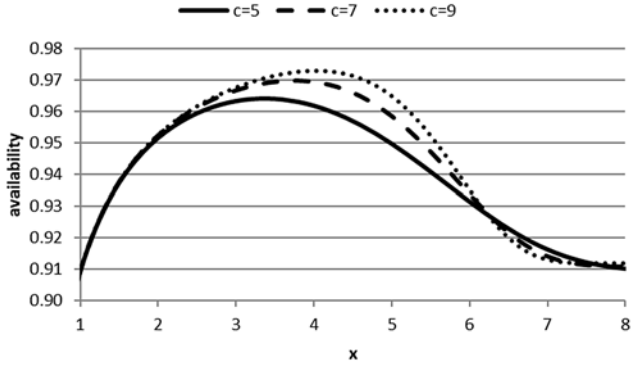


Fig. 2. Charts of changes in the value of availability rate depending on the time of preventive replacement x, for c ϵ {5, 6, 7)

Example 2. In this example the value of the function g(x) is determined, when g(x) is profit per time unit. The calculations included mean values for times of remaining at states, probability matrix P and the parameter of Weibull distribution scale the same as in Example 1. The values of parameter c form (shape) for Weibull distribution cϵ{5, 6, 7} were assumed. For calculations, unit profits were assumed as $z_1 = 6$, $z_2 = - 0.1$, $z_3 = - 0.8$, $z_4 = - 0.2$.

For all values of parameter c of the form of Weibull distribution, the criterion function reaches the maximum value. Analysis of the dependence of the point x_{max}, in which the criterion function g(x) reaches maximum value shows that as the value of parameter c increases,

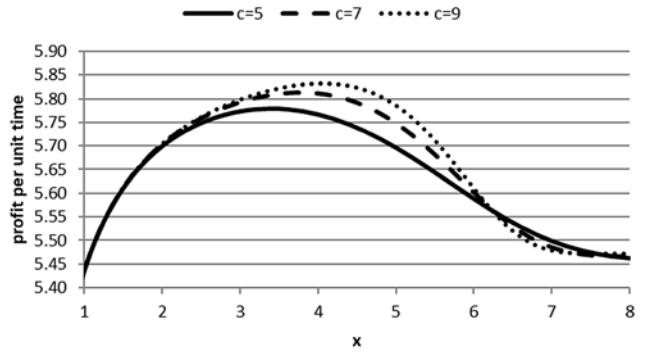


Fig. 3. Charts of changes in the value of profit per unit time rate depending on the time of preventive replacement x, for c ϵ {5, 6, 7)

the value of x_{max} and the maximum value of the criterion function increase.

6. Conclusions

Maintenance systems performing two types of repairs: minimal repairs and perfect repairs are covered by a wide range of literature. However, the use of semi-Markov processes is rare. This paper shows that the application of semi-Markov processes in determining optimal strategies for preventive actions in systems with minimal repair allows for the formulation of interesting conclusions. For the criterion functions analyzed in this paper (availability and profit per unit of time), sufficient conditions for the occurrence of maximum of these criterion functions were formulated. Criterion functions are analyzed in an infinite time horizon. Formulating stronger conditions requires establishing the relation between the mean times of technical object remaining at individual states as well as unit profits at states of minimal repair and preventive replacement.

References

1. Block H W, Borges W, Savits T H. Age-dependent minimal repair. Journal of Applied Probability 1985; 22: 370-386, https://doi.org/10.2307/3213780.
2. Brown M, Prochan F. Imperfect Repair. Journal of Applied Probability 1983; 20: 851-859, https://doi.org/10.2307/3213596.
3. Chan P K W, Downs T. Two criteria for preventive maintenance. IEEE Transaction on Reliability 1978; 27: 272-273, https://doi.org/10.1109/TR.1978.5220367.
4. Chang C C, Sheu S H, Chen Y L, Zhang Z G. A multi-criteria optimal replacement policy for a system subject to shocks. Computer and Industrial Engineering 2011; 61: 1035-1043, https://doi.org/10.1016/j.cie.2011.06.017.
5. Chang C C, Sheu S H, Chen Y L. A bivariate optimal replacement policy for system with age-dependent minimal repair and cumulative repair-cost limit. Communications in statistics – Theory and methods 2013; 42(22): 4108-4126, https://doi.org/10.1080/03610926.2011.648789.
6. Chin-Chih, Chang. Optimum preventive maintenance policies for system subject to random working time, replacement, and minimal repair. Computers & Industrial Engineering 2014; 67: 185-194, https://doi.org/10.1016/j.cie.2013.11.011.
7. Grabski F. Semi-markowskie modele niezawodności i eksploatacji [Semi-Markov models of reliability and maintenance]. Warszawa: IBS PAN, 2002.
8. Grabski F. Semi-Markov Processes: Applications in System Reliability and Maintenance. Elsevier, Amsterdam, 2014.
9. Knopik L. Some results on ageing class. Control and Cybernetics 2005; 34(4): 1175-1180.
10. Knopik L. Characterization of a class of lifetime distributions. Control and Cybernetics 2006; 35(2): 1175-1180.
11. Knopik L, Migawa K. Multi-state model of maintenance policy. Eksploatacja i Niezawdonosc - Maintenance and Reliability 2018; 20(1): 125-130, https://doi.org/10.17531/ein.2018.1.16.
12. Knopik L, Migawa K. Optimal age-replacement policy for non-repairable technical objects with warranty. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19(2): 172-178, https://doi.org/10.17531/ein.2017.2.4.
13. Pham H, Wang H. Imperfect maintenance. European Journal of Operational Research 1996; 94: 425-438, https://doi.org/10.1016/S0377-2217(96)00099-9.
14. Pham H, Wang H. Probability and Optimal Maintenance. Springer Verlag, London, 2006.
15. Tadj L, Ouali M S, Yacount S, Ait-Kadi D. Replacement Models with minimal Repair, Chapter: A Survey of Replacement Models with Minimal Repair, Springer Verlag, London, 2011, https://doi.org/10.1007/978-0-85729-215-5.
16. Xia, Xi, Zhou. Modeling and optimizing maintenance schedule for energy systems subject to degradation. Computer and Industrial Engineering 2012; 63(3): 607-614, https://doi.org/10.1016/j.cie.2012.01.014.
17. Xu, Chen, Yang. Optimal replacement policy for safety-related multi-component multi-state systems. Reliability Engineering and System Safety 2012; 99: 87-95, https://doi.org/10.1016/j.ress.2011.11.010.

Leszek KNOPIK
Faculty of Management
UTP University of Science and Technology
ul. Fordońska 430, 85-890 Bydgoszcz, Poland

Klaudiusz MIGAWA
Faculty of Mechanical Engineering
UTP University of Science and Technology
Al. prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland

E-mails: knopikl@utp.edu.pl, klaudiusz.migawa@utp.edu.pl

Article citation info:
RYMARCZYK T, KŁOSOWSKI G. Innovative methods of neural reconstruction for tomographic images in maintenance of tank industrial reactors. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 261–267, http://dx.doi.org/10.17531/ein.2019.2.10.

Tomasz RYMARCZYK
Grzegorz KŁOSOWSKI

INNOVATIVE METHODS OF NEURAL RECONSTRUCTION FOR TOMOGRAPHIC IMAGES IN MAINTENANCE OF TANK INDUSTRIAL REACTORS

NOWATORSKIE METODY NEURONOWEJ REKONSTRUKCJI OBRAZÓW TOMOGRAFICZNYCH W EKSPLOATACJI ZBIORNIKOWYCH REAKTORÓW PRZEMYSŁOWYCH*

The article presents an innovative concept of improving the monitoring and optimization of industrial processes. The developed method is based on a system of many separately trained neural networks, in which each network generates a single point of the output image. Thanks to the elastic net method, the implemented algorithm reduces the correlated and irrelevant variables from the input measurement vector, making it more resistant to the phenomenon of data noises. The advantage of the described solution over known non-invasive methods is to obtain a higher resolution of images dynamically appearing inside the reactor of artifacts (crystals or gas bubbles), which essentially contributes to the early detection of hazards and problems associated with the operation of industrial systems, and thus increases the efficiency of chemical process control.

Keywords: *electrical tomography, industrial processes, process control, neural networks, machine learning.*

W artykule przedstawiono nowatorską koncepcję usprawnienia monitoringu i optymalizacji procesów przemysłowych. Opracowana metoda bazuje na systemie osobno wytrenowanych wielu sieci neuronowych, w którym każda sieć generuje pojedynczy punkt obrazu wyjściowego. Dzięki zastosowaniu metody elastic net zaimplementowany algorytm redukuje z wejściowego wektora pomiarowego zmienne skorelowane i nieistotne, czyniąc go bardziej odpornym na zjawisko zaszumienia danych. Przewagą opisywanego rozwiązania nad znanymi metodami nieinwazyjnymi jest uzyskanie wyższej rozdzielczości obrazów dynamicznie pojawiających się wewnątrz reaktora artefaktów (kryształów lub pęcherzy gazowych), co zasadniczo przyczynia się do wczesnego wykrycia zagrożeń i problemów związanych z eksploatacją systemów przemysłowych, a tym samym zwiększa efektywność sterowania procesami chemicznymi.

Słowa kluczowe: *tomografia elektryczna, procesy przemysłowe, sterowanie procesami, sieci neuronowe, uczenie maszynowe.*

1. Introduction

Many chemical engineering products are created as a result of processes carried out using technological lines in which chemical reactors play a key role. A chemical reactor is a vessel adapted to carry out reactions taking place inside it. The purpose of industrial tank reactors is to ensure optimal economic parameters of chemical processes [18]. It can be achieved by the appropriate reactor design and by the skillful overlap of the three types of sub-processes occurring inside the reactor, namely the transfer of mass, momentum and heat. In this way, process control can be based on a dynamic selection of parameters such as: mixing intensity, temperature, pressure, substrate ratios and others. The presented research included reactors in which reactions take place between solid and liquid as well as gas and liquid. The tank reactor diagram is shown in Fig. 1.

The first type of reaction concerns the crystallization of solids in a liquid environment. It covers industrial processes of synthesis and purification of solid substances and changes in particle properties. The reactors in which crystallization occurs are used in many branches of the economy, including: chemical, food [9], metallurgy and waste utilization [27]. One of the monitoring systems tasks of such processes is to provide precise information on the quantity, size and location of crystals formed in the liquid in real time.

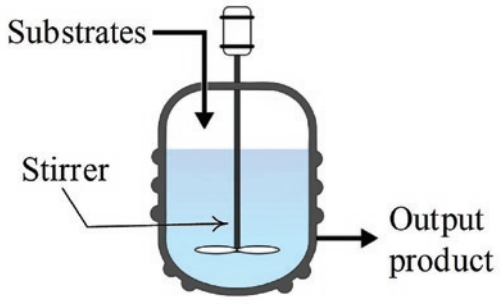


Fig. 1. Diagram of a tank reactor with a mixing system

The second type of reaction relates to the gas phase and the liquid phase. Such processes type are used, inter alia, in the production of biogas. Physicochemical fermentation reactors are a key element of biogas installations. Methane fermentation of organic waste takes place inside these reactors. The correct operation of technical systems is one of the crucial conditions for obtaining an adequate level of reliability of industrial processes [10].

There are two main reasons to monitor the state of dynamic processes. The first one is the detection of impending failures [11] which

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

can include damage to technical infrastructure, excessive deviation of critical process parameters or interruption of its continuity. An effective monitoring system is designed to identify the problem early enough for effective corrective action to be taken.

The second reason for monitoring industrial processes is the need to control the course of the industrial process [29]. This is necessary for ensuring an adequate level of quality. In order to effectively control multiphase processes in which participate substances that can dynamically change the states of aggregation, effective monitoring methods should be applied. This is a difficult task, taking into account the aggressive conditions in which reactions take place inside the reactor. Using invasive sensors, the problem is the inability to directly examine any part of the interior of the reactor, point accuracy of measurements, the need to use many different monitoring systems simultaneously and high uncertainty in determining the dynamic state of the process based on incomplete data (indirect method). Among the non-invasive methods used in monitoring industrial processes can be distinguished: electric capacitive tomography [2, 4, 5, 13, 14, 16, 20, 26] electrical impedance tomography [3, 8, 28], magnetoacoustic tomography [30], ultrasound and radio tomography [21], X-ray tomography [1] and many more. Recently, more and more research in the field of industrial systems takes into account the use of various computational methods, such as: intelligent prediction methods [25], fuzzy logic [6], machine learning [22], numerical modeling [15], deep learning [8, 19] and binary programming [12].

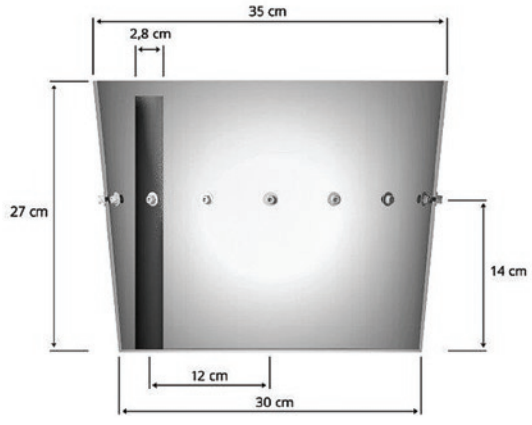
Currently used techniques of non-invasive monitoring of industrial processes do not fully meet current operational needs. Obtained mapping of images of studied phenomena and processes can be unsharp, ambiguous, difficult to interpret, burdened with inaccuracies both in terms of the number of artefacts (crystals or gas bubbles) detected in the reactor as well as their size and location. As a result, redundant systems are used to obtain precise information about the status of the monitored process, which significantly increases the operating costs.

The above-mentioned difficulties and imperfections of the monitoring methods of chemical tank reactors are the reason for their improvement. The use of an improved monitoring method will increase the reliability of processes occurring inside the reactors and will reduce the operating costs of industrial systems.

The aim of this study is to introduce an improved method of monitoring and optimization of chemical processes occurring in heterogeneous tank reactors in which reactions take place between solid and liquid as well as gas and liquid. The applied method concerns electrical tomography [22]. The innovation is an original method of parallel use of a hybrid system that is a combination of the elastic net method and artificial neural networks [7, 23]. An advantage of the described concept over other known non-invasive methods is increased resistance to interference during measurements, higher accuracy of reconstruction, unlimited imaging resolution, low cost and high speed of operation. The description of the multiply neural system enabling effective monitoring of chemical reactions using electrical tomography is presented in the further part of this study.

Fig. 2. The test stand: a) hybrid tomograph, b) reactor with connected electrodes

a)



b)

Fig. 3. The physical EIT model with the electrodes: a) schematic drawing, b) reactor with artifacts immersed in liquid electrodes

2. Models, methods, algorithms

Electrical impedance tomography (EIT), like as electrical capacitance tomography (ECT), are non-invasive diagnostic methods of technical objects [15, 22]. The EIT method used in the described studies processes data generated by a system of 16 electrodes installed on the surface of the reactor. Fig. 2 and 3 present a research position with a physical model with an electrode system and a hybrid tomograph adapted for EIT measurements.

Fig. 4 presents the method of measuring the voltages generated by the arrangement of 16 electrodes. Due to the unknown value of voltage drops between the tested object and the electrodes to which the electric current source (I) is connected, these electrodes are not taken into account during the measurements. For each projection angle 12 independent voltage measurements (V) can be obtained between individual pairs of adjacent electrodes. This number results from the following calculation: n-4 = 12, where n = 16 is the total number of electrodes in the system, while 4 is the number of electrodes excluded from measurements within the projection angle. In this way, the total number of measurements is $(n-4)(n/2)=12\times 8=96$.

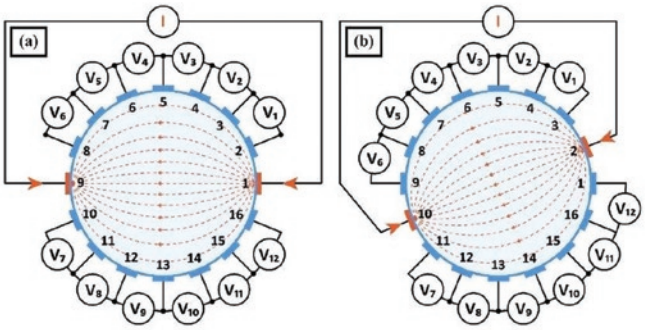


Fig. 4. The method of voltage measurement in a system of 16 electrodes: a) first measuring cycle, b) next measurement cycle

2.1. Hybrid neural algorithm

Dynamic progress in the field of low-cost technological solutions and access to advanced computational methods cause that the costs of using computing power and storage media are falling [16]. The popularity of computational techniques using parallel computations and requiring the processing of large data sets is growing [3]. As a result, there are less and less frequent situations in which researchers and designers of tomographic systems are forced to shorten the calculation time at the cost of the quality of results [17]. The presented tomographic system converts a vector of 96 voltages into an image with a resolution of 2883 pixels. Because the designed algorithm uses 2883 parallel-acting multiply neural networks, computing power is an important factor determining the speed of the measurement system. Fig. 5 shows the model of one of the 2883 hybrid subsystems to generate the actual value of a single point of the reconstructed image.

The algorithm for training a hybrid multiply neural system has the following sequence:

1. set-up the initial conditions:
   a) number of pixels in the mesh of the output image: m=2883;
   b) the size of the measurement vector $\boldsymbol{X_n}$ for each of the reconstructions: n=96;
   c) the number of cases that is required for neural networks training: N=99900;
   d) neural networks structure: $\tilde{n}_\upsilon$ -10-1, where $\tilde{n}_\upsilon$ means the reduced by the elastic net number of measurements at the input of the neural network generating υ-th pixel, 10 neurons in the hidden layer and 1 neuron in the output layer. Logistic transfer functions were used in both layers;
2. using the elastic net method, generate a matrix of individually reduced measurement vectors for each pixel of the output image separately:
   $\boldsymbol{X_n} \rightarrow elastic\,net \rightarrow \boldsymbol{X}_{\tilde{n},\upsilon}$, where $n=96$, $n>\tilde{n}>0$. You can use a subset with a number of cases much smaller than N, for example, 3000 randomly selected cases from the training set;
3. for υ=1 to m train neural network ANN_υ using the training set of the cardinality N;
4. save trained 2883 neural networks to the structural variable.

The mentioned above algorithm was implemented using the Matlab code, while the grid of the output image was modeled using the Eidors toolbox. Table 1 presents the results of the training process of one of 2883 neural networks, generating a real number determining the color of an example pixel on a tomographic image mesh. The pixel 600 was randomly chosen for the following test. Input data are included in the matrix $\boldsymbol{X_n}$ with dimensions of 96x99900 (96 measurements, 99900 cases). The input file was a vector $\boldsymbol{Y_\upsilon}$ with dimensions of 1x99900 (a single pixel of the image υ = 600, 99900 cases). Before training the network, the input vector was reduced from $n=96$ to $\tilde{n}=30$ using the elastic net method. The structure of the single neural network for point 600 was therefore: ANN_{600} = 30-10-1.

Table 1. The results of the training process along with the division of data

Division of the data set	The number of cases in a given set	Mean Square Error (MSE)	Regression (R)
1	2	3	4
Training set (70%)	69930	$7.65053\cdot10^{-3}$	0.813877
Validating set (15%)	14985	$7.27605\cdot10^{-3}$	0.806954
Testing set (15%)	14985	$8.27152\cdot10^{-3}$	0.822387

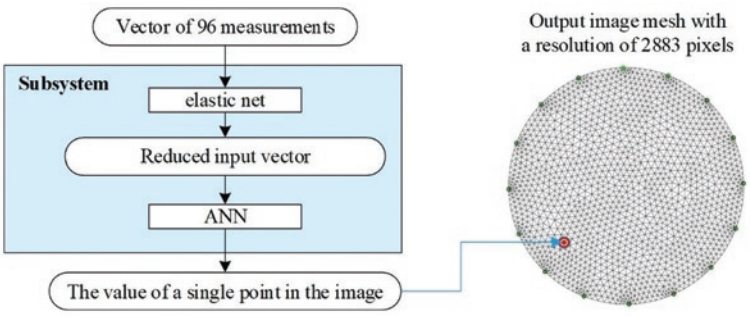


Fig. 5. The model of one of 2883 subsystems generating a single point of the image

A simulated collection of 99900 data have been divided into 3 parts: training, validating and testing in proportions 70/15/15, which is reflected in the first two columns of Table 1. Columns 3 and 4 contain information about error MSE and regression R for all 3 subsets that have been used in the network learning

process. The low value of MSE and R value close to 1 testify to the good quality of the trained network.

Formula (1) presents the method of calculating MSE:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i' - y^*_i\right)^2 \quad (1)$$

where: n – image resolution; y_i' – the reference value of the i-th pixel; y^*_i – the reconstructive value of the i-th pixel.

The method of calculating the regression coefficient R is the formula (2):

$$R\left(y', y^*\right) = \frac{cov\left(y', y^*\right)}{\sigma_{y'}\sigma_{y^*}} \quad (2)$$

where: $\sigma_{y'}$ – standard deviation of reference values, σ_{y^*} – standard deviation of reconstructed values.

Fig. 6a shows the training progress of the selected ANN. The homogeneous hyperbolic shape of the graph line indicates that the network has not been overtrained. The lack of fluctuations and line compliance for all 3 data sets (learning, validation and test) indicates lack of overfitting and the network's ability to generalize. The graph also shows the moment when network learning process stopped. It happened after the 53-th epoch, when the MSE of the validating set reached the value of 0.0072761. The same value can also be read from column 3 of Table 1. The validating set was used to determine the condition of stop training ANN. This condition is met if the MSE of validation set does not decrease for 6 successive epochs.

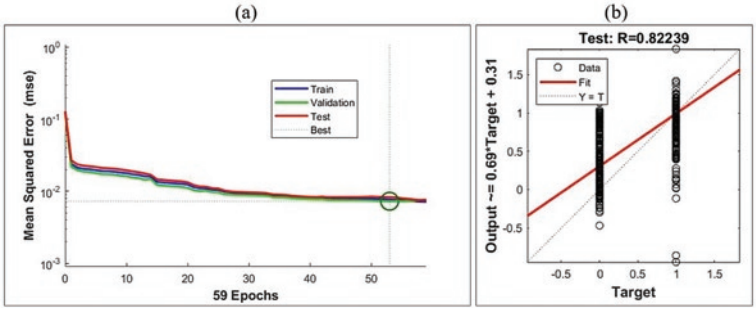


Fig. 6. Training process results for the selected network: a) MSE charts for the training, validating and testing set, b) regression statistics for the testing set

Information on the trained network is supplemented by Fig. 6b, which shows the graph of regression statistics for the testing set. The testing set gives the most reliable results regarding the quality of the received network, because the data contained therein did not have any impact on the learning process. As you can see, R = 0.82239 is consistent with the information from Table 1 contained in the bottom row of column 4. The characteristic way of data distribution in Fig. 6b results from the fact that the reference output image reflects only 2 values: 1 for background color and 0 for artifact (crystals or gas bubbles).

2.2. Improving the quality of measurement data using the elastic net

In the case of reconstruction of tomographic images of real objects with relatively low conductivity, the data from the electrodes are usually noisy. This is the result of the imperfection of electrode insulation, the effects of fast-changing and low currents generated by multiplexers, the influence of electromagnetic fields and many other factors. An example of technical facilities, from which tomographic data show a high level of noise are also industrial chemical reactors [18]. Disturbances of electrical signals are one of the main barriers hindering the development of tomographic methods for such objects [23].

To make the input data resistant to distortions and noise, the elastic net regularization was used [22]. In this method, we assume a certain linear system that can be described by means of the state equation (3):

$$Y = X\beta + \varepsilon \quad (3)$$

where $Y \in R^n$ is a matrix of output variables (reconstruction), $X \in R^{n\times(k+1)}$ represents the matrix of input variables, the coefficient $\beta \in R^{k+1}$ means a vector with unknown parameters, and $\varepsilon \in R^n$ reflects the sequence interference. If the linear task (4) has a solution in which the regression line intersects the y axis, then the first column of the matrix X in the linear equation (3) is a unitary column vector.

In the case when the predictors introduced into the regression model are strongly correlated with each other, a possible way to determine linear regression is to solve the problem (4):

$$\min_{(\beta_0,\beta')\in R^{k+1}} \frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \beta_0 - x_i\beta'\right)^2 + \lambda P_\alpha\left(\beta'\right) \quad (4)$$

where: $x_i = \left(x_{i1},\ldots,x_{ik}\right)$, $\beta' = \left(\beta_1,\ldots,\beta_k\right)$ dla $1 \le i \le n$

P_α – elastic net penalty function is defined by the equation (3):

$$P_\alpha\left(\beta'\right) = \left(1-\alpha\right)\frac{1}{2}\left\|\beta'\right\|_{L_2} + \alpha\left\|\beta'\right\|_{L_1} = \sum_{j=1}^{k}\left(\frac{1-\alpha}{2}\beta_j^2 + \alpha\left|\beta_j\right|\right) \quad (5)$$

It can be seen that the penalty in this case is a linear combination of the L_1 and L_2 norms from unknown β' parameters. The introduction of the parameter-dependent penalties function to the objective function reduces the predictions of unknown parameters. The parameter λ in the task (5) represents the penalty factor. Entering the $0 \le \alpha \le 1$ parameter causes the task (4) to compromise between Lasso (least absolute shrinkage and selection operator) and ridge regression. If $\alpha = 0$ we deal with pure ridge regression called Tikhonov regularization. Ridge regression is a popular method of regularization of linear models [22].

If $\alpha = 1$ there is a pure Lasso method in task (4). Lasso is a statistical method that allows selection of independent variables

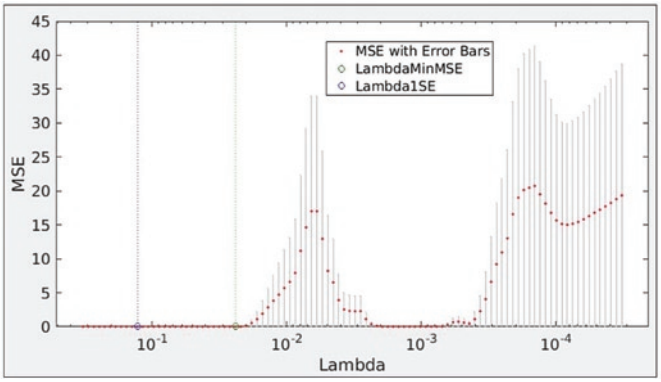


Fig. 7. Selection of Lambda parameter in the elastic net method with α = 0.25 based on the minimum mean square error (MSE)

and regularization of linear models. In the case of ridge regression, the coefficient of penalties is calculated in the L_1 norm and for Lasso in L_2 norm. Lasso is also indifferent to the correlation of predictors.

Figure 7 presents the process of selecting the parameter λ (lambda) in the elastic net method with the assumed α = 0.25 for one of the pixels of the output image. Two lambda values are marked with a green and blue dotted lines. The green line indicates the lambda value with the minimum mean square error (MSE) obtained using the cross-validation. This value has been marked with the LambdaMinMSE variable. The blue dotted line indicates the largest Lambda value, which is within the standard deviation of the minimum MSE (Lambda1SE). This lambda value means that even the worst of the tested models (with the largest MSE) still has a relatively low error value. In the described case λ = 0.129 for MSE = 0.0154 was assumed. The measurement vector has 96 values of voltage drops. In the considered case, thanks to the use of elastic net, the input vector has been reduced from 96 to 30 measurements.

3. Results

Fig. 8 presents a comparison of tomographic images obtained with two methods: ANN and a hybrid method combining elastic net with ANN.

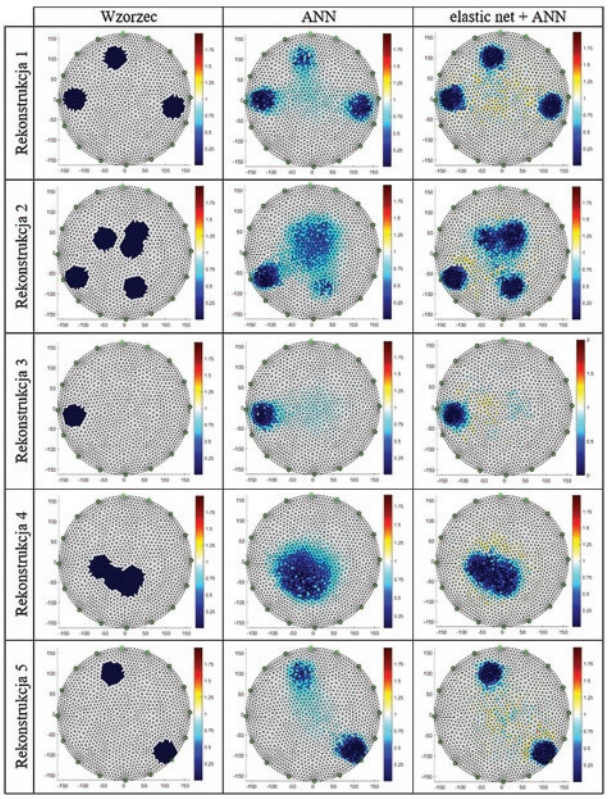


Fig. 8. Comparison of reconstructive images for ANN methods and "elastic net + ANN"

The first column contains reference images for the five tested cases. The test stand was calibrated in such a way that the value in the output, corresponding to the conductance of the reference liquid (water) equals 1. The values of the image pixels corresponding to the artefacts is 0. Next, using the physical model, real measurements were made by the ANN and elastic net + ANN (hybrid algorithm).

By comparing the images obtained with two methods, it can be noticed that in the hybrid approach, despite the reduction of the input measurement vector by almost 70% for each neural network, the visual quality of reconstruction is not worse than for pure ANN. It should also be noted that the algorithm reconstructs the worst artifacts located near the center of the reactor. This is most likely related to the distance of the recognized objects from the electrodes. Better reconstructions are obtained for artifacts that are localized closer to the electrodes.

Visual evaluation of reconstruction is not precise and insufficient for an objective comparison of the two analyzed methods. For this reason, three quantitative imaging quality indicators were introduced: MSE (Mean Squared Error), RIE (Relative Image Error) and ICC (Image Correlation Coefficient).

The relative RIE imaging error is calculated according to the formula (6):

$$RIE = \frac{\left\| Y^* - Y' \right\|}{\left\| Y' \right\|} \tag{6}$$

where: Y' – probability distribution of the pattern image pixels; Y^* – values of reconstructed pixels.

Formula (7) shows how to calculate the ICC image correlation coefficient:

$$ICC = \frac{\sum_{i=1}^{l} \left(y_i^* - \bar{Y}^* \right) \left(y_i' - \bar{Y}' \right)}{\sqrt{\sum_{i=1}^{l} \left(y_i^* - \bar{Y}^* \right)^2 \sum_{i=1}^{l} \left(y_i' - \bar{Y}' \right)^2}} \tag{7}$$

where: $\bar{Y}'$ – average probability distribution of the pattern image pixels; $\bar{Y}^*$ – average value of reconstructed pixels.

The smaller are the values of MSE and RIE coefficients, the better is the reconstruction quality. In the case of ICC, the opposite is true - the closer 1, the better the correlation of the output image with the reference image, which means a more accurate reconstruction.

Table 2 contains a summary of all three reconstruction quality indicators divided into tested methods and cases. The upper part of the table contains precisely calculated MSE, RIE and ICC values. In the right-hand column there are averaged values giving the possibility to make a generalized assessment of all five cases with one indicator. In order to facilitate the choice of the better method, the bottom part of the table contains a boolean classification of all indicators. Three questions have been formulated in such a way that in the case of an affirmative answer ("truth") a better method is "elastic net + ANN". In the case of a negative response ("false"), the index would prefer the pure ANN method.

As can be seen from Table 2, in all 18 cases, including responses and average values, the hybrid method „elastic net + ANN" turned out to be better.

4. Discussion and conclusions

The paper presents the results of research on the development of an improved algorithm for the reconstruction of images in the field of process tomography. In particular, the focus has been placed on imaging the cross-section of industrial tank reactors in which the processes of crystallization and anaerobic digestion occur in biogas installations. To make an objective assessment of the quality of the hybrid algorithm (elastic net + ANN), comparative studies were carried out using the physical laboratory model. Equal cases of artifact systems were prepared, and then voltage drops were read from the electrode system. Due to the fact that while reading data, many current-voltage states are transient, the data contained a certain level of noise. Af-

Table 2. Comparison of reconstruction indicators

Quality indicator of reconstruction		Number of reconstruction					Mean
		1	2	3	4	5	
ANN	MSE	0.0232	0.0573	0.0061	0.0337	0.0198	0.0280
	RIE	0.1587	0.2547	0.0795	0.1913	0.1451	0.1658
	ICC	0.8319	0.6715	0.8694	0.7301	0.7982	0.7802
elastic net + ANN	MSE	0.0158	0.0388	0.0047	0.0238	0.0115	0.0189
	RIE	0.1312	0.2096	0.0695	0.1610	0.1107	0.1364
	ICC	0.8804	0.7892	0.8987	0.8177	0.8907	0.8553
Is the MSE of the hybrid algorithm smaller than the MSE of pure ANN?		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Is the RIE hybrid algorithm smaller than RIE of pure ANN?		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Is the ICC of the hybrid algorithm greater than the ICC of pure ANN?		TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

ter filtering the data using the background reference values (reactor without artefacts) obtained results were illustrated and converted into quality indicators: MSE, RIE and ICC.

Normally, if the clear measurement data used for the imaging were simulated and did not contain interference, removing 70% of the predictors (eg, reduction from 96 to 30) would result in poor image quality. The carried out experiments showed that it was different in the case of data with a certain level of noise. Although the obtained reconstructions, both for pure ANN and for the hybrid method, are still not perfect, they are nevertheless sufficiently accurate to determine the amount, shape and location of the artifacts. The time of reconstruction in all investigated cases with the Intel Core i7 processor machine did not exceed 1 second. This means that the developed algorithm can also be used in processes with high reaction dynamics and even in flow systems [24].

It is significant that in Table 2, in all cases tested and for all three quality indicators, better results were obtained using the "elastic net + ANN" algorithm, not pure ANN. It turns out that the application of elastic net caused that along with the removal of 70% of input variables, a lot of noisy data was deleted, generating various types of disturbances in the output image. This fact is proof that the newly developed hybrid algorithm "cleanses" the data and makes the tomographic system immune to various types of interference and noisy data.

The results of the research have proved that the use of the described solution enables a higher resolution of images of crystals or gas bubbles appearing inside the reactor, which essentially contributes to the early detection of hazards and facilitates solving the operational problems of industrial systems.

Acknowledgements

the authors would like to thank the authorities and employees of the Faculty of Mathematics, Physics and Computer Science of UMCS in Lublin for sharing supercomputing resources.

References

1. Babout L, Grudzień K, Wiącek J, Niedostatkiewicz M, Karpiński B, Szkodo M. Selection of Material for X-Ray Tomography Analysis and DEM Simulations: Comparison between Granular Materials of Biological and Non-Biological Origins. Granular Matter 2018; 20 (3): 38, https://doi.org/10.1007/s10035-018-0809-y.
2. Banasiak R, Wajman R, Sankowski D, Soleimani M. Three-Dimensional Nonlinear Inversion of Electrical Capacitance Tomography Data Using a Complete Sensor Model. Progress In Electromagnetics Research (PIER) 2010; 100: 219-234, https://doi.org/10.2528/PIER09111201.
3. Dusek J, Hladky D, Mikulka J. Electrical Impedance Tomography Methods and Algorithms Processed with a GPU. Progress In Electromagnetics Research Symposium - Spring (PIERS) 2017; 1710–14, https://doi.org/10.1109/PIERS.2017.8262025.
4. Garbaa H, Jackowska-Strumiłło L, Grudzień K, Romanowski A. Application of Electrical Capacitance Tomography and Artificial Neural Networks to Rapid Estimation of Cylindrical Shape Parameters of Industrial Flow Structure. Archives of Electrical Engineering 2016; 65 (4): 657–69, https://doi.org/10.1515/aee-2016-0046.
5. Grudzien K, Chaniecki Z, Romanowski A, Sankowski D, Nowakowski J, Niedostatkiewicz M. Application of Twin-Plane ECT Sensor for Identification of the Internal Imperfections inside Concrete Beams. IEEE International Instrumentation and Measurement Technology Conference Proceedings 2016; May, 1–6, https://doi.org/10.1109/I2MTC.2016.7520512.
6. Kłosowski G, Gola A, Świć A. Application of Fuzzy Logic Controller for Machine Load Balancing in Discrete Manufacturing System. In International Conference on Intelligent Data Engineering and Automated Learning 2015; 256–63, https://doi.org/10.1007/978-3-319-24834-9_31.
7. Kłosowski G, Rymarczyk T, Gola A. Increasing the Reliability of Flood Embankments with Neural Imaging Method. Applied Sciences 2018; 8 (9): 1457, https://doi.org/10.3390/app8091457.
8. Kłosowski G, Rymarczyk T. Using neural networks and deep learning algorithms in electrical impedance tomography. Informatyka Automatyka Pomiary w Gospodarce i Ochronie Środowiska 2017; 7 (3): 99–102, https://doi.org/10.5604/01.3001.0010.5226.

9. Korzeniewska E, Gałązka-Czarnecka I, Czarnecki A, Piekarska A, Krawczyk A. Influence of PEF on Antocyjans in Wine. Przegląd Elektrotechniczny 2018; 1 (1): 59–62, https://doi.org/10.15199/48.2018.01.15.
10. Korzeniewska E, Walczak M, Rymaszewski J. Elements of Elastic Electronics Created on Textile Substrate, Proceedings of the 24th International Conference Mixed Design of Integrated Circuits and Systems - MIXDES 2017; 2017, 447-45, https://doi.org/10.23919/MIXDES.2017.8005250.
11. Kosicka E, Kozłowski E, Mazurkiewicz D. Intelligent Systems of Forecasting the Failure of Machinery Park and Supporting Fulfilment of Orders of Spare Parts. In: Burduk A., Mazurkiewicz D. (eds) Intelligent Systems in Production Engineering and Maintenance – ISPEM 2017. ISPEM 2017. Advances in Intelligent Systems and Computing, vol 637. Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-64465-3_6.
12. Kozłowski E., Mazurkiewicz D., Kowalska B., Kowalski, D. Binary Linear Programming as a Decision-Making Aid for Water Intake Operators. In: Burduk A., Mazurkiewicz D. (eds) Intelligent Systems in Production Engineering and Maintenance – ISPEM 2017. ISPEM 2017. Advances in Intelligent Systems and Computing, vol 637. Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-64465-3_20.
13. Kryszyn J, Smolik W T, Radzik B, Olszewski T, Szabatin R. Switchless Charge-Discharge Circuit for Electrical Capacitance Tomography. Measurement Science and Technology 2014; 25 (11): 115009, https://doi.org/10.1088/0957-0233/25/11/115009.
14. Kryszyn J, Waldemar S. Toolbox for 3d Modelling and Image Reconstruction in Electrical Capacitance Tomography. Informatics Control Measurement in Economy and Environment Protection 2017; 7 (1).
15. Lopato P, Tomasz C, Sikora R, Gratkowski S, Ziolkowski M. Full Wave Numerical Modelling of Terahertz Systems for Nondestructive Evaluation of Dielectric Structures. COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering 2013; 32 (3): 736–49, https://doi.org/10.1108/03321641311305719.
16. Majchrowicz M, Kapusta P, Jackowska-Strumiłło L, Sankowski D. Acceleration of image reconstruction process in the electrical capacitance tomography 3d in heterogeneous, multi-GPU system. Informatics Control Measurement in Economy and Environment Protection 2017; 7 (1): 37–41, https://doi.org/10.5604/01.3001.0010.4579.
17. Mikulka J. Accelerated Reconstruction of T2 Maps in Magnetic Resonance Imaging. Measurement Science Review 2015; 4: 210–18, https://doi.org/10.1515/msr-2015-0029.
18. Park S, Na J, Kim M, Lee J M. Multi-Objective Bayesian Optimization of Chemical Reactor Design Using Computational Fluid Dynamics. Computers & Chemical Engineering 2018; 119 : 25–37, https://doi.org/10.1016/j.compchemeng.2018.08.005.
19. Psuj G. Multi-Sensor Data Integration Using Deep Learning for Characterization of Defects in Steel Elements. Sensors 18 (2): 292, https://doi.org/10.3390/s18010292.
20. Romanowski A. Big Data-Driven Contextual Processing Methods for Electrical Capacitance Tomography. IEEE Transactions on Industrial Informatics 2018; 1–1, https://doi.org/10.1109/TII.2018.2855200.
21. Rymarczyk T, Adamkiewicz P, Polakowski K, Sikora J. Effective Ultrasound and Radio Tomography Imaging Algorithm for Two-Dimensional Problems. Przegląd Elektrotechniczny 2018; 94 (6): 62–69.
22. Rymarczyk T, Kłosowski G, Kozłowski E. A Non-Destructive System Based on Electrical Tomography and Machine Learning to Analyze the Moisture of Buildings. Sensors 2018; 18 (7): 2285.
23. Rymarczyk T, Kłosowski G. Application of Neural Reconstruction of Tomographic Images in the Problem of Reliability of Flood Protection Facilities. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (3): 425–34, https://doi.org/10.17531/ein.2018.3.11.
24. Rymarczyk T, Sikora J. Applying Industrial Tomography to Control and Optimization Flow Systems. Open Physics 2018; 16 (1): 332–45, https://doi.org/10.1515/phys-2018-0046.
25. Sobaszek Ł, Gola A, Świć A. Predictive Scheduling as a Part of Intelligent Job Scheduling System: in, 358–67. Springer, Cham 2018, https://doi.org/10.1007/978-3-319-64465-3_35.
26. Soleimani M, Mitchell C N, Banasiak R, Wajman R, Adler A. Four-dimensional electrical capacitance tomography imaging using experimental data. Progress In Electromagnetics Research 2009; 90: 171–86, https://doi.org/10.2528/PIER09010202.
27. Tian G, Yang B, Dong M, Zhu R, Yin F, Zhao X, Wang Y, Xiao W, Wang Q, Zhang W. The Effect of Temperature on the Microbial Communities of Peak Biogas Production in Batch Biogas Reactors. Renewable Energy 2018; 123: 15–25,https://doi.org/10.1016/j.renene.2018.01.119.
28. Voutilainen A, Lehikoinen A, Vauhkonen M, Kaipio J P. Three-Dimensional Nonstationary Electrical Impedance Tomography with a Single Electrode Layer. Measurement Science and Technology 2010; 21 (3): 035107, https://doi.org/10.1088/0957-0233/21/3/035107.
29. Wang Mi. Industrial Tomography: Systems and Applications. Edited by Elsevier Ltd. Woodhead Publishing 2015.
30. Ziolkowski M, Gratkowski S, Zywica A R. Analytical and Numerical Models of the Magnetoacoustic Tomography with Magnetic Induction. COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering 2018; 37 (2): 538–48, https://doi.org/10.1108/COMPEL-12-2016-0530.

Tomasz RYMARCZYK
University of Economics and Innovation
ul. Projektowa 4, 20-209 Lublin, Poland
Research and Development Center, Netrix S.A.
ul. Związkowa 26, 20-148 Lublin, Poland

Grzegorz KŁOSOWSKI
Lublin University of Technology
Department of Organization of Enterprise
ul. Nadbystrzycka 38D, 20-618 Lublin, Poland

E-mails: tomasz@rymarczyk.com, g.klosowski@pollub.pl

Article citation info:
KHANIYEV T, BASKIR MB, GOKPINAR F, MIRZAYEV F. Statistical distributions and reliability functions with type-2 fuzzy parameters. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 268–274, http://dx.doi.org/10.17531/ein.2019.2.11.

Tahir KHANIYEV
M. Bahar BASKIR
Fikri GOKPINAR
Farhad MIRZAYEV

STATISTICAL DISTRIBUTIONS AND RELIABILITY FUNCTIONS WITH TYPE-2 FUZZY PARAMETERS

ROZKŁADY STATYSTYCZNE I FUNKCJE NIEZAWODNOŚCI O PARAMETRACH ROZMYTYCH TYPU-2

Type-2 fuzzy sets were initially given by Zadeh as an extension of type-1 fuzzy sets. There is a growing interest in type-2 fuzzy set and its memberships (named secondary memberships) to handle the uncertainty in type-1 fuzzy set and its primary membership values. However, arithmetical operators on type-2 fuzzy sets have computational complexity due to third dimension of these sets. In this study, we present some mathematical operators which can be easily applied to type-2 fuzzy sets and numbers. Also, mathematical functions of type-2 fuzzy numbers are given according to their monotonicity. These functions are adapted to reliability and distribution functions of the random variables with the type-2 fuzzy parameters. These functions are applied to Exponential, Chi-square, Weibull distributions with respect to monotonicity of the parameters of these distributions.

Keywords: *type-2 fuzzy number, type-2 fuzzy parameters, (α,β)-cuts, fuzzy probability distribution, fuzzy reliability function.*

Zbiory rozmyte typu 2 po raz pierwszy zaproponował Zadeh jako rozszerzenie zbiorów rozmytych typu 1. Zbiory rozmyte typu 2 oraz ich funkcje przynależności (zwane wtórnymi funkcjami przynależności) cieszą się rosnącym zainteresowaniem, ponieważ pozwalają na modelowanie niepewności w zbiorze rozmytym typu 1 oraz wartości pierwotnych funkcji przynależności do takiego zbioru. Ich wadą jest złożoność obliczeniowa operatorów arytmetycznych wynikająca z trójwymiarowości tych zbiorów. W artykule przedstawiono operatory matematyczne, które można z powodzeniem stosować w odniesieniu do zbiorów i liczb rozmytych typu 2. Podano również funkcje matematyczne liczb rozmytych typu 2 zgodnie z ich monotonicznością. Funkcje te są dostosowane do funkcji niezawodności i rozkładu zmiennych losowych z parametrami rozmytymi typu 2. Można je stosować do opisu rozkładów wykładniczych, chi-kwadrat, oraz Weibulla w odniesieniu do monotoniczności parametrów tych rozkładów.

Słowa kluczowe: *liczba rozmyta typu-2, parametry rozmyte typu-2, (α, β) -cięcia, rozmyty rozkład prawdopodobieństwa, rozmyta funkcja niezawodności.*

1. Introduction

Fuzzy sets are useful and effective tools to model the uncertainty problem in real-life applications. The most common fuzzy sets used in these applications are known as type-1 fuzzy sets (T1FSs). Since the membership degrees of T1FSs are crisp numbers, recently, type-2 fuzzy sets (T2FSs) are also preferred by many researchers to express uncertainty in T1FSs. Zadeh [11] introduced T2FS as an extension version of the conventional T1FS. Some important studies about T2FSs can be given as Aisbett et al. [1], Hamrawi [3], Karnik and Mendel [5], Wu and Mendel [9]. Furthermore, some applications of T2FSs can be found in Tao et al. [6], Türkşen [7], Wagenkneckt and Hartmann [8], Wu and Mendel [10]. Also, basic operations on T2FSs were studied by Blewitt et.al. [2], Karnik and Mendel [5]. However, type-2 fuzzy numbers (T2FNs) are required to make theoretical inference about modelling uncertainty. In the literature, limited number of studies can be found related to the operators on T2FNs, e.g., Kardan et al. [4]. Mostly, it is difficult to use these operators on T2FNs due to the computational complexity of T2FSs.

This study introduces practical and innovative solutions for arithmetical operations on T2FN using the (α, β)-cut definition. Thus, type-2 fuzzy parameter-based distributions and reliability functions are proposed with regards to their monotonicity. Therefore, we present a novel perspective to perform various arithmetical operators on type-2 fuzzy numbers. The basis of this perspective is structured by an (α, β)-cut definition. This definition is derived from the type-1 operations on three type-1 membership functions (lower, upper and type-1 membership functions) of T2FN. Some operations (such as sum, subtraction, multiplication, division) for T2FNs are determined using the (α, β)-cuts. Then, the membership functions of T2FNs are structured by the (α, β)-cuts. Besides, we utilize this (α, β)-cut definition to form fuzzy function of T2FN under some assumptions. Finally, we give some applications of probability distributions when some parameters of the distributions are the T2FNs.

This paper is organized as follows: Section 2 provides mathematical background of type-2 fuzzy sets and numbers, (α, β)-cut definition of T2FN, fuzzy function of T2FN with its monotonicity. The applications based on the statistical distributions and reliability functions of T2FN are given in Section 3. Conclusion is drawn in Section 4.

2. Methodology

In this section, we give a brief overview of type-2 fuzzy sets and fuzzy numbers. Also, we propose some arithmetical operators on type-2 triangular fuzzy number (T2TFN) using (α, β)-cuts which are

easily obtained by operators on type-1 fuzzy number, fuzzy function of T2TFN with its monotonicity.

2.1. Type-1 and Type-2 Fuzzy Sets

A fuzzy set, B, in a universe of discourse X is defined as $B = \{(x,\mu_B(x)); x \in X\}$, where $\mu_B : X \rightarrow [0,1]$ and $\mu_B(x)$: the membership value of $x \in X$ in the fuzzy set B. The well-known issue is the type-1 fuzzy sets (T1FSs) are often represented by crisp numbers ranging from [0, 1]. Zadeh [11] defined the concept of a type-2 fuzzy set (T2FS) as an extension version of an ordinary fuzzy set, i.e., T1FS. A T2FS, $\tilde{B}$, is defined as in (1):

$$\tilde{B} = \int_{x\in X}\int_{u\in J_x} \mu_{\tilde{B}}(x,u)/(x,u), \qquad J_x \subseteq [0,1] \tag{1}$$

where J_x: primary membership function ($u \in J_x \subseteq [0,1]$), $\mu_{\tilde{B}}(x,u)$: secondary membership function and $0 \le \mu_{\tilde{B}}(x,u) \le 1$, $\int\int$ denotes union overall admissible domain values x and u. Interval T2FS (IT2FS) is a special case of the T2FS where all $\mu_{\tilde{B}}(x,u) = 1$ in (1).

The footprint of uncertainty (*FOU*) is obtained from union of all primary memberships as shown in (2):

$$FOU(\tilde{B}) = \int_{x\in X} J_x \tag{2}$$

FOU is bounded by two type-1 membership functions (type-1 MFs): i) lower MF $\underline{\mu_{\tilde{B}}(x)}$ and ii) upper MF $\overline{\mu_{\tilde{B}}(x)}$ as shown in (3):

$$FOU(\tilde{B}) = \int_{x\in X}[\underline{\mu_{\tilde{B}}(x)}, \overline{\mu_{\tilde{B}}(x)}] \tag{3}$$

2.2. Type-2 Triangular Fuzzy Numbers

In this study, particularly, we utilized T2TFNs to define some parameters of the probability distributions. A T2TFN, $\tilde{\tilde{B}}$, as an extension version of type-1 triangular fuzzy number $\tilde{B}$ T1TFN), and its type-1 MFs are illustrated in Fig. 1.

T2TFN has three membership functions as shown in Fig. 1: i) Lower membership function (LMF) of $\tilde{\tilde{B}}$ ii) Upper membership function (UMF) of $\tilde{\tilde{B}}$ and iii) Type-1 membership function (T1MF) of $\tilde{B}$. The notations of these functions are given in Table 1.

2.3. (α, β)-Cuts of Type-2 Triangular Fuzzy Numbers

In this section, we present (α, β)-cut definition of a T2TFN. We perform arithmetical operations on T2TFNs using (α, β)-cuts.

α-cut of a T1TFN $B=(a_{10},a,a_{20})$ is defined as $B_\alpha = \{\forall x \in U | \mu_B(x) \ge \alpha\}$. Lower and upper α-cut of a T1FN B can be written as follows:

Table 1. The type-1 membership functions of T2TFN

$LMF\left(\tilde{\tilde{B}}\right), \tilde{\tilde{B}}$	$UMF\left(\tilde{\tilde{B}}\right), \tilde{\tilde{B}}$	$T1MF\left(\breve{B}\right), \breve{B}$
$\underline{\mu}(x) = \begin{cases} 0, x \le \underline{a}_{12} \text{ and } x \ge \underline{a}_{21} \\ \dfrac{x-\underline{a}_{12}}{a-\underline{a}_{12}}, \underline{a}_{12} \le x < a \\ \dfrac{\underline{a}_{21}-x}{\underline{a}_{21}-a}, a \le x \le \underline{a}_{21} \end{cases}$	$\overline{\mu}(x) = \begin{cases} 0, x \le \overline{a}_{11} \text{ and } x \ge \overline{a}_{22} \\ \dfrac{x-\overline{a}_{11}}{a-\overline{a}_{11}}, \overline{a}_{11} \le x < a \\ \dfrac{\overline{a}_{22}-x}{\overline{a}_{22}-a}, a \le x \le \overline{a}_{22} \end{cases}$	$\mu(x) = \begin{cases} 0, x \le a_{10} \text{ and } x \ge a_{20} \\ \dfrac{x-a_{10}}{a-a_{10}}, a_{10} \le x < a \\ \dfrac{a_{20}-x}{a_{20}-a}, a \le x \le a_{20} \end{cases}$

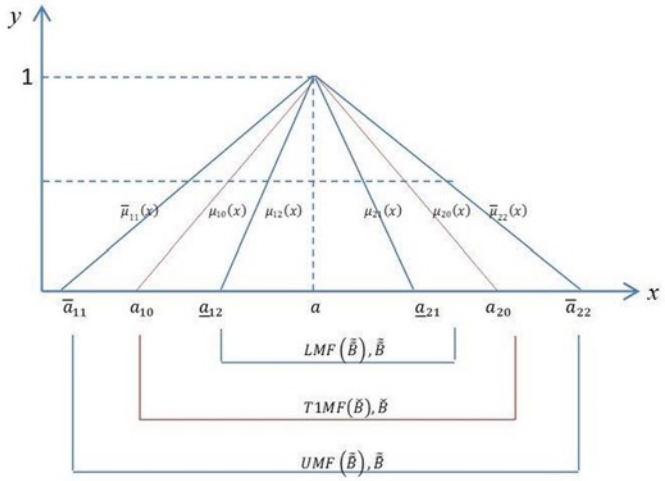


Fig. 1. T2TFN and its type-1 membership functions

$$a_{10}(\alpha) = a_{10} + (a - a_{10})\alpha \quad a_{20}(\alpha) = a_{20} - (a_{20} - a)\alpha$$

Therefore, B_α can be represented as follows:

$$B_\alpha = \left[a_{10}(\alpha), a_{20}(\alpha)\right], \alpha \in [0,1]$$

Definition 2.1. (*(α, β)-cut of T2TFN*) Based on α-cuts of T1FNs, (α, β)-cut of a T2TFN $\tilde{\tilde{B}}$ on R can be given as in (4), where $\alpha, \beta \in [0,1]$:

$$\left[\tilde{\tilde{B}}\right]_\alpha^\beta = \left[\left[\underline{\underline{a}}_{11}(\alpha,\beta), \overline{\overline{a}}_{12}(\alpha,\beta)\right]; \left[\underline{\underline{a}}_{21}(\alpha,\beta), \overline{\overline{a}}_{22}(\alpha,\beta)\right]\right] \tag{4}$$

Here,

$$\underline{\underline{a}}_{11}(\alpha,\beta) = \overline{a}_{11}(\alpha) + \beta\left(a_{10}(\alpha) - \overline{a}_{11}(\alpha)\right) \tag{5}$$

$$\overline{\overline{a}}_{12}(\alpha,\beta) = \underline{a}_{12}(\alpha) + \beta\left(a_{10}(\alpha) - \underline{a}_{12}(\alpha)\right) \tag{6}$$

$$\underline{\underline{a}}_{21}(\alpha,\beta) = \underline{a}_{21}(\alpha) + \beta\left(a_{20}(\alpha) - \underline{a}_{21}(\alpha)\right) \tag{7}$$

$$\overline{\overline{a}}_{22}(\alpha,\beta) = \overline{a}_{22}(\alpha) + \beta\left(a_{20}(\alpha) - \overline{a}_{22}(\alpha)\right) \tag{8}$$

Moreover, $\underline{a}_{ij}(\alpha), \overline{a}_{ii}(\alpha), a_{i0}(\alpha)$ can be represented as follows:

$$\underline{a}_{ij}(\alpha)=\underline{a}_{ij}+\alpha\left(a-\underline{a}_{ij}\right), i=1,\,2;\ j=1,\,2,\ i\neq j; \tag{9}$$

$$\overline{a}_{ii}(\alpha)=\overline{a}_{ii}+\alpha\left(a-\overline{a}_{ii}\right),\ i=1,\,2; \tag{10}$$

$$a_{i0}(\alpha)=a_{i0}+\alpha\left(a-a_0\right),\ \textit{for}\ i=1,\,2 \tag{11}$$

Note that, the initial constants $\underline{a}_{ij}$, $\overline{a}_{ii}$, a_{i0} are given in Table 1. The β-cuts for vertical slice of T2TFN are illustrated in Fig. 2.

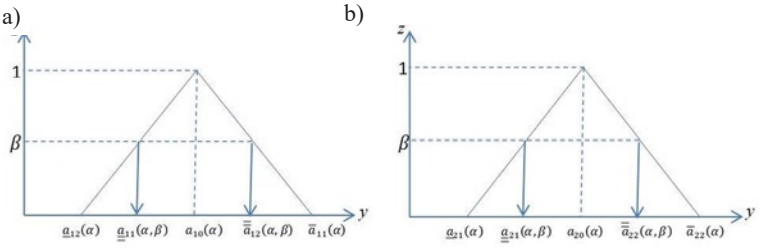


Fig. 2. The β-cuts of (a) LMF, (b) UMF of T2TFN

Finally, (α,β) cuts of T2TFN $\left(\left[\tilde{\tilde{B}}\right]_\alpha^\beta \text{ shown as in Eq.}(4)\right)$ are found as in (12)-(13) for LMF and in (14)-(15) for UMF:

$$\underline{\underline{a}}_{11}(\alpha,\beta)=\overline{a}_{11}+\alpha\left(a-\overline{a}_{11}\right)-\beta(1-\alpha)\left(\overline{a}_{11}-a_{10}\right), \tag{12}$$

$$\overline{\overline{a}}_{12}(\alpha,\beta)=\underline{a}_{12}+\alpha\left(a-\underline{a}_{12}\right)+\beta(1-\alpha)\left(\underline{a}_{12}-a_{10}\right), \tag{13}$$

$$\underline{\underline{a}}_{21}(\alpha,\beta)=\underline{a}_{21}+\alpha\left(a-\underline{a}_{21}\right)+\beta(1-\alpha)\left(\underline{a}_{21}-a_{20}\right), \tag{14}$$

$$\overline{\overline{a}}_{22}(\alpha,\beta)=\overline{a}_{22}+\alpha\left(a-\overline{a}_{22}\right)-\beta(1-\alpha)\left(\overline{a}_{22}-a_{20}\right). \tag{15}$$

2.4. Fuzzy Function of Type-2 Triangular Fuzzy Number

In this section, we present how to determine fuzzy function of a T2TFN under some assumptions.

Definition 2.2. Let *f(x)* be a function of a variable *X* and *X* be a type-2 fuzzy number. The following rules can be obtained for *f(.)* at the point of *X=a* under the monotonicity of the function *f*.

Rule-1: Let *f(x)* be monotone increasing function.

Consider $\left[\tilde{\tilde{a}}\right]_\alpha^\beta=\left[\left[\underline{\underline{a}}_{11}(\alpha,\beta),\overline{\overline{a}}_{12}(\alpha,\beta)\right];\left[\underline{\underline{a}}_{21}(\alpha,\beta),\overline{\overline{a}}_{22}(\alpha,\beta)\right]\right]$. The fuzzy function of $\left[\tilde{\tilde{a}}\right]_\alpha^\beta$ is described as follows:

$$\left[f\left(\tilde{\tilde{a}}\right)\right]_\alpha^\beta=\left[\left[f\left(\underline{\underline{a}}_{11}(\alpha,\beta)\right),f\left(\overline{\overline{a}}_{12}(\alpha,\beta)\right)\right];\left[f\left(\underline{\underline{a}}_{21}(\alpha,\beta)\right),f\left(\overline{\overline{a}}_{22}(\alpha,\beta)\right)\right]\right].$$

$$f\left(\underline{\underline{a}}_{11}(\alpha,\beta)\right)\Big|_{\beta=0}=x \text{ then } \underline{\underline{a}}_{11}(\alpha,0)=f^{-1}(x),\Rightarrow\alpha=\mu_{11}\left(f^{-1}(x)\right)\equiv\mu_{f11}(x)$$

Similarly, the following membership functions are found:

$$\mu_{f12}(x)=\mu_{12}\left(f^{-1}(x)\right),$$

$$\mu_{f21}(x)=\mu_{21}\left(f^{-1}(x)\right),$$

$$\mu_{f22}(x)=\mu_{22}\left(f^{-1}(x)\right),$$

Rule-2: Let *f(x)* be monotone decreasing function. Consider $\left[\tilde{\tilde{a}}\right]_\alpha^\beta=\left[\left[\underline{\underline{a}}_{11}(\alpha,\beta),\overline{\overline{a}}_{12}(\alpha,\beta)\right];\left[\underline{\underline{a}}_{21}(\alpha,\beta),\overline{\overline{a}}_{22}(\alpha,\beta)\right]\right]$. The fuzzy function of $\left[\tilde{\tilde{a}}\right]_\alpha^\beta$ is described as follows:

$$\left[f\left(\tilde{\tilde{a}}\right)\right]_\alpha^\beta=\left[\left[f\left(\overline{\overline{a}}_{22}(\alpha,\beta)\right),f\left(\underline{\underline{a}}_{21}(\alpha,\beta)\right)\right];\left[f\left(\overline{\overline{a}}_{12}(\alpha,\beta)\right),f(\underline{\underline{a}}_{11}(\alpha,\beta))\right]\right]$$

If $f\left(\overline{\overline{a}}_{22}(\alpha,\beta)\right)\Big|_{\beta=0}=x$ then $\overline{\overline{a}}_{22}(\alpha,0)=f^{-1}(x)$,

$\Rightarrow\alpha=\mu_{22}\left(f^{-1}(x)\right)=\mu_{f11}(x)$.

Similarly, the following membership functions are found:

$$\mu_{f12}(x)=\mu_{21}\left(f^{-1}(x)\right),$$

$$\mu_{f21}(x)=\mu_{12}\left(f^{-1}(x)\right),$$

$$\mu_{f22}(x)=\mu_{11}\left(f^{-1}(x)\right),$$

Rule-3: Let *f(x)*be non-monotone function. However, consider that *f(x)* is monotone function for each $[d_i,d_{i+1}]$, where $-\infty<d_1<d_2<\ldots<d_i<\ldots<\infty$. Accordingly, Rule-1 is applied to *f(x)* if *f(x)* is monotone increasing function at $[d_i,d_{i+1}]$, Rule-2 is applied to *f(x)* if *f(x)* is monotone decreasing function at $[d_i,d_{i+1}]$. An illustration of non-monotone function is given in Fig. 3.

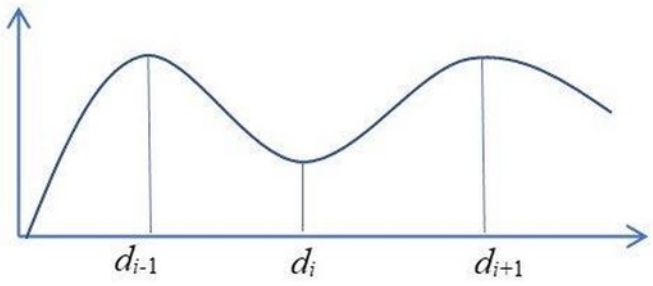


Fig. 3. An illustration of non-monotone function

Definition 2.3. Let $F(x;a)$ be distribution function with parameter-*a* defined by T2TFN. (α,β)-cut of the parameter-a ($\left[\tilde{\tilde{a}}\right]_\alpha^\beta$) is defined as in (4). Then, reliability function with T2TFN-defined parameter ($R(x;a)$) can be determined as in (16):

$$R(x;a)=1-F(x;a) \tag{16}$$

The following rules can be obtained for whether R is monotone by the parameter-a.

Rule-1: Let $R(x;a)$ be monotone increasing function by the parameter-a. If the parameter is defined as in (4) then:

$$\left[R\left(x;\tilde{\tilde{a}}\right)\right]_{\alpha}^{\beta}=\left[\left[R_{11}\left(x;a;(\alpha,\beta)\right),R_{12}\left(x;a;(\alpha,\beta)\right)\right];\left[R_{21}\left(x;a;(\alpha,\beta)\right),R_{22}\left(x;a;(\alpha,\beta)\right)\right]\right].$$

Here:

$$R_{11}\left(x;a;(\alpha,\beta)\right)=R\left(x;\underline{\underline{a}}_{11}(\alpha,\beta)\right)$$

$$R_{12}\left(x;a;(\alpha,\beta)\right)=R\left(x;\overline{\overline{a}}_{12}(\alpha,\beta)\right)$$

$$R_{21}\left(x;a;(\alpha,\beta)\right)=R\left(x;\underline{\underline{a}}_{21}(\alpha,\beta)\right)$$

$$R_{22}\left(x;a;(\alpha,\beta)\right)=R\left(x;\overline{\overline{a}}_{22}(\alpha,\beta)\right)$$

Rule-2: Let $R(x;a)$ be monotone decreasing function by the parameter-a. According to T2TFN-definedparameter as in (4), reliability function will be obtained as follows:

$$\left[R\left(x;\tilde{\tilde{a}}\right)\right]_{\alpha}^{\beta}=\left[\left[R_{11}\left(x;a;(\alpha,\beta)\right),R_{12}\left(x;a;(\alpha,\beta)\right)\right];\left[R_{21}\left(x;a;(\alpha,\beta)\right),R_{22}\left(x;a;(\alpha,\beta)\right)\right]\right]$$

Here

$$R_{11}\left(x;a;(\alpha,\beta)\right)=R\left(x;\overline{\overline{a}}_{22}(\alpha,\beta)\right)$$

$$R_{12}\left(x;a;(\alpha,\beta)\right)=R\left(x;\underline{\underline{a}}_{21}(\alpha,\beta)\right)$$

$$R_{21}\left(x;a;(\alpha,\beta)\right)=R\left(x;\overline{\overline{a}}_{12}(\alpha,\beta)\right)$$

$$R_{22}\left(x;a;(\alpha,\beta)\right)=R\left(x;\underline{\underline{a}}_{11}(\alpha,\beta)\right)$$

Rule-3: Let *R(x;a)* be non-monotone function. However, consider that *R(x;a)* is monotone function for each $[d_i,d_{i+1}]$, where $-\infty<d_1<d_2<\ldots<d_i<\ldots<\infty$. Accordingly, Rule-1 is applied to *R(x;a)* if *R(x;a)* is monotone increasing function at $[d_i,d_{i+1}]$, Rule-2 is applied to *R(x;a)* if *R(x;a)* is monotone decreasing function at $[d_i,d_{i+1}]$.

Distribution function with T2TFN-defined parameter (*F(x;a)*) can be determined by definition 2.3. In this study, Matlab R2018b programming language was used for the illustrations and calculations in section 3.

3. Experimental Results

In this section, we demonstrate how to apply to the well-known three probability distributions when some parameters of these distributions are chosen as T2TFN.

3.1. Experiment-1

Let the distribution function of exponential variate X be as $F(x;a)$ and the parameter a be type-2 fuzzy number. The distribution function $F(x;a)$ can be written as in (17):

$$F(x;a)=P(X\le x)=1-e^{-xa},x>0 \tag{17}$$

Thus, this distribution function is an increasing function according to a. Therefore, the (α,β)-cut of distribution function F can be written as in (18)-(21):

$$F_{11}\left(x;a;(\alpha,\beta)\right)=F\left(x;\underline{\underline{a}}_{11}(\alpha,\beta)\right)=1-e^{-x\underline{\underline{a}}_{11}(\alpha,\beta)}, \tag{18}$$

$$F_{12}\left(x;a;(\alpha,\beta)\right)=F\left(x;\overline{\overline{a}}_{12}(\alpha,\beta)\right)=1-e^{-x\overline{\overline{a}}_{12}(\alpha,\beta)}, \tag{19}$$

$$F_{21}\left(x;a;(\alpha,\beta)\right)=F\left(x;\underline{\underline{a}}_{21}(\alpha,\beta)\right)=1-e^{-x\underline{\underline{a}}_{21}(\alpha,\beta)}, \tag{20}$$

$$F_{22}\left(x;a;(\alpha,\beta)\right)=F\left(x;\overline{\overline{a}}_{22}(\alpha,\beta)\right)=1-e^{-x\overline{\overline{a}}_{22}(\alpha,\beta)}, \tag{21}$$

(α,β)-cuts for distribution function F of an Exponential variate is illustrated in Fig. 4.

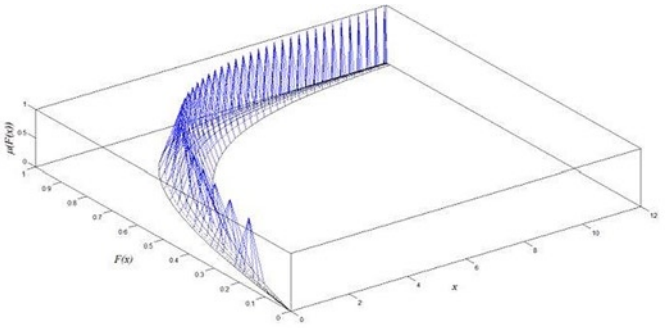


Fig. 4. (α,β)-cuts for distribution function F of an Exponential variate

The reliability function R for exponential distribution can be written as follows:

$$R(x;a)=P(X>x)=e^{-xa},\ x>0 \tag{22}$$

It is easy to see that, $R(x;a)$ is a decreasing function according to parameter a. Therefore, in this case, the (α,β)-cut of reliability function $R(x;a)$ can be written as in (23)-(26):

$$R_{11}\left(x;a;(\alpha,\beta)\right)=R\left(x;\overline{\overline{a}}_{22}(\alpha,\beta)\right)=e^{-x\overline{\overline{a}}_{22}(\alpha,\beta)}, \tag{23}$$

$$R_{12}\left(x;a;(\alpha,\beta)\right)=R\left(x;\underline{\underline{a}}_{21}(\alpha,\beta)\right)=e^{-x\underline{\underline{a}}_{21}(\alpha,\beta)}, \tag{24}$$

$$R_{21}\left(x;a;(\alpha,\beta)\right)=R\left(x;\overline{\overline{a}}_{12}(\alpha,\beta)\right)=e^{-x\overline{\overline{a}}_{12}(\alpha,\beta)}, \tag{25}$$

$$R_{22}\left(x;a;(\alpha,\beta)\right)=R\left(x;\underline{\underline{a}}_{11}(\alpha,\beta)\right)=e^{-x\underline{\underline{a}}_{11}(\alpha,\beta)}. \qquad (26)$$

(α,β) cuts for reliability function *R* of an Exponential variate is illustrated in Fig.5.

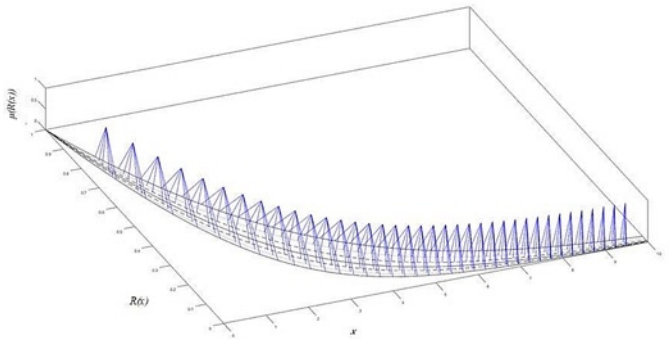


Fig. 5. (α,β)*-cuts for Reliability function R of an Exponential variate.*

3.2. Experiment-2

Let the distribution function of chi-square variate *X* be as $F(x;a)$ and the parameter *a* be type-2 fuzzy number (T2FN). The distribution function $F(x;a)$ can be written as in (27):

$$F(x;a)=P(X\le x)=\frac{1}{2^{\frac{a}{2}}\Gamma\left(\frac{a}{2}\right)}\int_0^x t^{\frac{a}{2}-1}e^{-t/2}dt \qquad (27)$$

Thus, this distribution function is an decreasing function according to *a*. The (α,β)-cut of the parameter *a* is as in (4).

By using the (α,β)-cut of parameter a, the (α,β)-cut of distribution function *F* can be written as in (28)-(31):

$$F_{11}\left(x;a;(\alpha,\beta)\right)=F\left(x;\bar{\bar{a}}_{22}(\alpha,\beta)\right)=\frac{1}{2^{\frac{\bar{\bar{a}}_{22}(\alpha,\beta)}{2}}\Gamma\left(\frac{\bar{\bar{a}}_{22}(\alpha,\beta)}{2}\right)}\int_0^x t^{\frac{\bar{\bar{a}}_{22}(\alpha,\beta)}{2}-1}e^{-t/2}dt, \qquad (28)$$

$$F_{12}\left(x;a;(\alpha,\beta)\right)=F\left(x;\underline{\underline{a}}_{21}(\alpha,\beta)\right)=\frac{1}{2^{\frac{\underline{\underline{a}}_{21}(\alpha,\beta)}{2}}\Gamma\left(\frac{\underline{\underline{a}}_{21}(\alpha,\beta)}{2}\right)}\int_0^x t^{\frac{\underline{\underline{a}}_{21}(\alpha,\beta)}{2}-1}e^{-t/2}dt, \qquad (29)$$

$$F_{21}\left(x;a;(\alpha,\beta)\right)=F\left(x;\bar{\bar{a}}_{12}(\alpha,\beta)\right)=\frac{1}{2^{\frac{\bar{\bar{a}}_{12}(\alpha,\beta)}{2}}\Gamma\left(\frac{\bar{\bar{a}}_{12}(\alpha,\beta)}{2}\right)}\int_0^x t^{\frac{\bar{\bar{a}}_{12}(\alpha,\beta)}{2}-1}e^{-t/2}dt, \qquad (30)$$

$$F_{22}\left(x;a;(\alpha,\beta)\right)=F\left(x;\underline{\underline{a}}_{11}(\alpha,\beta)\right)=\frac{1}{2^{\frac{\underline{\underline{a}}_{11}(\alpha,\beta)}{2}}\Gamma\left(\frac{\underline{\underline{a}}_{11}(\alpha,\beta)}{2}\right)}\int_0^x t^{\frac{\underline{\underline{a}}_{11}(\alpha,\beta)}{2}-1}e^{-t/2}dt, \qquad (31)$$

(α,β)-cuts for distribution function *F* of a Chi-square variate is illustrated in Fig. 6.

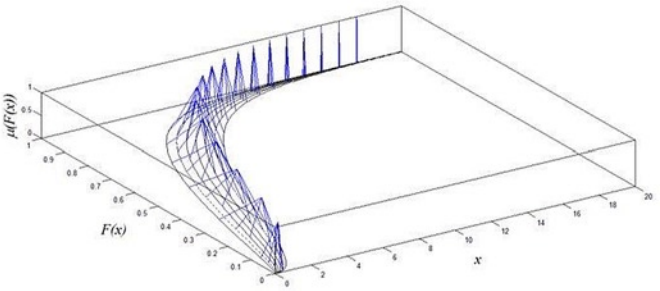


Fig. 6. (α,β)*-cuts for distribution function F of a Chi-square variate*

The reliability function *R* for chi-square distribution is an increasing function according to parameter *a*.

$$R(x;a)=P(X>x)=1-\frac{1}{2^{\frac{a}{2}}\Gamma\left(\frac{a}{2}\right)}\int_0^x t^{\frac{a}{2}-1}e^{-t/2}dt \qquad (32)$$

Therefore, (α,β)-cut of *R* can be written as in (33)-(36):

$$\mathrm{R}\left(\mathrm{x};\mathrm{a};(\alpha,\beta)\right)=\mathrm{R}\left(\mathrm{x};\underline{\underline{\mathrm{a}}}_{11}(\alpha,\beta)\right)=1-\frac{1}{2^{\frac{\underline{\underline{\mathrm{a}}}_{11}(\alpha,\beta)}{2}}\Gamma\left(\frac{\underline{\underline{\mathrm{a}}}_{11}(\alpha,\beta)}{2}\right)}\int_0^x \mathrm{t}^{\frac{\underline{\underline{\mathrm{a}}}_{11}(\alpha,\beta)}{2}-1}\mathrm{e}^{-\mathrm{t}/2}\mathrm{dt}, \qquad (33)$$

$$R_{12}\left(x;a;(\alpha,\beta)\right)=R\left(x;\bar{\bar{a}}_{12}(\alpha,\beta)\right)=1-\frac{1}{2^{\frac{\bar{\bar{a}}_{12}(\alpha,\beta)}{2}}\Gamma\left(\frac{\bar{\bar{a}}_{12}(\alpha,\beta)}{2}\right)}\int_0^x t^{\frac{\bar{\bar{a}}_{12}(\alpha,\beta)}{2}-1}e^{-t/2}dt, \qquad (34)$$

$$R_{21}\left(x;a;(\alpha,\beta)\right)=R\left(x;\underline{\underline{a}}_{21}(\alpha,\beta)\right)=1-\frac{1}{2^{\frac{\underline{\underline{a}}_{21}(\alpha,\beta)}{2}}\Gamma\left(\frac{\underline{\underline{a}}_{21}(\alpha,\beta)}{2}\right)}\int_0^x t^{\frac{\underline{\underline{a}}_{21}(\alpha,\beta)}{2}-1}e^{-t/2}dt, \qquad (35)$$

$$R_{22}\left(x;a;(\alpha,\beta)\right)=R\left(x;\bar{\bar{a}}_{22}(\alpha,\beta)\right)=1-\frac{1}{2^{\frac{\bar{\bar{a}}_{22}(\alpha,\beta)}{2}}\Gamma\left(\frac{\bar{\bar{a}}_{22}(\alpha,\beta)}{2}\right)}\int_0^x t^{\frac{\bar{\bar{a}}_{22}(\alpha,\beta)}{2}-1}e^{-t/2}dt. \qquad (36)$$

(α,β)-cuts for reliability function R of a Chi-square variate is illustrated in Fig. 7.

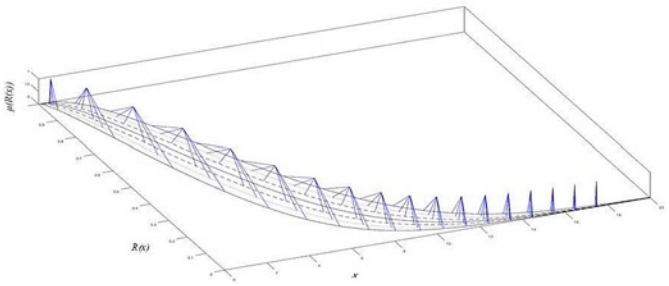


Fig. 7. (α,β)*-cuts for Reliability function R of a Chi-Square variate*

3.3. Experiment-3

Let the distribution function of Weibull variate X be as $F\left(x;(a,\sigma)\right)$ and the parameter a be type-2 fuzzy number. The distribution function $F\left(x;(a,\sigma)\right)$ can be written as follows:

$$F\left(x;(a,\sigma)\right)=P\left(X\le x\right)=1-e^{-\left(\frac{x}{a}\right)^{\sigma}} \tag{37}$$

It is seen that, this distribution function is a decreasing function according to parameter a. The (α,β)-cut of the parameter a is as in (4).

In this case, the (α,β)-cut of distribution function F can be written as in (38)-(41):

$$F_{11}\left(x;a;(\alpha,\beta)\right)=F\left(x;\overline{\overline{a}}_{22}(\alpha,\beta)\right)=1-e^{-\left(\frac{x}{\overline{\overline{a}}_{22}(\alpha,\beta)}\right)^{\sigma}}, \tag{38}$$

$$F_{12}\left(x;a;(\alpha,\beta)\right)=F\left(x;\underline{\underline{a}}_{21}(\alpha,\beta)\right)=1-e^{-\left(\frac{x}{\underline{\underline{a}}_{21}(\alpha,\beta)}\right)^{\sigma}}, \tag{39}$$

$$F_{21}\left(x;a;(\alpha,\beta)\right)=F\left(x;\overline{\overline{a}}_{12}(\alpha,\beta)\right)=1-e^{-\left(\frac{x}{\overline{\overline{a}}_{12}(\alpha,\beta)}\right)^{\sigma}}, \tag{40}$$

$$F_{22}\left(x;a;(\alpha,\beta)\right)=F\left(x;\underline{\underline{a}}_{11}(\alpha,\beta)\right)=1-e^{-\left(\frac{x}{\underline{\underline{a}}_{11}(\alpha,\beta)}\right)^{\sigma}}, \tag{41}$$

(α,β)-cuts for distribution function F of a Weibull variate is illustrated in Fig. 8.

The reliability function $R\left(x;(a,\sigma)\right)$ of Weibull distribution has the following form:

$$R\left(x;(a,\sigma)\right)=P\left(X>x\right)=e^{-\left(\frac{x}{a}\right)^{\sigma}} \tag{42}$$

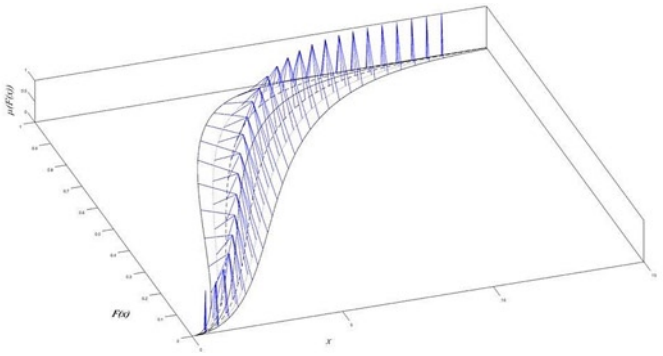


Fig. 8. (α,β)*-cuts for distribution function F of a Weibull variate*

It is seen from (42) that the reliability function $R\left(x;(a,\sigma)\right)$ is an increasing function according to parameter a. Therefore, (α,β)-cut of reliability function R can be written as in (43)-(46):

$$R_{11}\left(x;a;(\alpha,\beta)\right)=R\left(x;\underline{\underline{a}}_{11}(\alpha,\beta)\right)=e^{-\left(\frac{x}{\underline{\underline{a}}_{11}(\alpha,\beta)}\right)^{\sigma}} \tag{43}$$

$$R_{12}\left(x;a;(\alpha,\beta)\right)=R\left(x;\overline{\overline{a}}_{12}(\alpha,\beta)\right)=e^{-\left(\frac{x}{\overline{\overline{a}}_{12}(\alpha,\beta)}\right)^{\sigma}} \tag{44}$$

$$R_{21}\left(x;a;(\alpha,\beta)\right)=R\left(x;\underline{\underline{a}}_{21}(\alpha,\beta)\right)=e^{-\left(\frac{x}{\underline{\underline{a}}_{21}(\alpha,\beta)}\right)^{\sigma}} \tag{45}$$

$$R_{22}\left(x;a;(\alpha,\beta)\right)=R\left(x;\overline{\overline{a}}_{22}(\alpha,\beta)\right)=e^{-\left(\frac{x}{\overline{\overline{a}}_{22}(\alpha,\beta)}\right)^{\sigma}} \tag{46}$$

(α,β)-cuts for reliability function R of a Weibull variate is illustrated in Fig. 9.

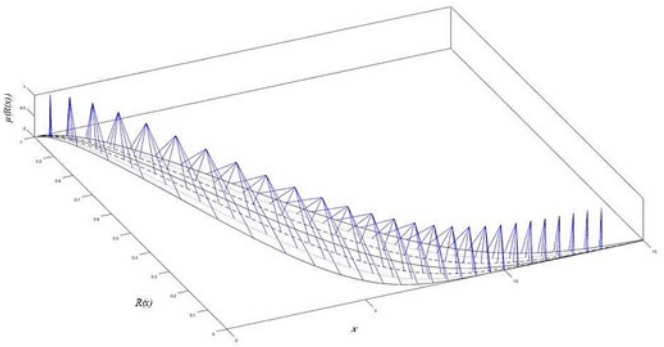


Fig. 9. (α,β)*-cuts for Reliability function R of a Weibull variate*

4. Conclusion

In this study, we presented (α, β)-cut definition of a T2TFN which is easily defined by type-1 membership functions of the T2TFN. Therefore, the computational complexity due to third dimension of T2FSs was reduced by using this (α, β)-cut definition. According to this definition, we proposed fuzzy function of T2TFN under its structural rules: i) monotone decreasing, ii) monotone increasing, iii) non-monotone functions. We provided how to apply the (α, β)-cut to some well-known statistical distributions and reliability functions,

where the parameters are defined as T2TFN. The proposed T2TFN parameter-based distributions and reliability functions can be used for various application areas related to survival analysis, reliability-based system design, machine productivity. In future studies, the reliability function with type-2 fuzzy parameter, expressing by non-triangular membership function, will be investigated.

Acknowledgement

The authors would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper. They would also like to thank the Editors for their generous comments and support during the review process.

References

1. Aisbett J, Rickard J T, Morgenthaler D. Multivariate modeling and type-2 fuzzy sets. Fuzzy Sets and Systems 2011; 163(1): 78–95, https://doi.org/10.1016/j.fss.2010.10.001.
2. Blewitt W, Coupland S, Zhou S M. A novel approach to type-2 fuzzy addition. In 'Proc. FUZZ-IEEE 2007', London, 2007; 1456 – 1461, https://doi.org/10.1109/FUZZY.2007.4295581.
3. Hamrawi H. Type-2 fuzzy alpha-cuts, PhD. Thesis. De Montfort University 2011.
4. Kardan I, Akbarzadeh-T M R, Akbarzadeh K, Kalani H. Quasi type 2 fuzzy differential equations. Journal of Intelligent & Fuzzy Systems 2017; 32(1): 551–563, https://doi.org/10.3233/JIFS-152470.
5. Karnik N N, Mendel J M. Operations on type-2 fuzzy sets. Fuzzy Sets and Systems 2001; 122(2): 327–348, https://doi.org/10.1016/S0165-0114(00)00079-8.
6. Tao C W, Taur J S, Chang C W, Chang Y H. Simplified type-2 fuzzy sliding controller for wing rock system. Fuzzy Sets and Systems 2012; 207: 111–129, https://doi.org/10.1016/j.fss.2012.02.015.
7. Türkşen I B. Type-2 representation and reasoning for CWW. Fuzzy Sets and Systems 2002; 127(1): 17–36, https://doi.org/10.1016/S0165-0114(01)00150-6.
8. Wagenknecht M, Hartmann K. Application of fuzzy sets of type-2 to the solution of fuzzy equations systems. Fuzzy Sets and Systems 1988; 25(2): 183–190, https://doi.org/10.1016/0165-0114(88)90186-8.
9. Wu D, Mendel J. Uncertainty measures for interval type-2 fuzzy sets. Information Sciences 2007a; 177(23): 5378–5393, https://doi.org/10.1016/j.ins.2007.07.012.
10. Wu D, Mendel J. Aggregation using the linguistic weighted average and interval type-2 fuzzy sets. IEEE Transactions on Fuzzy Systems 2007b; 15(6): 1145–1161, https://doi.org/10.1109/TFUZZ.2007.896325.
11. Zadeh L A. The concept of a linguistic variable and its applications in approximate reasoning (I). Information Sciences 1975; 8(3): 199–249, https://doi.org/10.1016/0020-0255(75)90036-5.

Tahir KHANIYEV
Department of Industrial Engineering
TOBB University of Economics and Technology
Sogutozu str., 06560 Ankara, Turkey

M. Bahar BASKIR
Department of Statistics
Bartin University
Kutlubey-Yazicilar campus, 74100 Bartin, Turkey

Fikri GOKPINAR
Department of Statistics
Gazi University
Teknikokullar str., 06500 Ankara, Turkey

Farhad MIRZAYEV
Department of Economic Cybernetics
Baku State University
Khalilov str. 23, AZ 1148 Baku, Azerbaijan

E-mails: tahirkhaniyev@etu.edu.tr, mbaskir@bartin.edu.tr, fikri@gazi.edu.tr, farhad_1958@mail.ru

Article citation info:
KILIKEVIČIENĖ K, MATIJOŠIUS J, FURSENKO A, KILIKEVIČIUS A. Tests of hail simulation and research of the resulting impact on the structural reliability of solar cells. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 275–281, http://dx.doi.org/10.17531/ein.2019.2.12.

Kristina KILIKEVIČIENĖ
Jonas MATIJOŠIUS
Antanas FURSENKO
Artūras KILIKEVIČIUS

TESTS OF HAIL SIMULATION AND RESEARCH OF THE RESULTING IMPACT ON THE STRUCTURAL RELIABILITY OF SOLAR CELLS

BADANIA SYMULACYJNE WPŁYWU OPADÓW GRADU NA NIEZAWODNOŚĆ KONSTRUKCJI OGNIW SŁONECZNYCH

The mandatory tests of resistance to hail is carried out in order to qualify solar cell modules according to standards (IEC 61215 and IEC 61646). The efficiency of modern photovoltaic systems decreases significantly, when the crystalline structure of solar cells is damaged as a result of climatic factors, such as wind, hail, etc., which are similar to mechanical-dynamic effects. This work presents a conducted research of efficiency and reliability of solar cells, simulating hail effects. A testbed was created specifically for carrying out experimental research. During the research, solar elements were exposed to impact, cyclic dynamic loads, with the frequency of revolutions of the balls simulating hail ranging from 5 to 20 Hz, the amplitude of the impact excitation acceleration of the solar cell - up to 986 m / s2 and the force amplitude - up to 1129 N. Experimental research results revealed the reliability of photovoltaic modules of different sizes during the simulation of hail. The proposed assessment methodology of hail effects can be successfully applied in studies of the influence of mechanical-dynamic effects of solar cells of different structures.

Keywords: *solar cells, degradation, measurement of mechanical stress, reliability, hail simulation.*

Przy kwalifikacji modułów ogniw słonecznych do użytkowania przeprowadza się obowiązkowe badania odporności na gradobicie zgodnie z normami IEC 61215 i IEC 61646. Wydajność nowoczesnych systemów fotowoltaicznych znacznie spada, gdy struktura krystaliczna ogniw słonecznych ulega uszkodzeniu w wyniku czynników klimatycznych, takich jak wiatr, gradobicie itp., które przypominają w swoim działaniu obciążenia mechaniczno-dynamiczne. W pracy przedstawiono wyniki badań symulacyjnych wpływu gradobicia na wydajność i niezawodność ogniw słonecznych. Badania prowadzono na specjalnie do tego celu skonstruowanym stanowisku testowym. Podczas badań, elementy słoneczne były wystawiane na cykliczne obciążenia dynamiczne wywoływane uderzeniami kulek symulujących grad o częstotliwości obrotów od 5 do 20 Hz przy amplitudzie przyspieszenia wzbudzenia uderzeniowego ogniwa słonecznego wynoszącej do 986 m/s2 oraz amplitudzie siły do 1129 N. Wyniki symulacji pozwoliły ocenić niezawodność modułów fotowoltaicznych o różnych rozmiarach. Proponowaną metodologię oceny wpływu opadów gradu można z powodzeniem stosować w badaniach oddziaływania obciążeń mechaniczno-dynamicznych na ogniwa słoneczne o różnych strukturach.

Słowa kluczowe: *ogniwa słoneczne, degradacja, pomiar naprężenia mechanicznego, niezawodność, symulacja opadów gradu.*

1. Introduction

Changes in weather conditions, such as the impact of hail or effects of severe snow loads, can lead to glass cracks on PV modules, thus damaging them. Such extreme weather conditions (severe hail effects and high snow loads) can destroy modules that cause glass damage due to mechanical impacts, also unveiling hidden damage, because internal solar cell tears and cracks can form along with thermo-mechanical stresses caused by wind and temperature changes, leading to energy losses and negative loss of stiffness of the structure.

Not only industrialists, but also private users use solar energy today, because the price of solar cells has been decreasing due to improving technology and mass production [12, 18]. House roofs and walls are covered in solar cells [2]. Efficiency is a very important indicator for photovoltaic systems [15], while weather conditions decrease the efficiency of solar cells.

Such factors as wind and temperature fluctuations are very important in assessing the longevity and performance efficiency of a particular solar cell and the entire system [5, 8, 14, 16, 17]. Dynamic-mechanical loads result in micro-cracks, which damage the crystal structure of photovoltaic modules [18], which adversely affects the efficiency of light energy conversion and the quality of work of the solar module [3, 4, 9, 20]. In order to reduce the loss of efficiency, knowing the impact of weather conditions is necessary [13, 14], and environmental vibration tests perfectly serve this purpose [6]. Environmental vibration tests are dynamic tests, when the wake-up occurs naturally, for example, from the air flow caused by oscillations of the earth, transport-driven excitation and the like. Excitation forces are not measurable, measuring the response of the experimental object instead. Time and amplitude dependence algorithms have been composed on the basis of the information received [7, 11, 21]. The source of excitation may be wind, periodic or accidental seismic effects or other sources [1].

The performance of the hail test requires expensive space-consuming equipment, which is often difficult to implement. Thus one of the main objectives of this work is to present a simple methodology to simulate hail. The proposed methodology for the assessment

of hail effects can be successfully applied in studying the influence of mechanical-dynamic effects of solar cells of different structures.

2. Research object and testbed

The object of research is a photovoltaic (polycrystalline) solar cell, the geometric parameters and other data of which are presented in Table 1.

Table 1. Photovoltaic (polycrystalline) module solet

Specifications	
Cell size	37x156 mm
Number of cells	5x2
Front side glass	3,2 mm hardened solar glass
Dimensions (L x W)	322 x 204 mm

A testbed, the main element of which is a motor with a frequency convertor and balls simulating ice cubes, was created for hail simulation. The main purpose of the testbed was to reach the average speed of ice cubes during hail (with the diameter of the cube being 40 mm and its mass – 30 g).

The danger level posed by hail is classified according to the energy of the impact of hail cubes. Kinetic energy generated by a hail cube depends on the mass and speed of the hail cube before the impact (1):

$$E = 1/2\ m\ v^2 \tag{1}$$

Where: E kinetic energy [J], *m* mass of hail cube [kg], *v* impact speed of hail cube [m/s].

The speed of a hail cube (before the impact) is calculated according to the formula (2):

$$v = r\ 2\ \pi f \tag{2}$$

Where: r – radius (from the axis of revolution to the point of impact), m; *f* – frequency of revolutions per second, Hz.

Changing the radius r and the frequency of revolutions f was possible during the research. Results of the speed of a hail cube (*v*) and kinetic energy (*E*) when changing the radius and frequency of revolutions are presented in Table 2. The speed of a hail cube (*v*) and kinetic energy (*E*) recommended according to the standard [10] with the hail cube weighing 40 g is 27.5 m/s and 11.1 J, respectively.

Table 2. Results of the hail cube speed (v) and kinetic energy (E) when changing radius r and frequency of revolutions f

Seq. No.	r, m	f, Hz	v, m/s	E, J
1	0,2	5	6,28	0,790
2	0,25	5	7,85	1,234
3	0,3	5	9,42	1,777
4	0,35	5	10,99	2,418
5	0,4	5	12,57	3,158
6	0,2	10	12,57	3,158
7	0,25	10	15,71	4,935
8	0,3	10	18,86	7,106
9	0,35	10	21,99	9,672
10	0,4	10	25,13	12,63
11	0,2	12	15,07	4,543
12	0,25	12	18,84	7,099
13	0,3	12	22,61	10,222
14	0,35	12	26,38	13,914
15	0,4	12	30,14	18,173
16	0,2	14	17,58	6,184
17	0,25	14	21,98	9,662
18	0,3	14	26,38	13,914
19	0,35	14	30,77	18,938
20	0,4	14	35,17	24,736
21	0,2	16	20,10	8,0770
22	0,25	16	25,12	12,620
23	0,3	16	30,14	18,173
24	0,35	16	35,17	24,736
25	0,4	16	40,19	32,308
26	0,2	20	25,12	12,62029
27	0,25	20	31,4	19,7192
28	0,3	20	37,68	28,39565
29	0,35	20	43,96	38,64963
30	0,4	20	50,24	50,48115

Fig. 1. General views of solar cells

The general view of the research solar cells is presented in Fig. 1.

A testbed was used for experimental simulation of actual hail weather conditions. The change of the impact amplitude and frequency of oscillations allows simulating dynamic-mechanical loads that are typical of hail. During the experiments, an accelerometer and a force transducer were attached to the middle of the solar cells.

Fig. 2 presents the general view of the experimental hail simulation testbed (a) and points of attachment of the accelerometer and the force transducer (b). Figure 2a illustrates the part of the testbed simulating hail: a motor Vela STM (position 3) with a frequency convertor Lenze SMVector (position 4); researched solar cell (position 2); data collection and analysis equipment "Machine Diagnostics Toolbox – Type 9727" (position 1); hail simu-

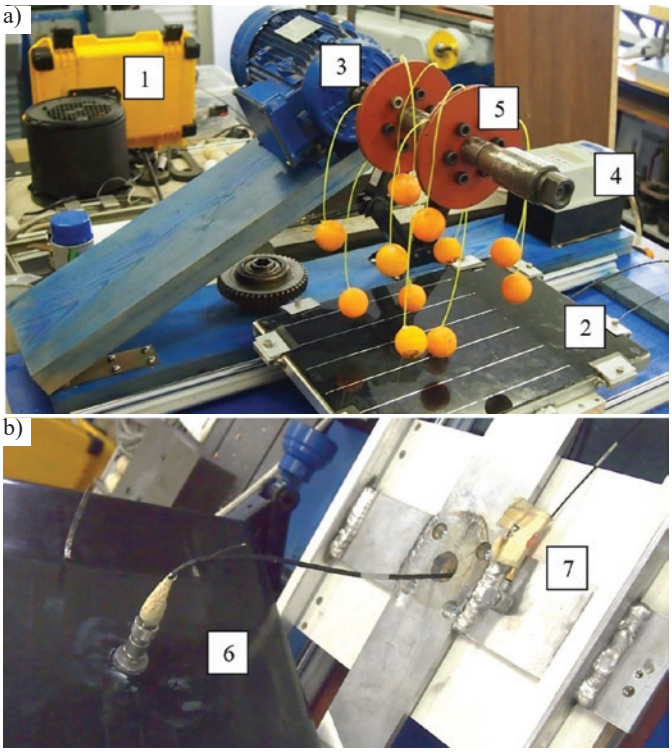


Fig. 2. (a) general view of the testbed and points of attachment of the accelerometer and the force transducer (b) of the experimental hail simulation testbed

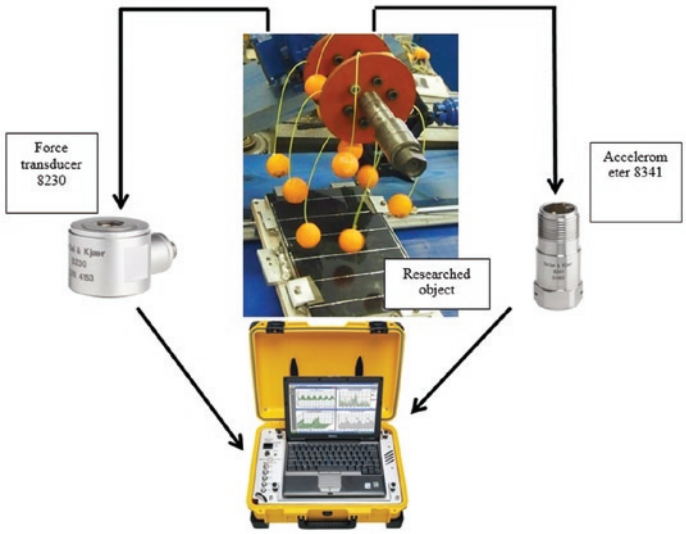


Fig. 3. Block diagram of the dynamic mechanical load generating facility

lation cube system (position 5). Figure 2 b illustrates the accelerometer (position 6) and the force transducer (position 7). Sensors were inspected using the best combination: small mass, small dimensions and high sensitivity (of acceleration: Piezoelectric Accelerometer Types 8341 (Voltage Sensitivity 100 mV/g; Amplitude Response 0.5 to 10000 Hz); Force Transducer Types 8230-002 (Voltage Sensitivity 2,2 mV/N).

Block diagram of the testbed is presented in Fig. 3.

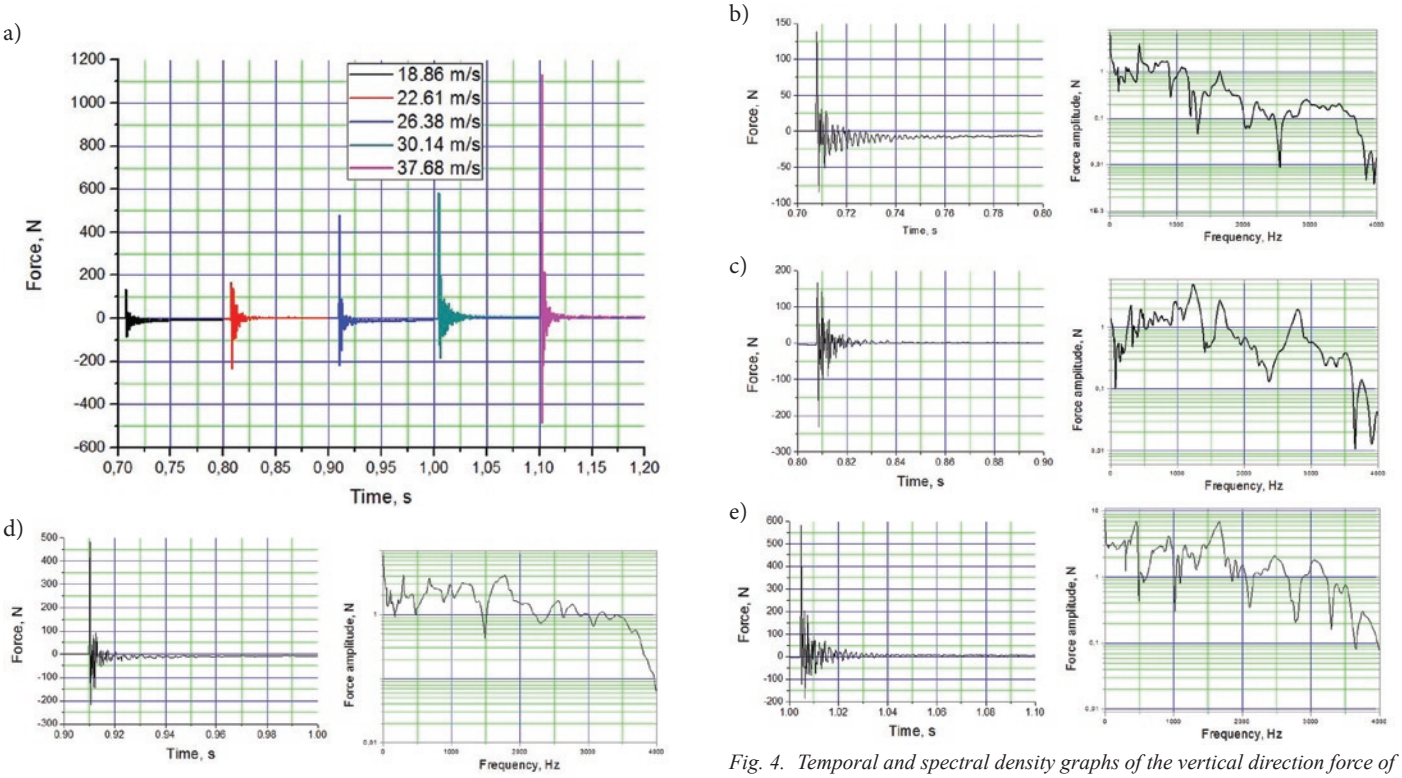


Fig. 4. Temporal and spectral density graphs of the vertical direction force of the solar cell midpoint in presence of one impact excitation: a – temporal excitation force graphs (at the linear ball speed v (18.86, 22.61, 26.38, 30.14 and 37.68 m/s)); b – temporal graphs of excitation force and its spectral density at the linear ball speed of 18.86 m/s; c –temporal graphs of excitation force and its spectral density at the linear ball speed of 22.61 m/s; d – temporal graphs of excitation force and its spectral density at the linear ball speed of 26.38 m/s; e – temporal graphs of excitation force and its spectral density at the linear ball speed of 30.14 m/s; f – temporal graphs of excitation force and its spectral density at the linear ball speed of 37.68 m/s

During the experiment, a cyclic mechanical load was applied when the frequency of revolutions of balls simulating hail ranged from 5 to 20 Hz, the amplitude of the impact excitation acceleration of the solar cell was up to 986 m /s^2 and the force amplitude - up to 1129 N. Such changes in parameters were sufficient to simulate hail conditions. The equipment used allows determining dynamic parameters of the researched object, which should be evaluated at the time of design and operation of solar cells. This would allow reducing the formation of micro-cracks in the crystalline structure of the solar cell.

3. Research results and discussions

The experiment was carried out simulating the impact of hail. The mechanical-dynamic effect was measured (excitation force and acceleration). The maximum amplitudes of excitation force are illustrated in Fig. 4, which presents the results of the measurement of the solar cell midpoint. Fig. 4 presents temporary force graphs and graphs of spectral density of force. The maximum received acceleration amplitudes are presented in Figure 5. Figure 5 illustrates temporal acceleration graphs and graphs of their spectral density. During the experiment, the radius r (from the axis of rotation to the point of impact) was 0.3 m, changing the linear speed of the ball *v* (18.86, 22.61, 26.38, 30.14 and 37.68 m / s).

Statistical parameters of the results of measurement of force and acceleration illustrated in Figures 4 and 5 are presented in Table 3.

Data presented in Figures 4 and 5 and Table 3 show experimental measurement results of the impact excitation force and response of the solar cell midpoint acceleration. The assessment of changes in the impact excitation force (Figure 4) at different speeds reveals that the increase in speed every 3.76 m/s (from 18.86 to 30.14 m/s) increases force values (by 20 % comparing the maximum values 137.927 and 166,073 N; 290 % comparing the maximum values 166,073 and 480,814 N; 21 % comparing the maximum values 480,814 and 582,064 N). Speed increase from 30.14 to 37.68 m/s increases force values by

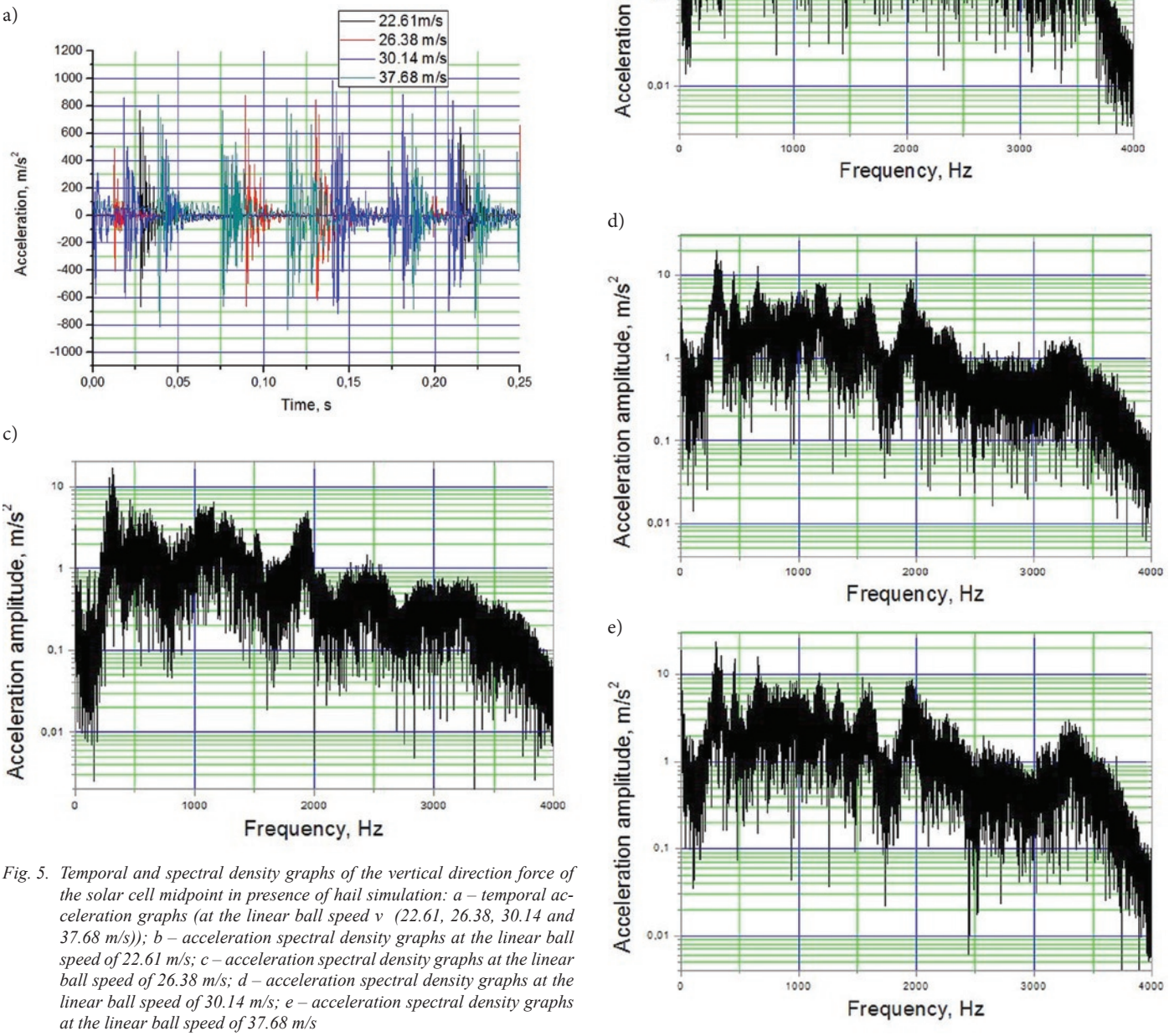


Fig. 5. Temporal and spectral density graphs of the vertical direction force of the solar cell midpoint in presence of hail simulation: a – temporal acceleration graphs (at the linear ball speed v (22.61, 26.38, 30.14 and 37.68 m/s)); b – acceleration spectral density graphs at the linear ball speed of 22.61 m/s; c – acceleration spectral density graphs at the linear ball speed of 26.38 m/s; d – acceleration spectral density graphs at the linear ball speed of 30.14 m/s; e – acceleration spectral density graphs at the linear ball speed of 37.68 m/s

Table 3. Statistical parameters of the results of measurement of excitation force and acceleration

Linear speed v, m/s	Excitation force, N		
	Standard Deviation	Minimum	Maximum
18.86	10,425	-84,536	137,927
22.61	17,795	-231,507	166,073
26.38	25,816	-216,714	480,814
30.14	32,211	-184,353	582,064
37.68	54,633	-483,972	1128,91
	Acceleration, m/s^2		
22.61	63,686	-781,358	832,698
26.38	78,589	-675,184	944,293
30.14	132,255	-729,196	986,152
37.68	156,956	-877,289	977,222

a) b)

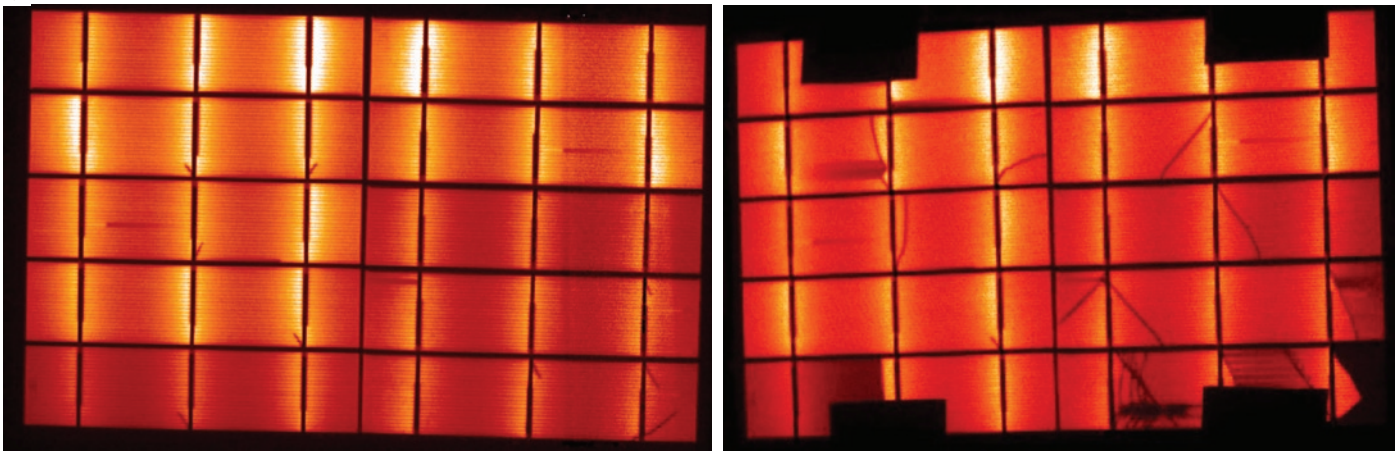


Fig. 6. Images of solar cells (a – before external effects; b – after external effects)

a) b)

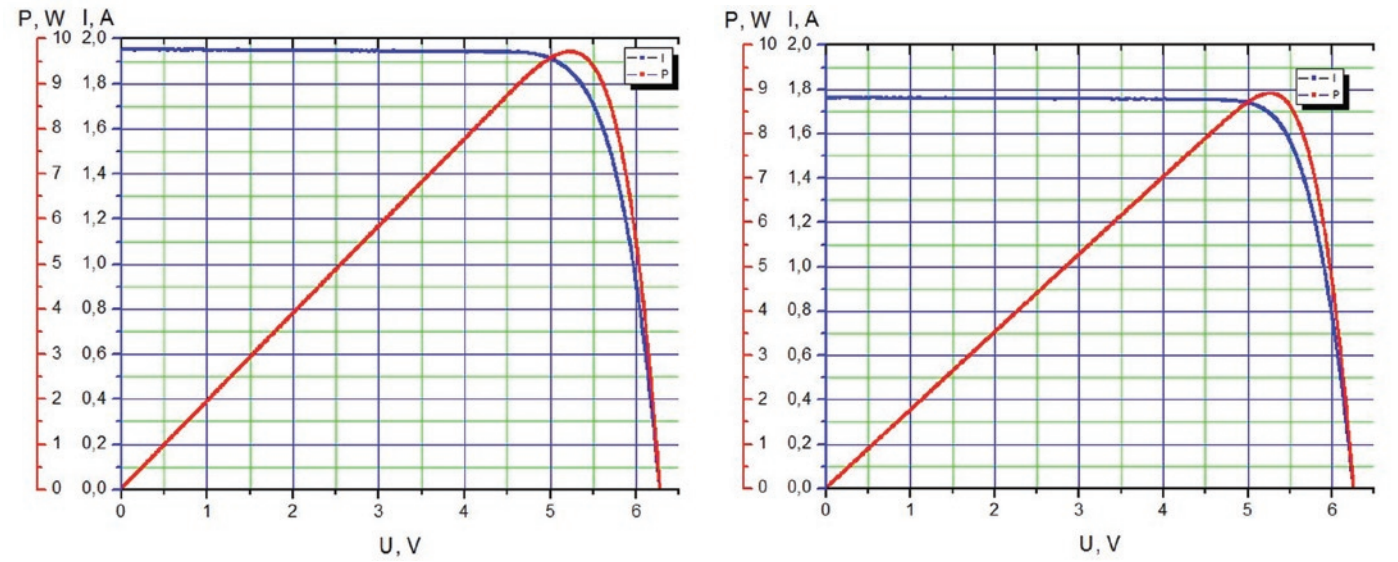


Fig. 7. Solar cell (from Figure 6) current and power curves (Current I(A); Power (W)) (a – before external effects; b – after external effects)

194 % comparing the maximum values 582,064 and 1128,91 N. In the assessment of standard deviation values of the solar cell midpoint response, the increase in speed from 22.61 to 26.38 m/s increases acceleration values by 23.4 % comparing the maximum values 63.686 and 78.589 m/s^2; respectively, the increase of speed from 26.38 to 30.14 m/s increases acceleration values by 68,3 % comparing the maximum values 78,589 and 132,255 m/s^2; respectively, the increase of speed

from 30.14 to 37.68 m/s increases acceleration values by 68.3 % comparing the maximum values 132,255 and 156,956 m/s^2.

Fig. 6 presents images of a photovoltaic module before and after hail simulation. It shows limitation use of solar elements in different areas of use for example in new solar panel for the air condition systems in new vehicles [22] etc.

Fig. 7 presents results of measurement of power of the solar cell.

As per Fig. 6 and 7, cracks formed in the crystalline structure of solar cells after dynamic-mechanical loads. Such defects may lead to minor failures in the solar cell energy current or, in the worst case scenario, the solar cell may not function at all. The similar results are presented in the another author's researches [23,24].

In the course of the experimental research, power before and after the external oscillation excitation effects was measured. The average power of the examined solar cell (6 pcs. were evaluated) before hail simulation was 9.8 W, and power after hail simulation was 8.8 W. The power decreased by about 11.2 %.

4. Conclusion

The use of the special testbed, which ensured the creation of mechanical-dynamic loads equivalent to hail conditions, allowed conducting a test of dynamic excitation of the solar cell. The advantage of the testbed is the fact that impact excitation is performed at an angle, which is close to actual hail conditions, because usually the operated elements are affixed at an angle.

Having carried out the hail simulation test, large damage was done to the researched solar cells, which manifested by cracks formed in the crystalline structure and a significant power loss by about 11.2 %.

References

1. Anders C, Urbassek H M. Solar wind ion impacts into ice surfaces: A molecular-dynamics study using the REAX force field. Icarus 2017; 282: 351–362, https://doi.org/10.1016/j.icarus.2016.09.037.
2. Atkinson C, Sansom C L, Almond H J, Shaw C P. Coatings for concentrating solar systems – A review. Renewable and Sustainable Energy Reviews 2015; 45: 113–122, https://doi.org/10.1016/j.rser.2015.01.015.
3. Du B, Yang R, He Y, Wang F, Huang S. Nondestructive inspection, testing and evaluation for Si-based, thin film and multi-junction solar cells: An overview. Renewable and Sustainable Energy Reviews 2017; 78: 1117–1151, https://doi.org/10.1016/j.rser.2017.05.017.
4. El Mghouchi Y, Chham E, Krikiz M S, Ajzoul T, El Bouardi A. On the prediction of the daily global solar radiation intensity on south-facing plane surfaces inclined at varying angles. Energy Conversion and Management 2016; 120: 397–411, https://doi.org/10.1016/j.enconman.2016.05.005.
5. Guo B, Javed W, Pett C, Wu C-Y, Scheffe J R. Electrodynamic dust shield performance under simulated operating conditions for solar energy applications. Solar Energy Materials and Solar Cells 2018; 185: 80–85, https://doi.org/10.1016/j.solmat.2018.05.021.
6. Gunduz H, Jayaweera D. Reliability assessment of a power system with cyber-physical interactive operation of photovoltaic systems. International Journal of Electrical Power & Energy Systems 2018; 101: 371–384, https://doi.org/10.1016/j.ijepes.2018.04.001.
7. Hills J M, Michalena E. Renewable energy pioneers are threatened by EU policy reform. Renewable Energy 2017; 108: 26–36, https://doi.org/10.1016/j.renene.2017.02.042.
8. Kilikevičius A, Čereška A, Kilikevičienė K. Analysis of external dynamic loads influence to photovoltaic module structural performance. Engineering Failure Analysis 2016; 66: 445–454, https://doi.org/10.1016/j.engfailanal.2016.04.031.
9. Kulesza G, Panek P, Zięba P. Time efficient texturization of multicrystalline silicon in the HF/HNO3 solutions and its effect on optoelectronic parameters of solar cells. Archives of Civil and Mechanical Engineering 2014; 14(4): 595–601, https://doi.org/10.1016/j.acme.2014.02.007.
10. Martins A C, Chapuis V, Virtuani A, Li H-Y, Perret-Aebi L-E, Ballif C. Thermo-mechanical stability of lightweight glass-free photovoltaic modules based on a composite substrate. Solar Energy Materials and Solar Cells 2018; 187: 82–90, https://doi.org/10.1016/j.solmat.2018.07.015.
11. Picotti G, Borghesani P, Cholette M E, Manzolini G. Soiling of solar collectors – Modelling approaches for airborne dust and its interactions with surfaces. Renewable and Sustainable Energy Reviews 2018; 81: 2343–2357, https://doi.org/10.1016/j.rser.2017.06.043.
12. Polimeno M R, Roselli I, Luprano V A M, Mongelli M, Tatì A, De Canio G. A non-destructive testing methodology for damage assessment of reinforced concrete buildings after seismic events. Engineering Structures 2018; 163: 122–136, https://doi.org/10.1016/j.engstruct.2018.02.053.
13. Popoola I K, Gondal M A, Qahtan T F. Recent progress in flexible perovskite solar cells: Materials, mechanical tolerance and stability. Renewable and Sustainable Energy Reviews 2018; 82: 3127–3151, https://doi.org/10.1016/j.rser.2017.10.028.
14. Punge H J, Kunz M. Hail observations and hailstorm characteristics in Europe: A review. Atmospheric Research 2016; 176–177: 159–184, https://doi.org/10.1016/j.atmosres.2016.02.012.
15. Reshma Gopi R, Sreejith S. Converter topologies in photovoltaic applications – A review. Renewable and Sustainable Energy Reviews 2018; 94: 1–14, https://doi.org/10.1016/j.rser.2018.05.047.
16. Smyth M, Pugsley A, Hanna G, Zacharopoulos A, Mondol J, Besheer A, et al. Experimental performance characterisation of a Hybrid Photovoltaic/Solar Thermal Façade module compared to a flat Integrated Collector Storage Solar Water Heater module. Renewable Energy [Internet]. 2018 Apr [cited 2018 Oct 26]; Available from: https://linkinghub.elsevier.com/retrieve/pii/S0960148118304245.
17. Sun L, Gu X H, Song P, Di Y. A generalized equivalent temperature model in a time-varying environment. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017, 19 (3): 432–440, https://doi.org/10.17531/ein.2017.3.14.
18. Stratakis E, Kymakis E. Nanoparticle-based plasmonic organic photovoltaic devices. Materials Today 2013; 16(4): 133–146, https://doi.org/10.1016/j.mattod.2013.04.006.
19. Teotia M, Soni R K. Applications of finite element modelling in failure analysis of laminated glass composites: A review. Engineering Failure Analysis 2018; 94: 412–437, https://doi.org/10.1016/j.engfailanal.2018.08.016.
20. Višniakov N, Kilikevičius A, Novickij J, Grainys A, Novickij V. Low-cost experimental facility for evaluation of the effect of dynamic mechanical loads on photovoltaic modules. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2015; 17(3): 334–337, https://doi.org/10.17531/ein.2015.3.2.

21. Zhang C, Wu J, Long C, Cheng M. Review of Existing Peer-to-Peer Energy Trading Projects. Energy Procedia 2017; 105: 2563–2568, https://doi.org/10.1016/j.egypro.2017.03.737.
22. Setiyo M, Soeparman S, Wahyudi S, Hamidi N. The Alternative Way to Drive the Automobile Air-Conditioning, Improve Performance, and Mitigate the High Temperature: A Literature Overview. Periodica PolytechnicaTransportation Engineering 2018; 46(1): 36-41, https://doi.org/10.3311/PPtr.8892.
23. Trzmiel G. Determination of a mathematical model of the thin-film photovoltaic panel (CIS) based on measurement data. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19 (4): 516–521, https://doi.org/10.17531/ein.2017.4.4.
24. Smolinski M., Perkowski T.,Mystkowski A., Dragašius E., Eidukynas D., Jastrzebski R P. AMB flywheel integration with photovoltaic system for household purpose – modelling and analysis. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19 (1): 86–94, https://doi.org/10.17531/ein.2017.1.12.

Kristina KILIKEVIČIENĖ
Department of Mechanical and Material Engineering
Faculty of Mechanical Engineering
Vilnius Gediminas Technical University
Basanavičiaus str. 28, LT-03224 Vilnius, Lithuania

Jonas MATIJOŠIUS
Department of Automobile Engineering
Faculty of Transport Engineering
Vilnius Gediminas Technical University
Basanavičiaus str. 28, LT-03224 Vilnius, Lithuania

Antanas FURSENKO
Department of Mechanical and Material Engineering
Faculty of Mechanical Engineering
Vilnius Gediminas Technical University
Basanavičiaus str. 28, LT-03224 Vilnius, Lithuania
E-mail:

Artūras KILIKEVIČIUS
Institute of Mechanical Science
Faculty of Mechanical Engineering
Vilnius Gediminas Technical University
Basanavičiaus str. 28, LT-03224 Vilnius, Lithuania

E-mails: kristina.kilikeviciene@vgtu.lt, jonas.matijosius@vgtu.lt, antanas.fursenko@vgtu.lt, arturas.kilikevicius@vgtu.lt

Article citation info:
GOŁAŚ A, CIESIELKA W, SZOPA K, ZYDROŃ P, BĄCHOREK W, BENESZ M, KOT A, MOSKWA S. Analysis of the possibilities to improve the reliability of a 15 kV overhead line exposed to catastrophic icing in Poland. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 282–288, http://dx.doi.org/10.17531/ein.2019.2.13.

Andrzej GOŁAŚ
Wojciech CIESIELKA
Krystian SZOPA
Paweł ZYDROŃ
Wojciech BĄCHOREK
Mariusz BENESZ
Aleksander KOT
Szczepan MOSKWA

ANALYSIS OF THE POSSIBILITIES TO IMPROVE THE RELIABILITY OF A 15 KV OVERHEAD LINE EXPOSED TO CATASTROPHIC ICING IN POLAND

ANALIZA MOŻLIWOŚCI POPRAWY NIEZAWODNOŚCI NAPOWIETRZNEJ LINII 15 KV NARAŻONEJ NA KATASTROFALNE OBLODZENIE W WARUNKACH POLSKICH*

The paper is a result of a synergic cooperation of two academic teams, i.e. power engineering and mechanical teams, and a distribution system operator. A real 15 kV overhead line exposed to a catastrophic load of ice and rime was analyzed and three solutions to improve the reliability of the tested object in such conditions were examined. Authors considered: shortening the length of the line spans, heating the main line with increased current and rebuilding the overhead line to a cable line. The researches worked out a FEM model taking into account the newest normatives, simulated the model, experimentally increased the load on the real line with measured wire temperature, and performed multi-variant calculations to determine indicators of reliability, i.e. SAIDI and SAIFI. The analyses were followed by conclusions thanks to which the reliability of power lines exposed to catastrophic icing could be increased. These inferences should be considered and applied by all distribution system operators in Poland.

***Keywords**: power distribution networks, overhead lines, reliability, icing, rime.*

Praca jest efektem synergicznej współpracy dwóch zespołów akademickich: elektroenergetycznego i mechanicznego oraz operatora systemu dystrybucyjnego. Analizie poddano rzeczywistą, napowietrzną linię średniego napięcia 15 kV narażoną na katastrofalne obciążenia lodem i szadzią. Zbadano możliwość zastosowania trzech rozwiązań mogących poprawić niezawodność badanego obiektu w takich warunkach. Rozważono: skrócenie długości przęseł linii, podgrzewanie magistrali zwiększonym prądem roboczym oraz przebudowę linii do linii kablowej. W celu realizacji pracy wykonano badania modelowo-symulacyjne MES z uwzględnieniem najnowszych wytycznych normatywnych, zrealizowano eksperyment dociążenia linii wraz z pomiarem temperatury przewodów oraz przeprowadzono wielowariantowe obliczenia niezawodnościowe prowadzące do wyznaczenia wskaźników SAIDI i SAIFI. W wyniku szczegółowych analiz sprecyzowano wnioski końcowe pozwalające na zwiększenie niezawodności linii elektroenergetycznych narażonych na katastrofalne oblodzenie, które powinny być rozważone i stosowane przez wszystkich operatorów systemów dystrybucyjnych w Polsce.

***Słowa kluczowe**: elektroenergetyczne sieci dystrybucyjne, linie napowietrzne, niezawodność, oblodzenie, szadź.*

1. Introduction

Electricity supplies are a very important element of living now. Reliability and continuity of energy supply, as well as minimization of power outages, are currently one of the key priorities of distribution system operators.

In view of the observed climate extremization, the problem of overhead network infrastructure exposure to difficult environmental conditions (icing, winds) is of particular importance.

The most important problem encountered by the designers and constructors of overhead lines lies in precise determining of additional loads of lines caused by ice and/or rime deposited on wires and supporting structures. This issue was dealt with, among other things, during the standardization of these loads in IEC [15].

In recent years, a lot of blackouts caused by the impact of adverse weather conditions on the elements of power systems have been noted over the whole world. Most of the world's countries located on the northern hemisphere, such as China, Germany [2], [11], north-east part of the USA and Canada [5], Czech Republic [17], North Caucasus in Russia [10], Japan [1], Finland [14], Norway [7], Romania [9], Hungary [13], United Kingdom [20], Iceland [4], [8] or part of India are struggling with the problem of ice formation on overhead lines. Also Poland experienced catastrophic failures caused by extreme weather conditions, as described in [2] and [11].

The gravity of this problem was noted and discussions followed on the international forum, e.g. CIGRE materials paid much attention to this issue [18], [19]. It is also the main topic of periodic meetings within the International Workshop on Atmospheric Icing of Structures.

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

With this experience in mind, national committees of numerous countries decided to change the standards to be used by the designers of the power lines in their calculations. In Poland, the ice loading zones have been also changed [16].

Reports in the world literature document a number of methods and techniques aimed at cooling ice and removing ice and/or raking from overhead lines [6], however, their effectiveness is negligibly low in the extreme conditions of catastrophic icing. These solutions relate mainly to the transmission networks.

The present work examines three solutions that improve the reliability of the object exposed to catastrophic load icing. Special attention was paid to: reducing span length (consolidation of supporting structures), heating of the main line by an increased current and rebuilding of the overhead line to the cable line. The researches included a FEM model, taking into account the newest policy normative, simulation based on the very model, experimental increase of load on the real line with measured wire temperature, and multi-variant calculations of reliability indicators, i.e. SAIDI and SAIFI. The final conclusions give information on how to increase the reliability of the analyzed power lines.

2. Characteristics of the object

The analysis was performed on a real 15 kV overhead distribution line located in the southern part of Poland. The schematic of the line is shown in Figure 1. This is an overhead line which supplies 11 lateral branches. Three of them include cable sections. Lateral branches are marked with the successive letters. The main line is supplied from the substation PZ1, and the tie point (normally open switch) is located in the substation PZ2.

Selected data characterizing the tested MV line are presented in Table 1.

Table 1. Basic data of the analyzed 15 kV network

	Type of construction / type	Length/Number	Sum
Main line	overhead line with AFL wires	9.9 km	9.9 km
	cable line	0 km	
Lateral branches	overhead line with AFL wires	7.8 km	12.3 km
	cable line	4.5 km	
Support structures (poles)	reinforced concrete	12	174
	steel	4	
	prestressed	41	
	centrifuged	114	
	wooden	3	
Switches	manually controlled	25	28
	radio controlled	2	
	reclosers	1	
Number of MV/LV transformer stations	19		
Number of customers	945		
Annual load range (min - max)	6 A – 16 A		

In winter of 2010 the analyzed overhead line experienced catastrophic icing conditions that led to extensive destruction of the network infrastructure.

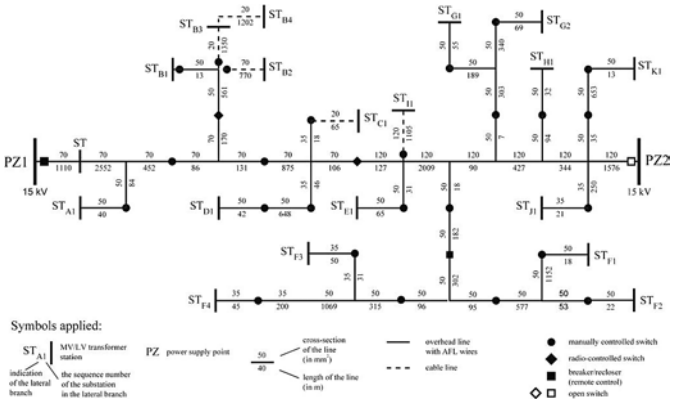


Fig. 1. Diagram of analyzed 15 kV power line

3. FEM simulation studies of selected elements of the line infrastructure for catastrophic weather conditions

The most important element of FEM model-simulation was to take into account the real catastrophic impacts that occurred in a large area on which the analyzed power line was located. In January 2010 a weather anomaly occurred contributing to the formation of icing on the overhead lines, significantly exceeding the normative assumptions (up to 18 kg/m and 15 cm in diameter). The effect of such catastrophic icing on the reaction forces at the suspension points of the wires and the force in the wire itself has been presented in this section. As the formation of such a large rime is favored by a relatively small wind, attention was mainly paid to cases involving only ice.

The purpose of the simulation tests was to determine the impact of atmospheric loads on the supporting structures and wires of the line. It was shown how the asymmetry of icing and length of the spans influenced the change of forces in the layout. The impact of these forces on the cross arms of the structure was not analyzed. The supporting structures were loaded with a static layout of forces equivalent to the forces coming from the wires to the cross arm.

First of all, a numerical model of the pole was built and the correctness of the assumptions verified.

The test was performed on a section of the analyzed 15 kV line. The pole E-12/10 was used as a support structure, which means that its length equaled to 12 m (Fig. 2), and the nominal peak force it carried was 10 kN. The structure was made of concrete class C 40/50. The steel reinforcing bars, running from the base to the apex, corresponded with the standards for reinforced concrete constructions.

The cross-section of the pole had a ring shape, the diameter of which decreased in proportion to the height of the structure (Fig. 2). The outer diameter of the pole at its base was 398 mm, and in the apical part 218 mm. 16 steel bars with a diameter of 16 mm were distributed evenly over the circumference of the pole. The size of the cross-section mesh was chosen 3 mm for reinforcing bars and 10 mm in the concrete area (Fig. 3).

The finite element model was verified and calculations were made for the destructive force $P_n = 18$ kN specified by the manufacturer. The maximum compressive stresses in concrete were 47.5 MPa (Fig. 4a),

Fig. 2. The geometry of pole E 12/10

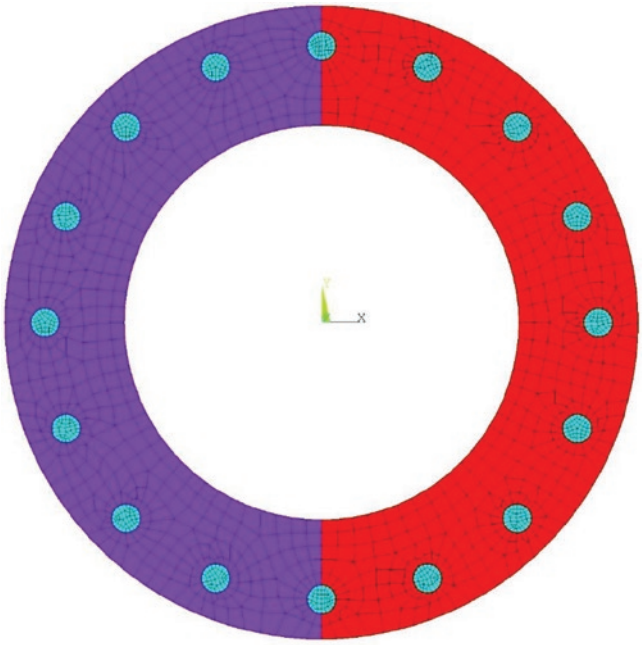


Fig. 3. The cross-section of E 12/10 pole viewed with the applied mesh

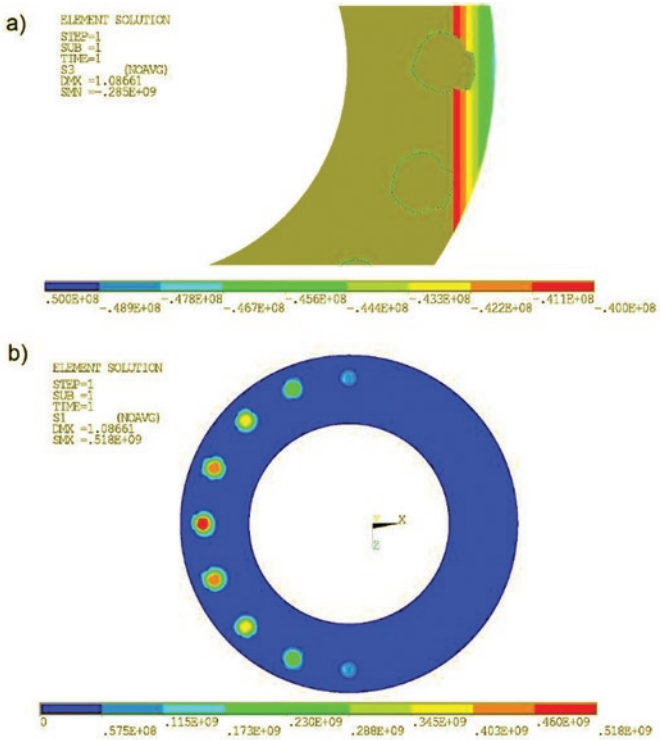


Fig. 4. Stress in power pole E 12/10 for tip bending force a) compressive in concrete, b) stretching in bars

which in principle coincided with the limit compressive stress for C 40/50 concrete equal to 48 MPa. So it should be assumed that the model was made correctly.

The parameters of the supporting structure E12/10 were calculated in view of the load determined for the suspension pole and the strain pole. Wire AFL-6 70 was selected for phase conductors because bare wires are more susceptible to the occurrence of high icing, and additionally, this type of cable is still being used in MV lines. The calculations were carried out for a replacement span length of 80 m. This value is approximately the average distance between poles used in zones with big characteristic icing, on the other hand, is adequate enough to show how large forces can occur in a relatively short span.

The results obtained for the described case are presented in table 2. Such a large mass of icing causes very large vertical forces and the axial tension in the wire. Again, with balanced ice in adjacent spans, compressive stress is small, but problems begin to appear for cases of asymmetrical icing. If the support structure is made as a suspension pole, as in case 2c (i.e., an unbalanced longitudinal load), the compressive stresses exceed the permissible values and crush the concrete and destroy the support structure. If the support structure is made as a strain pole, then it must also satisfy the conditions of group 5. Even in a situation when a single conductor is broken (case 5a), such a large transverse force appears that the compressive stresses in concrete are on the strength limit of the structure. In contrast, if the pole was designed to transfer a single-sided horizontal tension, the icing would generate a catastrophic failure of the structure. The situations described in this paragraph assume that the wire may break (cases: 5a, 5b, 5c), or withstand the icing load of 18 kg/m (cases: 2a, 2b). The design breaking force for AFL-6 70 is 22.75 kN. The table does not include sags due to the possibility of entering the scope of plastic deformation.

The given example presents results for a specific span length, sag and icing cases, therefore additional calculations were made to determine the relationship between the magnitude of lateral forces acting on the supporting structure E 12/10 and the span's length.

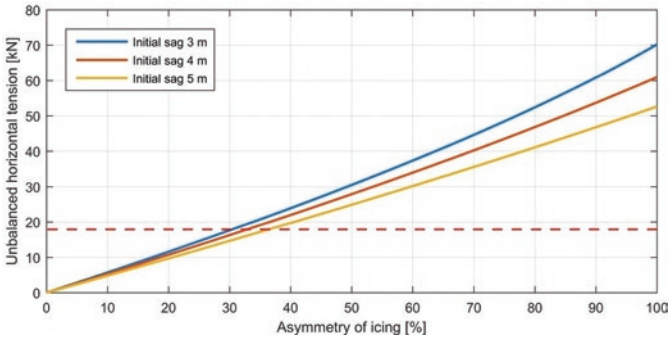


Fig. 5. The effect of the asymmetry of icing on the unbalanced horizontal tension

However, the first test was made on the effect of asymmetric icing on unbalanced tension. In this situation all the wires of one span are loaded with maximum catastrophic icing of 18 kg/m, while icing in the neighboring span changes from the maximum value ψ_I= 1 to 0. The asymmetry factor is taken here as $(1\text{-}\psi_I)\cdot 100\%$, so for the maximum icing of both spans, this coefficient is 0%, while if one span is noticed, the asymmetry of the icing load is 100%. The analysis of the dependencies in Figure 5 reveals that for spans ≥80 m long, already with an asymmetry coefficient of 30%-37% (depending on the initial sag), the unbalanced value of tension force (from three wires) reaches the critical value of the destructive force for the pole E12/10 equal to 18 kN. This confirms previous results, where the problem was usually not so much icing, but its uneven distribution in the neighboring spans.

Subsequently, the maximum length of the span was checked for a load of 18 kg/m, assuming that the strain pole was designed to transfer one-sided tension (case 5c). The graphs presented in Figure 6 show that already for a span length of 35 m and an initial sag of 3 m, the sum of the tension components coming from three wires reaches the value of the destructive force of the chosen support structure. For comparison, assuming the maximum load according to the

Table 2. Effect of forces exerted by 18 kg/m catastrophic icing conduits on E 12/10 support structure

Case	V	H	W	N	σ_C	σ_S
	[N]	[N]	[N]	[N]	[MPa]	[MPa]
0	106.0	706.3	0	714.2	0.19	0
2a	4996.4	24133.0	0	24645.0	0.51	0
ψ_1= 0.3	157 3.1	9258.7	0	9391.4	-	-
ψ_1= 0.5	2551.2	13979.2	0	14209.1	-	-
ψ_1= 0.7	3529.3	18264.1	0	18602.3	-	-
2b	Combination of loads ψ_1= 0.5 and ψ_1=1				7.65	48.9
2c	Combination of loads ψ_1= 0.3 and ψ_1=0.7				71.10	730.0
5a	Wire breakage for icing ψ_1= 0.7				48.21	492.2
5b	75% of single-sided tension with icing ψ_1= 1				142.25	1471.0
5c	Total unilateral tension at an icing ψ_1= 1				189.20	1962.8

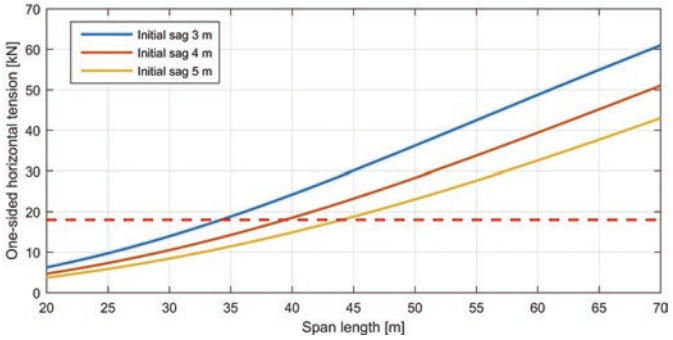


Fig. 6. The one-sided horizontal tension as a function of span length for strain pole and catastrophic icing 18 kg/m (case 5c)

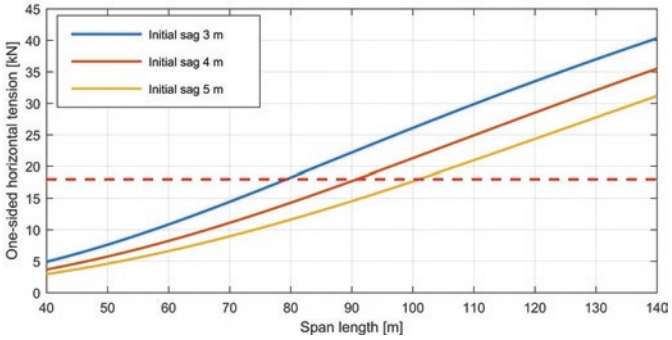


Fig. 7. The one-sided horizontal tension as a function of span length for strain pole and normative maximum icing (case 5c)

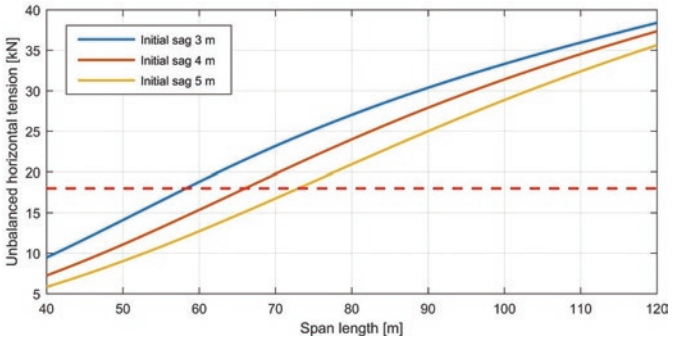


Fig. 8. The unbalanced horizontal tension as a function of span length for catastrophic icing 18 kg/m unbalanced longitudinally (case 2c)

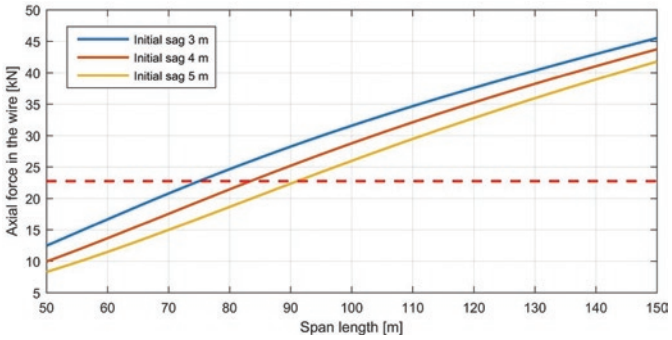


Fig. 9. The maximum axial force in a single wire as a function of span length for catastrophic icing of 18 kg/m

normative guidelines, the length of the span for pole type E12/10 could be 80 m (Fig. 7).

If the pole is a suspension pole, case 2c will be most dangerous for the considered conditions. A longitudinally unbalanced icing case will be a threat to a 58 m long equivalent span with an initial overhang of 3 m (Fig. 8).

When considering such a large icing one should obviously account for the maximum forces in wires as they are a potential source of breakage. For a wire AFL-6 70, the design breaking force of the conductor is 22.75 kN. If the wire is loaded with ice of 18 kg/m for an initial overhang of 3 m, it may break at the length of the span of 75 m (Fig. 9).

The analysis shows that the appearance of such unusual atmospheric conditions as in 2010, is accompanied by icing, the mass of which goes over the normative guidelines, even for the third ice load zone. E-type poles, which are characterized by high durability as compared to BSW constructions, may be too weak to carry such heavy loads and break. Of course, the worst case for a power line segment is when due to an excessive load, the wires are broken in one of the spans, the pole is loaded with one-sided tension and subjected to high bending forces. These are instances of 5 interaction systems. Although

they refer to strain poles, it may happen that due to large icing the wire will not be able to move freely in the support or disengage from the safety holder, thereby transferring the tension on the suspension pole, and leading to its destruction.

4. Field studies using thermovision

In the next step, authors concentrated on measuring the temperature of the wires of the analyzed 15 kV line under normal and increased load conditions. The measurements were conducted on phase wires AFL-6 70. The experiment was performed on 23 Nov. 2017 for a normal configuration of the system and a specially prepared network configuration with additional load.

The test results, i.e. current, ambient temperature and wire temperature recorded during the experiment are presented in Figure 10. For safety reasons, i.e. to ensure the continuity of power supply and guarantee adequate power quality, daytime hours were selected for the test. The measurements were carried out in windless conditions, with a little cloudy sky, clouds of the middle floor appearing after 9.30 a.m. The changing degree of insolation caused a change in the temperature during the experiment. In the initial experiments between 7:45 and 9:45 the ambient temperature increased from 0.3°C to 7.9°C. In during this period, the current value varied from 8 A to 10 A, and temperature of wire from 6.3°C to 9.2°C.

At 9:52 a.m. the MV line load was increased (change of network configuration), and the normal configuration restored at 01:25 p.m. At 10.00 a.m. a current of 47 A was observed. While increasing the load, the current value stayed in the interval 46 A to 51 A. At that time the observed temperature of the wire varied from 11.3°C to 17.3°C. In the observed period, the value of the ambient temperature also changed from 7.4°C to 15.6°C.

Particularly noteworthy are the measurements between 9:45 and 10:45 a.m., when the highest increase of wire temperature (8.5°C to 14.8°C) was observed. It was accompanied by an increase of current from 10 A to 48 A, and an ambient temperature increase by 2.3°C.

The analysis of the test results revealed that during the change of the load from 36 A to 40 A, the temperature of the wire increased from 4.3°C to 6.1°C.

In addition to the wire temperature rise, a significant deterioration of the voltage conditions in end-customers was also observed. Based on analyses of the year value of current and voltage waveforms in a selected network, thermal imaging experiments, and voltage drops, it was found that the implementation of the additional load for winter conditions not make the temperature of wires rise to positive values (guaranteeing the formation of an ice-and-rash coat). In this situation no main line heating is possible for stripping ice and/or rime from the wire.

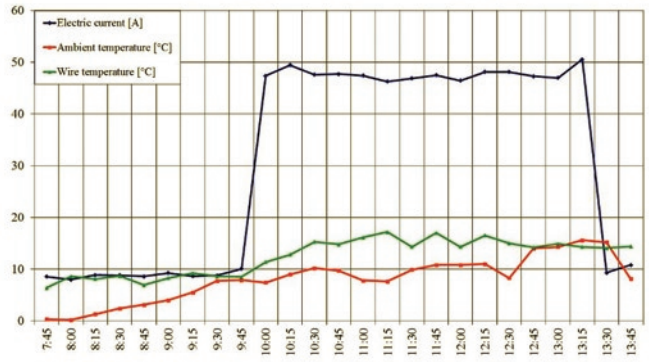


Fig. 10. The results of measurements of 23 Nov. 2017 for the tested 15 kV line during its experimental weighting

5. Structural reliability model

The reliability calculations were performed to determine SAIDI and SAIFI indicators for the analyzed MV line. For this purpose an author's dedicated software was employed. It was used for multivariate calculations, on the basis of which various scenarios of facility modernization could be quantitatively and reliably assessed.

The simulation requires information concerning the grid structures as well as the parameters of the reliability of components.

In terms of reliability parameters of the elements the same values of the average duration of failures were assumed. For the failure rate, two variants of calculations were implemented. Variant A was based on the relation of parameters after publication [3] and [21] whereas, in option B, the indicators were taken from [12]. The reliability parameters have been compiled in table 3.

The results of calculations of SAIDI and SAIFI reliability indicators are presented in Figure 11. The values of indicators for the current state constitute a reference level for all simulated cases.

Table 3. Used failure parameters of the MV distribution network

Element	Unit *j*	Failure rate -variant A [1/(*j*· year)]	Failure rate -variant B [1/(*j*· year)]	The average duration of failure [h]
MV overhead line (AFL)	100 km	8.14	11.02	5
MV cable line		0.814	10.16	5

In variant B, the successive rebuilt of the overhead line to the cable line, minimally improved the reliability, whereas in variant A, a significant improvement in the reliability of the facility was observed as a result of its reconstruction and modernization. Therefore the latter set of parameters seems more correct in the context of the operational realities of distribution networks.

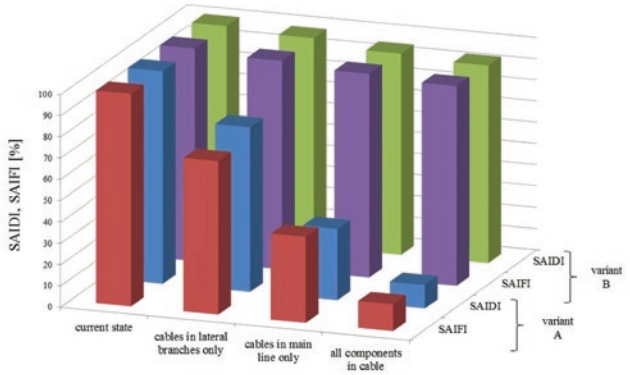


Fig. 11. The expected rate of reduction of reliability indicators for various variants of reconstruction of the analyzed 15 kV line

6. Summary and conclusions

The work tackles an extremely important problem of operation of medium voltage lines exposed to catastrophic icing. In Poland, that kind of line is located in the operation area of four distribution system operators, covering over 20000 km of length.

The aim of the work was to identify methods to improve the reliability of the analyzed object, counteracting the effects of disastrous icing. For this purpose FEM model simulation, infrared temperature measurement of wires and multivariate calculations of SAIDI and SAIFI indicators were performed.

The analysis of the FEM model simulation indicates that the analyzed object can withstand catastrophic loads, provided the span lengths are reduced to the length of 35 meters.

The experimental (planned) increase of current load in a wire and the thermovision measurements reveal that the effective use of preventive heating of the line with an additional current is not possible in the analyzed case. Line load in the normal work configuration is negligible, and reconfiguration thrust is significantly limited by voltage conditions.

Substantial elimination of climatic impacts (both winter and summer) may be achieved by the reconstruction of the analyzed object into a cable line. Apart from the effect of full climatic screening it will undoubtedly influence the operational reliability of the facility, and the degree of this impact will be significantly conditioned by the reliability parameters of elements of this network.

The overhead networks dominate in the national distribution network, therefore some strategy of dealing with the problem of impact of extreme atmospheric phenomena on overhead distribution infrastructure has to be worked out. Such a strategy should take into account a number of factors such as the state of technical infrastructure, percent of amortization of cushion property, local setting conditions, and functions of individual elements in the distribution network.

Concluding, the desired guidance presented in this paper allows for increasing the reliability of distribution power lines exposed to catastrophic icing and should be considered and implemented by all distribution system operators in Poland.

References

1. Admirat P, Sakamoto Y. Calibration of a wet snow model on real cases in Japan and France. 4th International Workshop on Atmospheric Icing of Structures (IWAIS), Paris, France 1988; 7-13.
2. Ciesielka W, Czajka I, Filipek R, Gołaś A, Hamiga W, Romik D, Suder-Dębska K, Szopa K, Wołoszyn J. Smart Grid in energetic facilities: modelling, monitoring and diagnostics. Monography of the Department of Power Systems and Environmental Protection. Faculty of Mechanical Engineering and Robotics AGH, Krakow 2017.
3. Damchi Y, Sadeh J. Effect of combined transmission line (overhead line/cable) on power system reliability indices. 4th International Power Engineering and Optimization Conference (PEOCO), Shah Alam, Malaysia 2010; 59–63, https://doi.org/10.1109/PEOCO.2010.5559223.
4. Elíasson A J, Thorsteins E, Ólafsson H. Study of wet snow events on the South Coast of Iceland. 9th International Workshop on Atmospheric Icing of Structures (IWAIS), Chester, United Kingdom 2000.
5. Farzaneh M, Savadjiev K. Icing Events Occurrence in Québec: Statistical analysis of field data. International Journal of Offshore and Polar Engineering 2001; 11(1): 9–15.
6. Farzaneh M. Atmosferic Icing of Power Networks. Springer Science+Business Media B.V., 2008.
7. Fikke S M, Johansen O S. Earlier Norwegian iceload research. A review of investigations and results. 2nd International Workshop on Atmospheric Icing of Structures (IWAIS), Trondheim, Norway 1984; 11-18.
8. Fikke S M et al. COST Action 727 Atmospheric icing on structures. Measurements and data collection on icing. State of the art. Veröffentlichung MeteoSchweiz 2007; 75.
9. Goia M L. Damages caused by icing and wind to the Romanian OEL. 9th International Workshop on Atmospheric Icing of Structures (IWAIS), Chester, United Kingdom 2000.
10. Golikova T N, Toporkava G D, Nikitina L G. Ascertaining ice-load maps of the USSR territory. Trans Improving the reliability of high voltage lines. Energoatomizdat, Moscow 1989; 107–122.
11. Gołaś A, Ciesielka W, Czajka I, Czechowski M, Filipek R, Suder-Dębska K, Szopa K, Śliwiński M, Wołoszyn J, Żywiec W. Mechanical engineering in Smart Grid technology. Monography of the Department of Power Systems and Environmental Protection. Faculty of Mechanical Engineering and Robotics AGH, Krakow 2015.
12 Kornatka M. Analysis of the exploitation failure rate in Polish MV networks. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20(3): 413–419. https://doi.org/10.17531/ein.2018.3.9
13. Krómer I. Hungarian icing activity survey. 6th International Workshop on Atmospheric Icing of Structures (IWAIS), Budapest, Hungary 1993; ix-x.
14. Lehtonen P, Ahti K, Makkonen L. The growth and disappearance of ice loads on a tall mast. 3rd International Workshop on Atmospheric Icing of Structures (IWAIS), Vancouver, Canada 1986; 363–368.
15. Overhead Lines – Meteorological Data for Assessing Climatic Loads, 1997; International Electrotechnical Commission Technical Report 61774, First edition: 1997–2008.
16. PN-EN 50341-2-22:2016-04 Elektroenergetyczne linie napowietrzne prądu przemiennego powyżej 1 kV - Część 2-22: Krajowe Warunki Normatywne (NNA) dla Polski.
17. Popolansky F. Economical aspects of ice failures caused in power transmission on the territory of former Czechoslovakia. 9th International Workshop on Atmospheric Icing of Structures (IWAIS), Chester, United Kingdom 2000.
18. Technical Brochure CIGRE - Guidelines for field measurement of ice loadings on power line conductors, 2001; CIGRE TB No 179.
19. Technical Brochure CIGRE - Big storm events. What we have learned?, 2008; CIGRE TB No 344.
20. Wareing B J, Chetwood P. Ice load data from Deadwater Fell. 9th International Workshop on Atmospheric Icing of Structures (IWAIS), Chester, United Kingdom 2000.
21. Zhu D, Broadwater R P, Tam K, Seguin R, Asgeirsson H. Impact of DG placement on reliability and efficiency with time-varying loads, IEEE Transactions on Power Systems 2006; 21(1): 419–427, https://doi.org/10.1109/TPWRS.2005.860943.

Andrzej GOŁAŚ
Wojciech CIESIELKA
Krystian SZOPA
AGH University of Science and Technology
Faculty of Mechanical Engineering and Robotics
Department of Power Systems and Environmental Protection Facilities
Al. Mickiewicza 30, 30-059 Kraków, Poland

Paweł ZYDROŃ
Wojciech BĄCHOREK
Mariusz BENESZ
Aleksander KOT
Szczepan MOSKWA
AGH University of Science and Technology
Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering
Department of Electrical and Power Engineering
Al. Mickiewicza 30, 30-059 Kraków, Poland

E-mails: ghgolas@cyf-kr.edu.pl, ghciesie@cyf-kr.edu.pl, kszopa@agh.edu.pl, pzydron@agh.edu.pl, wojbach@agh.edu.pl, mben@agh.edu.pl, akot@agh.edu.pl, szczepan@agh.edu.pl

Article citation info:
WANG Y, GUO L, WEN M, YANG Y. Availability analysis for a multi-component system with different *k*-out-of-*n*:G warm standby subsystems subject to suspended animation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 289–300, http://dx.doi.org/10.17531/ein.2019.2.14.

Yu WANG
Linhan GUO
Meilin WEN
Yi YANG

AVAILABILITY ANALYSIS FOR A MULTI-COMPONENT SYSTEM WITH DIFFERENT *K*-OUT-OF-*N*:G WARM STANDBY SUBSYSTEMS SUBJECT TO SUSPENDED ANIMATION

ANALIZA GOTOWOŚCI SYSTEMU WIELOELEMENTOWEGO SKŁADAJĄCEGO SIĘ Z RÓŻNYCH WYGASZALNYCH PODSYSTEMÓW TYPU *K*-Z-*N*:G STANOWIĄCYCH REZERWĘ CIEPŁĄ

Industrial equipment or systems are usually constructed as a multi-component series system with k-out-of-n:G subsystems to fulfill a specified function. As a common type of standby, warm standby is considered in the multi-component series system with k-out-of-n:G standby subsystems. When a subsystem fails, the non-failed subsystems are shut off and cannot fail, which is defined as suspended animation (SA). If the SA is ignored the non-failed subsystems are assumed to keep working in the SA time, which will cause inaccuracy in the availability analysis for the system. In this paper, we focus on the SA to construct an availability model for a multi-component series system with k-out-of-n:G warm standby subsystems. Multiple continuous time Markov chains are constructed to model the system availability. A Monte Carlo simulation has been carried out to verify our method. Several interesting findings are obtained. 1) The failure rates of subsystems with SA and their limits are derived. 2) The closed-form expressions for the stationary availability of the system and subsystems, mean time to failure, mean time to repair and stationary failure frequency are obtained considering SA. 3) The system stationary availability is a monotone function for its parameters. 4) The SA effect on the stationary availability should be emphasized in two cases, one is both the value of n/k and the failure rate of active components in a k-out-of-n subsystem are relatively large or small, the other is both the value of n/k and the repair rate are relatively small.

Keywords: *availability; multi-component series system; k-out-of-n:G warm standby subsystem; suspended animation; Markov process.*

Aby urządzenia i systemy przemysłowe mogły pełnić swoje określone funkcje, zwykle buduje się je w postaci wieloelementowych systemów szeregowych składających się z podsystemów typu k-z-n: G. W pracy rozważano zagadnienie rezerwy ciepłej w wieloelementowym systemie szeregowym składającym się z podsystemów rezerwowych typu k-z-n: G. W przypadku awarii jednego z takich podsystemów, pozostałe, działające podsystemy wyłącza się, dzięki czemu nie mogą one ulec uszkodzeniu. Procedurę taką określa się, przez analogię z organizmami żywymi mianem anabiozy (suspended animation) lub wygaszania. Pominięcie zjawiska wygaszania, prowadzi do założenia, że podsystemy, które nie uległy uszkodzeniu pracują w czasie wygaszenia, co skutkuje nieprawidłowościami w analizie gotowości systemu. W artykule koncepcję wygaszania podsystemów wykorzystano do budowy modelu gotowości wieloelementowego systemu szeregowego składającego się z podsystemów typu k-z-n: G stanowiących rezerwę ciepłą. W celu zamodelowania gotowości systemu, skonstruowano łańcuchy Markowa z czasem ciągłym. Przedstawioną metodę zweryfikowano za pomocą symulacji Monte Carlo. Uzyskano szereg interesujących wyników. 1) Obliczono intensywność uszkodzeń podsystemów wygaszonych i ich wartości graniczne. 2) Wyprowadzono, z uwzględnieniem procedury wygaszania, wyrażenia w postaci zamkniętej dla stacjonarnej gotowości rozważanego systemu i podsystemów oraz określono średni czas do uszkodzenia, średni czas do naprawy oraz stacjonarną częstotliwość uszkodzeń. 3) Gotowość stacjonarna systemu jest funkcją monotoniczną dla parametrów wejściowych systemu. 4) Należy podkreślić wpływ wygaszania na stacjonarną gotowość systemu w dwóch przypadkach: gdy stosunek n/k i intensywność uszkodzeń aktywnych elementów k-tego spośród n podsystemów są względnie niskie lub względnie wysokie, oraz gdy stosunek n/k i intensywność napraw są względnie niskie.

Słowa kluczowe: *gotowość; wieloelementowy system szeregowy; podsystem k-z-n:G; rezerwa ciepła; wygaszanie; proces Markowa.*

1. Introduction

Industrial equipment or systems are usually constructed with k-out-of-n:G subsystems in series to fulfill a specified function[14]. The k-out-of-n:G structure is a common type of redundancy used to improve the reliability and availability of engineering systems. A k-out-of-n:G system refers to a system that is functional if and only if at least k out of n components within the system are functional[31]. Two types of components exist in the k-out-of-n:G system, which are ac-

tive components and standby components[23]. A standby component switches into the active state upon an active component failure[29].

Warm standby has a general expression for the system reliability and availability. It is worthwhile to study the multi-component system with warm standby. According to the level of working load on the standby component, the standby component is classified into three types: hot, cold and warm standby[1, 6]. Hot standby implies that the standby component has the same failure rate as the active component. Cold standby implies that the standby component has a zero failure rate. The failure rate of warm standby is between cold and hot standby.

Subsystem-independence assumption can cause the inaccuracy of the system availability for the multi-component system. Some studies on a series system with k-out-of-n:G subsystems assumed that the subsystems work independently [7, 8, 11]. When a subsystem fails, the non-failed subsystems are shut off and cannot fail, which is defined as suspended animation (SA) [4, 18]. This phenomenon indicates that the subsystems are dependent. If we assume that the subsystems are independent, SA is then ignored. That is to say, the non-failed subsystems are assumed to keep working in SA time, which could result in an inaccurate estimation of the system availability.

In recent studies, some scholars have investigated the shut-off rule. The shut-off rules include SA and continuous operation (CO) [9, 12]. The SA rule specifies that no component operates when the system is down. The CO rule specifies that non-failed components continue to operate even if the system is down. The functional subsystem cannot be shut off because of the failed subsystem. The subsystem-independence assumption has no impact on the accuracy of the system availability subject to CO. However, to obtain more accurate availability, SA should be considered when we analyze the availability for a multi-component series system with different k-out-of-n:G warm standby subsystems. SA has been analyzed by some scholars in the series system and single k-out-of-n:G system.

The availability analysis for SA in a series system consisting of multiple components has been studied. Most of the studies obtained closed-form expressions for the system stationary availability. SA was first defined by Barlow and Proschan [4]. They analyzed the SA states of components in a series system and derived the system availability. The system structure has two levels. one is the system, the other is the components. Khalil [13] studied the shut-off rules of SA and CO in a series system. The availability model was constructed for the series system with exponential lifetime components. The closed-form availability was derived based on the convolution integral. Sherwin [24] discussed the calculation of the steady-state availability for a series system with SA. Pham [22] proved that the steady-state availability of a series system subject to SA is always larger than that subject to CO. Wang and Pham [26] analyzed a series system subject to SA considering the imperfect repair and the correlation of failure and repair. They assumed an arbitrary distribution of uptimes and downtimes of components and derived availability indices including system stationary availability, mean time to failure (MTTF), mean time to repair (MTTR) and stationary failure frequency (SFF).

The following studies have considered SA in single k-out-of-n:G systems with hot standby [3], cold standby [17, 25] or warm standby [27, 28] components. Moghaddass [20] et al. studied a k-out-of-n:G system with hot standby components and R repairmen. They investigated the system availability under different shut-off rules and derived closed-form expressions for the system stationary availability, MTTF, and mean time to first failure. Amiri and Ghassemi-Tari [2] performed a transient analysis for the k-out-of-n:G system subject to SA. A Markov model was constructed, and the diagonalization method was used for the transient analysis. They obtained the transient availability and MTTF of the system. Moghaddass [21] et al. analyzed the availability of a homogeneous k-out-of-n:G system with hot standby components under SA considering repair priority and finite repairmen. Moghaddass and Zuo [19] modeled the SA to analyze the availability of a k-out-of-n:G cold standby system considering repair priority. Kuo [15] et al. focused on SA to analyze the availability of a k-out-of-n:G system with warm and cold standby components. The availability model was constructed using a retrial queue at the repair facility, and the stationary availability, MTTF, and MTTR were derived. Zhang [32] et al. investigated a k-out-of-(M+N):G warm standby system with two different types of components subject to SA. Xie [30] et al. analyzed a k-out-of-n:G system jointly with hot standby redundancy and spare parts. The shut-off rule of the mixture of SA and CO was considered to analyze the system availability. An approximation of the system stationary availability is obtained.

Recently, some researches have studied the availability model of the multi-component series system with k-out-of-n: G subsystems. However, most of the models failed to consider subsystems dependence due to SA [7, 8, 11]. There are two articles most related to our work considering SA in such a system. One is that Cekyay and Ozekici [5] investigated the availability for a multi-component series system with k-out-of-n: G subsystems with exponential lifetime components considering SA. Only a continuous time Markov chain (CTMC) describing the system available states at the component level was constructed and the system stationary availability was obtained. The other is that Huffman [10] studied a multi-component series system with k-out-of-n: G hot standby subsystems considering SA. The repair begins if a failed component occurs and the repair makes the failed component brand new. They calculated the mean up-time and down-time of the subsystems based on the result derived by Li [18] et al. and substituted them in the equation proposed by Barlow and Proschan [4]:

$$A_S = \left[1 + \sum_{i=1}^{n} \frac{\theta_i}{\phi_i}\right]^{-1}. \tag{1}$$

where θ and ϕ_i are the failure rate and repair rate of components respectively.

Our work is different from the works mentioned above. Most existing studies considered SA to investigate the single k-out-of-n:G standby system or the series system consisting multiple components. We focus on the multi-component series system with different k-out-of-n:G warm standby subsystems considering SA. Although two studies [5, 10] are closely related to ours, the assumption of Eq. (1) was not satisfied in the Huffman's model[10]. In the Huffman's model, the repair time of a subsystem can be overlapped with that of the other subsystems. The assumption in Eq. (1) is that the repair time of each subsystem is independent of each other. The CTMC constructed by Cekyay and Ozekici[5] has the problem of state space explosion if the number of component type is large. Moreover, both studies failed to obtain closed-form solutions. In our paper, we model the dependence among the repair times of the multi-component k-out-of-n:G warm standby subsystem to analyze the system availability. We use multiple CTMCs to derive the system availability at the subsystem level to avoid the state space explosion. Moreover, the closed-form solution of the system availability is obtained.

The contributions of this paper can be summarized as follows.

1) We consider the SA in a multi-component series system with different k-out-of-n:G warm standby subsystems and use multiple CTMCs to model the system.
2) We effectively avoid the state explosion by constructing the CTMC model at the subsystem level.
3) We derive the closed-form expressions for the stationary availability of the system and subsystems, MTTF, MTTR, SFF based on the proposed CTMC model.
4) We discuss the property of the stationary availability function for k, n, failure rate, and repair rate.

The remainder of the paper is organized as follows. In section 2, the problem is described in detail, and the assumptions and notations are provided. In section 3, the mathematical model is given for the subsystem and system transition process. Then, the closed-form expressions of the stationary availability of the system and subsystems, SFF, MTTF, and MTTR are derived from the model. We also discuss the monotonicity of the system stationary availability function. In section 4, we carry out three numerical examples. The first example is a Monte Carlo simulation to verify our model. The second example is a comparison between the method with subsystem-independence assumption and the proposed method. The third example is a sensitivity analysis for the difference between the two methods in term of system stationary availability. Finally, conclusions and future research are presented in section 5.

2. Problem description

2.1. Notations

SA	Suspended animation
CO	Continuous operation
MTTF	Mean time to failure
MTTR	Mean time to repair
FF	Failure frequency
SFF	Stationary failure frequency
CTMC	Continuous-time Markov chain
m	Number of subsystems
i , j	Subsystem index
n_i	Number of components in subsystem i
k_i	Number of active components in subsystem i
$S(t), t \geq 0$	Transition process of the system, $S(t)$ indicates the system is functional or which subsystem fails at time t
$(X_i(t), I_i(t)), t \geq 0$	Transition process of subsystem i, $X_i(t)$ denotes the number of failed components in subsystem i at time t, $I_i(t) = 1, 0, -1$ denotes that subsystem i is in the operation, SA or repair state, respectively
$Y_i(T), T \geq 0$	CTMC of subsystem i without considering SA time, $Y_i(T)$ denotes the number of failed components in subsystem i at time T
λ_i	Failure rate of active components in subsystem i
λ_i^-	Failure rate of warm standby components in subsystem i
μ_i	Repair rate of subsystem i
r_i	Number of failed components in subsystem i
$\alpha_i(t)$	Failure rate of subsystem i considering SA at time t
α_i	Limit of $\alpha_i(t)$
$\mathrm{CWT}(t)$	Cumulative working time by time t
$\mathrm{CRT}_i(t)$	Cumulative repair time of subsystem i by time t
A_S	System stationary availability
NSA	No suspended animation (due to the subsystem-independence assumption)
A_S^{NSA}	System stationary availability based on NSA method
A_i^{NSA}	Stationary availability of subsystem i based on NSA method
A_S^{SA}	System stationary availability based on SA method
A_i^{SA}	Stationary availability of subsystem i based on SA method
Δ_i	Absolute difference between A_i^{NSA} and A_i^{SA}
Δ	Difference between A_S^{NSA} and A_S^{SA}

2.2. System description

We consider a multi-component series system consisting of m different k-out-of-n:G warm standby subsystems, as shown in Fig. 1. Subsystem $i\,(i \in \{1, \cdots, m\})$ has n_i identical and independent components. There are k_i active components and $n_i - k_i$ warm standby components in subsystem i. Subsystem i fails when less than k_i out of n_i components are functional. The system is functional only if all subsystems are functional. When a subsystem fails, the other subsystems are in the SA state. In the SA state, the non-failed subsystems cannot operate or fail. We assume that the repair of subsystem i begins upon the number of failed components being greater than $(n_i - k_i)$ in the subsystem. When the repair is complete, the system restarts to operate.

We define a transition process $S(t), t \geq 0$ to describe the system states with SA. Let $S(t) = 0$ denote that the system is operating at time t, and $S(t) = i\,(i \in \{1, 2, \cdots, m\})$ denote that the system is down at time t due to the failure of subsystem i. The operating state can transit to one of the failure states, vice versa. A failure state cannot transit to the other failure states since no failure occurs when the system is down.

The aim of this work is to consider SA for the availability analysis on a multi-component system with different k-out-of-n:G warm standby subsystems. The system availability is the probability that $S(t) = 0$. The state probability can be solved if we have the transition rates between state $S(t) = 0$ and $S(t) = i$. The transition rate from $S(t) = i$ to $S(t) = 0$ is the repair rate of subsystem i. In this paper, we assume the repair rate of subsystem i is a constant μ_i. However, the transition rate from state $S(t) = 0$ to $S(t) = i$ need to derive from the component failure rate in subsystem i. We define the transition rate from state $S(t) = 0$ to $S(t) = i$ as the failure rate of subsystem i, $\alpha_i(t)$. The failure rate $\alpha_i(t)$ is time-varying since the reliability of subsystem i depends on the time-varying number of functional components in subsystem i. Then, we present a binary-dimension state transition process derived from $S(t)$ to deduce the formula of $\alpha_i(t)$. Finally, we can solve the state probability based on $S(t)$ after obtaining $\alpha_i(t)$.

2.3. Assumptions

(1) The system fails when anyone of the subsystems fails. Subsystem i fails when less than k_i out of n_i components are functional.
(2) The lifetimes of active components and standby components in subsystem i are independent and exponentially distributed with the parameters λ_i and $\lambda_i^- \left(\lambda_i^- \le \lambda_i\right)$, respectively. The failure of active or standby components occurs only when the components operate.
(3) Upon a failure of an active component, a standby component instantly switches into the active state with 1 probability, if the standby component is available.
(4) The repair of a subsystem will not start until the number of failed components in the subsystem reaches $n_i - k_i + 1$ and the repair makes all failed components in the subsystem brand new. The repair time of subsystem i is exponentially distributed with the μ_i parameter.
(5) The occurrence of more than one failed subsystem is an impossible event.

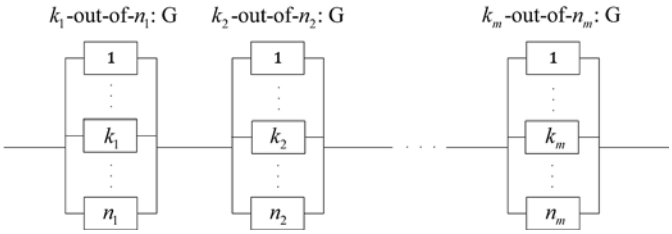


Fig. 1. Configuration of the multi-component series system consisting of m different k-out-of-n:G warm standby subsystems

3. System availability modeling and solution

3.1. Subsystem transition process and its failure rate

We define another process $\left(X_i(t), I_i(t)\right)$ derived from $S(t)$ to obtain the subsystem failure rate $\alpha_i(t)$. As the failure rate $\alpha_i(t)$ depends on the states of the components in subsystem i we define the binary-dimension state $\left(X_i(t), I_i(t)\right)$ to describe the operation, repair, SA states of components in subsystem i, where $X_i(t)$ denotes the number of failed components in subsystem i at time t and $I_i(t) = 1, 0, -1$ denotes the subsystem being in the operation, SA, or repair state, respectively. The stochastic process of $\left(X_i(t), I_i(t)\right)$ only considers the states of the components in subsystem i without the states of the components in other subsystems. The effect of SA among the subsystems can be described by $I_i(t)$. The relationship of $I_i(t)$ and $S(t)$ is as follows:

$$I_i(t) = \begin{cases} 1, & \text{if } S(t) = 0 \\ 0, & \text{if } S(t) = j, \ j = 1, 2, \cdots, i-1, i+1, \cdots, m \\ -1, & \text{if } S(t) = i \end{cases}. \quad (2)$$

As an example, we plot a sample path of $S(t)$ of a series system with 2 k-out-of-n:G subsystems and a corresponding path of $I_i(t)$.

We depict more detail of the transition process $\left(X_i(t), I_i(t)\right)$ as follows. Denote r_i as the number of failed components in subsystem i. Due to a failure of one component in subsystem i, the state transits from $(r_i, 1)$ to $(r_i + 1, 1)$ with rate $k_i\lambda_i + r_i\lambda_i^-$, for $r_i = 0, 1, \cdots, n_i - k_i - 1$, or transits from $(n_i - k_i, 1)$ to $(n_i - k_i + 1, -1)$ with rate $k_i\lambda_i$. The repair of the subsystem makes the state transit from $(n_i - k_i + 1, -1)$ to $(0, 1)$ with rate μ_i. The other subsystems failure leads to the transition from state $(r_i, 1)$ to SA state $(r_i, 0)$, for $r_i = 1, 2, \cdots, n_i - k_i$, and the repair of the failed subsystem leads to the inverse transition. The transition diagram of the process $\left(X_i(t), I_i(t)\right)$ is shown in Fig. 3.

We present two lemmas to derive the expression of $\alpha_i(t)$ and its limit based on the process $\left(X_i(t), I_i(t)\right)$.

Lemma 1 The failure rate of subsystem i can be calculated by

$$\alpha_i(t) = k_i\lambda_i \cdot \frac{\mathrm{P}\left(X_i(t) = n_i - k_i, I_i(t) = 1\right)}{\sum_{r_i=0}^{n_i-k_i} \mathrm{P}\left(X_i(t) = r_i, I_i(t) = 1\right)}. \quad (3)$$

Proof

At state $(n_i - k_i, 1)$, all standby components in subsystem i are failed, and one more failure of the active components in subsystem i results in the failure of subsystem i. Therefore, the subsystem failure

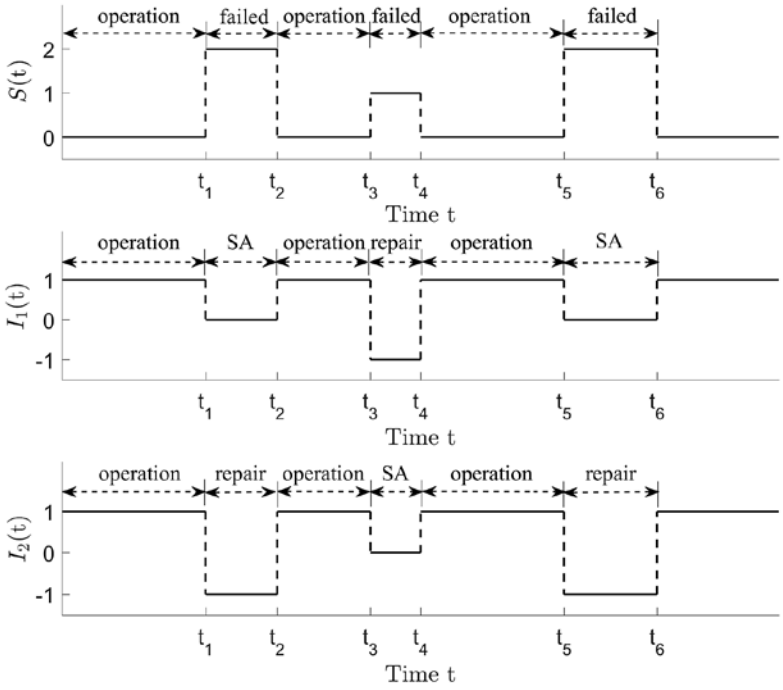


Fig. 2. Sample path of $S(t)$ and corresponding $I_i(t)$. The operation state of $S(t)$ corresponds to the operation state of $I_i(t)$. If $S(t) = i$ the failed state of $S(t)$ corresponds to the repair state of $I_i(t)$; if $S(t) \neq i$ the failed state of $S(t)$ corresponds to the SA state of $I_i(t)$

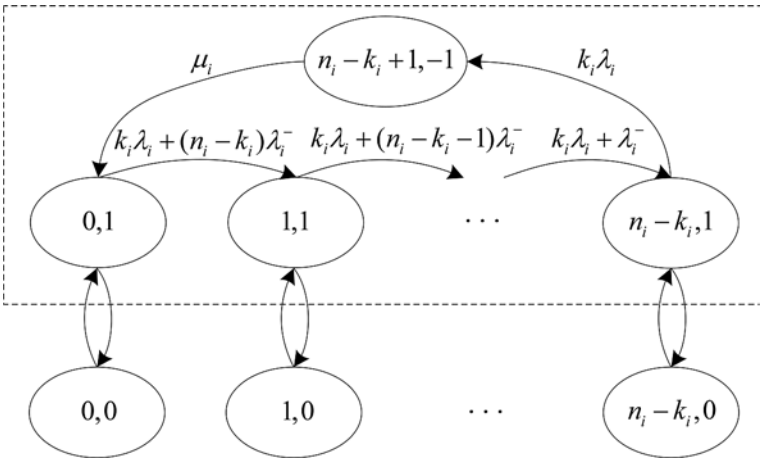


Fig. 3. Transition diagram of the process $(X_i(t),I_i(t))$*, and* $Y_i(T)$ *fenced in the dashed rectangle.*

rate $\alpha_i(t)$ is the transition rate from the state $(n_i-k_i,1)$ to the state $(n_i-k_i+1,-1)$ on condition that $(X_i(t),I_i(t))=(n_i-k_i,1)$ when the subsystem is operating at time t. The details are provided in the following deduction:

$$\begin{aligned}\alpha_i(t) &= \lim_{\Delta t\to 0}\frac{\mathrm{P}(S(t+\Delta t)=i\mid S(t)=0)}{\Delta t}\\ &= \lim_{\Delta t\to 0}\frac{\mathrm{P}(X_i(t+\Delta t)=n_i-k_i+1, I_i(t+\Delta t)=-1\mid X_i(t)<n_i-k_i+1, I_i(t)=1)}{\Delta t}\\ &= \lim_{\Delta t\to 0}\frac{\mathrm{P}(X_i(t+\Delta t)=n_i-k_i+1, I_i(t+\Delta t)=-1\mid X_i(t)=n_i-k_i, I_i(t)=1)}{\Delta t}\\ &\quad\cdot\mathrm{P}(X_i(t)=n_i-k_i, I_i(t)=1\mid X_i(t)<n_i-k_i+1, I_i(t)=1)\\ &= k_i\lambda_i\cdot\mathrm{P}(X_i(t)=n_i-k_i, I_i(t)=1\mid X_i(t)<n_i-k_i+1, I_i(t)=1)\\ &= k_i\lambda_i\cdot\mathrm{P}(X_i(t)=n_i-k_i, I_i(t)=1)\Big/\sum_{r_i=0}^{n_i-k_i}\mathrm{P}(X_i(t)=r_i, I_i(t)=1)\end{aligned} \tag{4}$$

Lemma 1 is proved.

To analyze the system stationary availability, we study the limit of $\alpha_i(t)$ by analyzing the limiting behavior of $(X_i(t),I_i(t))$. Note that Eq. (4) has no terms about the SA states except time t influenced by SA. We first do not consider the SA time spent in state $(r_i,0)$. In $(X_i(t),I_i(t))$, the operating time spent in state $(r_i,1)$ before transition to state $(r_i+1,1)$ or $(n_i-k_i+1,-1)$ follows an exponential distribution with parameter $k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-$. The repair time of the state $(n_i-k_i+1,-1)$ follows an exponential distribution with parameter μ_i. Thus, the operating and repair times can be analyzed in a new CTMC. The transition diagram of the new CTMC is shown in the dotted rectangle in Fig. 3. We denote the new CTMC as $Y_i(T),T\ge 0$ to describe the transition process among the states $(r_i,1)$, for all $r_i=1,2,\cdots,n_i-k_i$, and state $(n_i-k_i+1,-1)$.

As we omit the time spent in the SA states $(r_i,0)$, we need to analyze the relationship of the time spent in $Y_i(T)$ and in $(X_i(t),I_i(t))$, that is:

$$T=T_i(t)=\int_0^t|I_i(u)|\,\mathrm{d}u\,. \tag{5}$$

On the other hand, since the state space of $Y_i(T)$ does not include the SA states corresponding to $I_i(t)=0$, we have the conditional probability equation:

$$\mathrm{P}(Y_i(T_i(t))=r_i)=\mathrm{P}(X_i(t)=r_i, I(t)=1\mid I_i(t)\ne 0). \tag{6}$$

Multiplying $\mathrm{P}(I_i(t)\ne 0)$ at the two sides of Eq. (6), we have:

$$\mathrm{P}(X_i(t)=r_i, I_i(t)=1)=\mathrm{P}(Y_i(T_i(t))=r_i)\mathrm{P}(I_i(t)\ne 0). \tag{7}$$

Substituting Eq. (7) in Eq. (3) and dividing the top and bottom by $\mathrm{P}(I_i(t)\ne 0)$, we have:

$$\alpha_i(t)=k_i\lambda_i\cdot\frac{\mathrm{P}(Y_i(T_i(t))=n_i-k_i)}{\sum_{r_i=0}^{n_i-k_i}\mathrm{P}(Y_i(T_i(t))=r_i)}. \tag{8}$$

Based on the CTMC $Y_i(t)$ and Eq. (8), we propose Lemma 2 for obtaining the formula of the limit of $\alpha_i(t)$ when time t tends to infinity.

Lemma 2 The limit of $\alpha_i(t)$ is a constant, computed as follows:

$$\alpha_i=\lim_{t\to\infty}\alpha_i(t)=\left(\sum_{r_i=0}^{n_i-k_i}\frac{1}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1}. \tag{9}$$

Proof

According to Eq. (8), the limit of $\alpha_i(t)$ can be derived if the limit of $\mathrm{P}(Y_i(T_i(t))=r_i)$ is determined.

When the time t tends to infinity, we have:

$$\begin{aligned}\lim_{t\to\infty}T_i(t) &= \lim_{t\to\infty}t\cdot\frac{\mathrm{CWT}(t)+\mathrm{CRT}_i(t)}{t}\\ &> \lim_{t\to\infty}t\cdot\frac{\mathrm{CWT}(t)}{t} > A_S\lim_{t\to\infty}t=\infty\end{aligned}, \tag{10}$$

where $\mathrm{CWT}(t)$ and $\mathrm{CRT}_i(t)$ respectively denote the cumulative working time and the cumulative repair time of subsystem i by time t, and A_S is the system stationary availability. Then, we have for $r_i=0,1,\cdots,n_i-k_i$:

$$\lim_{t\to\infty}\mathrm{P}(Y_i(T_i(t))=r_i)=\lim_{T\to\infty}\mathrm{P}(Y_i(T)=r_i), \tag{11}$$

which can be calculated using the Chapman-Kolmogorov equation of $Y_i(T)$:

$$\begin{cases}\frac{\mathrm{d}}{\mathrm{d}T}\mathrm{P}(Y_i(T)=0)=-\left(k_i\lambda_i+(n_i-k_i)\lambda_i^-\right)\mathrm{P}(Y_i(T)=0)+\mu_i\mathrm{P}(Y_i(T)=n_i-k_i+1)\\ \frac{\mathrm{d}}{\mathrm{d}T}\mathrm{P}(Y_i(T)=r_i)=\left(k_i\lambda_i+(n_i-k_i-r_i+1)\lambda_i^-\right)\mathrm{P}(Y_i(T)=r_i-1)-\left(k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-\right)\mathrm{P}(Y_i(T)=r_i),\\ \qquad\text{for } r_i=1,2,\cdots,n_i-k_i\\ \frac{\mathrm{d}}{\mathrm{d}T}\mathrm{P}(Y_i(T)=n_i-k_i+1)=k_i\lambda_i\mathrm{P}(Y_i(T)=n_i-k_i)-\mu_i\mathrm{P}(Y_i(T)=n_i-k_i+1)\end{cases} \tag{12}$$

As $Y_i(T)$ is an irreducible and aperiodic CTMC with finite state space, the stationary state probability is regardless of the initial distribution. Solving Eq. (12), we obtain that:

$$\lim_{T\to\infty} \mathrm{P}(\mathrm{Y}_i(T)=r_i)=\left(\frac{1}{\mu_i}+\sum_{r_i=0}^{n_i-k_i}\frac{1}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1} \cdot \frac{1}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}, \quad (13)$$

for $r_i=1,2,\cdots,n_i-k_i$. Then, we obtain the limit of $\alpha_i(t)$ based on Eq. (8,13):

$$\begin{aligned}\alpha_i &= \lim_{t\to\infty}\alpha_i(t)\\ &= \lim_{t\to\infty} k_i\lambda_i\cdot\frac{\mathrm{P}\left(Y_i(T_i(t))=n_i-k_i\right)}{\sum_{r_i=0}^{n_i-k_i}\mathrm{P}\left(Y_i(T_i(t))=r_i\right)}\\ &= \lim_{T\to\infty} k_i\lambda_i\cdot\frac{\mathrm{P}\left(Y_i(T)=n_i-k_i\right)}{\sum_{r_i=0}^{n_i-k_i}\mathrm{P}\left(Y_i(T)=r_i\right)} \quad . \quad (14)\\ &= \left(\sum_{r_i=0}^{n_i-k_i}\frac{1}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1}\end{aligned}$$

Lemma 2 is proved.

Lemma 2 indicates that, after a long time, the transition rates from $S(t)=0$ to $S(t)=i$ are constant.

3.2. Stationary availability and other characteristics

In this subsection, we present the closed-form expressions for the stationary availability of the system and subsystems, SFF, MTTF, and MTTR.

The system availability is the probability that $S(t)=0$. The availability of subsystem i is the probability that $S(t)\neq i$. After a long time, the behavior of the stochastic process $\mathrm{S}(t)$ can be described using a CTMC where the system state $S(t)$ transits from 0 to i with the transition rate α_i and transits from i to 0 with the transition rate μ_i. By solving the limiting state probability, we derive the stationary availability of the system and subsystems.

Theorem the stationary availability of the system is:

$$A_S=\left[1+\sum_{i=1}^{m}\left(\sum_{r_i=0}^{n_i-k_i}\frac{\mu_i}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1}\right]^{-1}. \quad (15)$$

And the stationary availability of subsystem i is:

$$A_i=1-\left(\sum_{r_i=0}^{n_i-k_i}\frac{\mu_i}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1} \cdot\left[1+\sum_{j=1}^{m}\left(\sum_{r_j=0}^{n_j-k_j}\frac{\mu_j}{k_j\lambda_j+(n_j-k_j-r_j)\lambda_j^-}\right)^{-1}\right]^{-1}. \quad (16)$$

Proof

The stationary availability of the system is the probability that $S(t)=0$ as time t tends to infinity:

$$A_S=\lim_{t\to\infty}\mathrm{P}\left(S(t)=0\right). \quad (17)$$

As to the stationary availability of subsystem i, the suspended time of subsystem i is the available time of subsystem i because subsystem i is functional in the SA state. Then, we have the stationary availability of subsystem i as:

$$A_i=1-\lim_{t\to\infty}\mathrm{P}\left(S(t)=i\right). \quad (18)$$

The Chapman-Kolmogorov equation of the CTMC is as follows:

$$\begin{cases}\dfrac{\mathrm{d}}{\mathrm{d}t}\mathrm{P}\left(S(t)=0\right)=-\sum_{i=1}^{m}\alpha_i\mathrm{P}\left(S(t)=0\right)+\sum_{i=1}^{m}\mu_i\mathrm{P}\left(S(t)=i\right)\\ \dfrac{\mathrm{d}}{\mathrm{d}t}\mathrm{P}\left(S(t)=i\right)=\alpha_i\mathrm{P}\left(S(t)=0\right)-\mu_i\mathrm{P}\left(S(t)=i\right)\\ \quad \text{for } i=1,\cdots,m\end{cases} \quad . \quad (19)$$

As the limit behavior of $S(t)$ is an irreducible and aperiodic CTMC with finite state space, the stationary state probability is regardless of the initial distribution. Solving Eq. (19), we have:

$$\begin{cases}\lim_{t\to\infty}\mathrm{P}\left(S(t)=0\right)=\left(1+\sum_{i=1}^{m}\frac{\alpha_i}{\mu_i}\right)^{-1}\\ \quad=\left[1+\sum_{i=1}^{m}\left(\sum_{r_i=0}^{n_i-k_i}\frac{\mu_i}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1}\right]^{-1}\\ \lim_{t\to\infty}\mathrm{P}\left(S(t)=i\right)=\frac{\alpha_i}{\mu_i}\cdot\left(1+\sum_{j=1}^{m}\frac{\alpha_j}{\mu_j}\right)^{-1} \quad \text{for } i=1,\cdots,m\end{cases} \quad . \quad (20)$$

Substituting Eq. (20) in Eq.(17,18), we have the closed-form expression of A_S and A_i. Then the theorem is proved.

The other characteristics including SFF, MTTF, and MTTR can also be derived based on the limiting behavior of $S(t)$. The failure frequency of the system is a total of the failure rates of the subsystems on condition that the system is available. MTTF and MTTR are also related to the failure rates of the subsystems. Based on the two Lemma and theorem, we propose the following corollary for the formula of SFF, MTTF, and MTTR.

Corollary 1 Based on the limiting behavior of $S(t)$, we have the formula of SFF, MTTF, and MTTR as follows:

$$\mathrm{SFF}=\left[1+\sum_{i=1}^{m}\left(\sum_{r_i=0}^{n_i-k_i}\frac{\mu_i}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1}\right]^{-1} \cdot\sum_{i=1}^{m}\left(\sum_{r_i=0}^{n_i-k_i}\frac{1}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1}, \quad (21)$$

$$\mathrm{MTTF}=\left[\sum_{i=1}^{m}\left(\sum_{r_i=0}^{n_i-k_i}\frac{1}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1}\right]^{-1}, \tag{22}$$

$$\mathrm{MTTR}=\left[\sum_{i=1}^{m}\left(\sum_{r_i=0}^{n_i-k_i}\frac{1}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1}\right]^{-1}\cdot \sum_{i=1}^{m}\left(\sum_{r_i=0}^{n_i-k_i}\frac{\mu_i}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1}. \tag{23}$$

Proof

According to Lemma 1 and Lemma 2, α_i is defined as the failure rate of subsystem i on condition that the system is operating. SFF is the total failure rate of the system when the system is operating. Thus, SFF can be calculated as the sum of α_i multiplying the probability that $S(t)=0$ as time t to infinity:

$$\mathrm{SFF}=\sum_{i=1}^{m}\alpha_i\cdot A_S. \tag{24}$$

According to the concept of MTTF and MTTR, we have:

$$\mathrm{MTTF}=\frac{A_S}{\mathrm{SFF}}, \tag{25}$$

and:

$$\mathrm{MTTR}=\frac{1-A_S}{\mathrm{SFF}}. \tag{26}$$

Based on Lemma 2, Theorem and Eq.(23-25), we have the formula of SFF, MTTF, and MTTR. Corollary 1 is proved.

A multi-component series system with m components is a special case of the considered system. we can set $n_i=k_i=1$, and according to Eq. (15, 21-23), we have:

$$\begin{cases} A_S=\left(1+\sum_{i=1}^{m}\frac{\lambda_i}{\mu_i}\right)^{-1} \\ \mathrm{SFF}=\left(1+\sum_{i=1}^{m}\frac{\lambda_i}{\mu_i}\right)^{-1}\cdot\sum_{i=1}^{m}\lambda_i \\ \mathrm{MTTF}=\left(\sum_{i=1}^{m}\lambda_i\right)^{-1} \\ \mathrm{MTTR}=\left(\sum_{i=1}^{m}\lambda_i\right)^{-1}\cdot\sum_{i=1}^{m}\frac{\lambda_i}{\mu_i} \end{cases}, \tag{27}$$

The above results coincide with the works of Kuo and Zuo in which a CTMC at the component level is constructed to solve these characteristics [16].

3.3. Analysis of the system stationary availability function

To apply the proposed method in the product development, we need to analyze the effect of the input parameters, including the failure rate and repair rate of the components and redundancy level, on the system availability so that we can improve the reliability of the product. Then, we discuss the property of the system stationary availability function.

Corollary 2 The system stationary availability is a monotone decreasing function of k_i

Proof

For convenience, we denote $\alpha_i(k_i)$ as the function α_i of k_i. Then, we have:

$$\alpha_i(k_i)=\left(\sum_{r_i=0}^{n_i-k_i}\frac{1}{k_i\lambda_i+(n_i-k_i-r_i)\lambda_i^-}\right)^{-1}=\left(\sum_{r_i=0}^{n_i-k_i}\frac{1}{k_i\lambda_i+r_i\lambda_i^-}\right)^{-1}, \tag{28}$$

$$\alpha_i(k_i+1)=\left(\sum_{r_i=0}^{n_i-k_i-1}\frac{1}{(k_i+1)\lambda_i+r_i\lambda_i^-}\right)^{-1}. \tag{29}$$

Then, we have:

$$\frac{1}{\alpha_i(k_i)}-\frac{1}{\alpha_i(k_i+1)}=\sum_{r_i=0}^{n_i-k_i-1}\left(\frac{1}{k_i\lambda_i+r_i\lambda_i^-}-\frac{1}{(k_i+1)\lambda_i+r_i\lambda_i^-}\right)+\frac{1}{k_i\lambda_i+(n_i-k_i)\lambda_i^-}>0 \tag{30}$$

Therefore, $\alpha_i(k_i)$ is a monotone increasing function of k_i. Furthermore, A_S is a monotone decreasing function of α_i according to Eq. (19). Thus, A_S is a monotone decreasing function of k_i. Corollary 2 is proved.

We can similarly derive A_S is a monotone increasing function for n_i.

Corollary 3 The system stationary availability is a monotone decreasing function of λ_i.

Proof

The first order difference of A_S for λ_i is:

$$\frac{\partial A_S}{\partial\lambda_i}=-A_S^2\cdot\left(\sum_{r_i=0}^{n_i-k_i}\frac{\mu_i}{k_i+(n_i-k_i-r_i)\theta_i}\right)^{-1}<0, \tag{31}$$

where $\theta_i=\lambda_i^-/\lambda_i$. Corollary 3 is proved.

For the other parameters λ_i^- and μ_i, the monotone of A_S can be similarly derived by the partial difference.

4. Numerical examples

4.1. Inputs of the numerical examples

In this part, we present some examples to illustrate and better understand the proposed method. A multi-component series system with 10 k-out-of-n:G warm standby subsystems is considered as the object in the following examples. The input parameters including n_i, k_i, λ_i, λ_i^- / λ_i, and μ_i in each subsystem are shown in Table 1. We use the same inputs for verification using a Monte Carlo simulation developed by MATLAB 2016a. We compare the methods assuming subsystem independence and subsystem dependence due to SA to find the difference between the two methods. A sensitivity analysis is performed with respect to the input parameters, including n_i, k_i, λ_i, and μ_i to provide technical insight for the reliability engineer. We coded the numerical algorithm in MATLAB 2016a. The programs were run in a PC with a 2.50 GHz processor and 4 Gb of RAM.

Table 1. Input parameters of the system

Subsystem i	n_i	k_i	λ_i (per hour)	λ_i^- / λ_i	μ_i (per hour)
1	5	2	1/1000	0.2	1/30
2	4	1	1/900	0.4	1/30
3	7	3	1/800	0.3	1/25
4	2	1	1/2000	0.8	1/60
5	3	1	1/750	0.7	1/50
6	4	2	1/500	0.6	1/25
7	6	2	1/1000	0.4	1/20
8	3	1	1/600	0.1	1/35
9	8	3	1/700	0.2	1/30
10	6	1	1/500	0.1	1/20

4.2. Model verification by a Monte Carlo simulation

In this subsection, we calculate A_S, SFF, MTTF, and MTTR based on the analytical method proposed in section 3. A Monte Carlo simulation is carried out to verify our method. The number of replications in the simulation is 10^4, and the time period is 10^5 hours. The time unit is the hour. In the simulation, we record whether the system is working or under repair in each hour, so that we can carry out the instantaneous analysis. A_S is calculated by the average ratio of the cumulated working times to the time period; The instantaneous availability is calculated by the average ratio of the number of working systems to the number of replications at time t in the simulation. Failure frequency(FF) is calculated by the average ratio of the failure numbers to the time period; MTTF (MTTR) is calculated by the average ratio of the cumulated working (repair) time to the failure number in the time period.

We choose 5 to 10 subsystems as the object systems and make the comparison between the analytical model and simulation to verify the analytical model proposed in this paper. The results of the analytical model and simulation by $t = 10^5$ hours are shown in Table 2. All the relative errors between analytical model and simulation at or below 10^{-3} level. For the 7th combination, we plot the instantaneous availability, FF output by simulation in Fig. 4. The other combinations are similar to the 7th combination.

In fig.4 (a), the simulation curve drops quickly at first and then tends to be stable. The simulation curve covers the analytical result $A_S = 0.8259$ from time $t = 1.2 \times 10^3$ hours. We compute the mean of the instantaneous availability output by the simulation from $t = 1.2 \times 10^3$ hours to $t = 10^5$ hours, and the result is 0.8260. The relative error between the mean and analytical result of A_S is 0.012%. In fig. 4(b), the simulation curve changes significantly from $t = 0$ to $t = 2 \times 10^4$ hours, which indicates that the system tends to be stable from the initial state. The relative error between FF and analytical result of SFF is 0.5440% at $t = 10^5$ hours. Thus, it can be concluded that our method is accurate.

4.3. Comparison of the method assuming subsystem independence and the proposed method

In this subsection, we compare the methods assuming subsystem independence and subsystem dependence due to SA to find the error of the method assuming subsystem-independence. We calculate the stationary availability of the system and subsystem i under the condition of no suspended animation (NSA) and SA. We denote the stationary availabilities of the system and subsystem i based on the NSA method as A_S^{NSA} and A_i^{NSA},

Table 2. The results of the analytical model and simulation

Combinations		A_S	SFF (or FF for Simulation)	MTTF (hours)	MTTR (hours)
(1,2,3,4,5)	Analytical	0.9058	0.002583	350.7	36.49
	Simulation	0.9062	0.002578	351.9	36.38
	Relative error	0.0442%	0.1936%	0.3422%	0.3015%
(6,7,8,9,10)	Analytical	0.9036	0.003659	247.0	26.35
	Simulation	0.9040	0.003652	247.7	26.29
	Relative error	0.0443%	0.1913%	0.2834%	0.2277%
(1,3,5,7,9)	Analytical	0.9025	0.003144	287.0	31.00
	Simulation	0.9031	0.003136	288.2	30.91
	Relative error	0.0665%	0.2545%	0.4181%	0.2903%
(2,4,6,8,10)	Analytical	0.9068	0.003098	292.7	30.07
	Simulation	0.9073	0.003091	293.8	30.00
	Relative error	0.0551%	0.2260%	0.3758%	0.2328%
(1,2,3,5,7,10)	Analytical	0.9061	0.003149	287.8	29.83
	Simulation	0.9066	0.003138	289.1	29.75
	Relative error	0.0552%	0.3493%	0.4517%	0.2682%
(2,3,4,7,8,9,10)	Analytical	0.8939	0.003571	250.3	29.70
	Simulation	0.8946	0.003559	251.6	29.63
	Relative error	0.0783%	0.3360%	0.5194%	0.2357%
(1,2,3,4,5,6,7,8,9,10)	Analytical	0.8259	0.005699	144.9	30.54
	Simulation	0.8269	0.005668	145.7	30.50
	Relative error	0.1211%	0.5440%	0.5521%	0.1310%

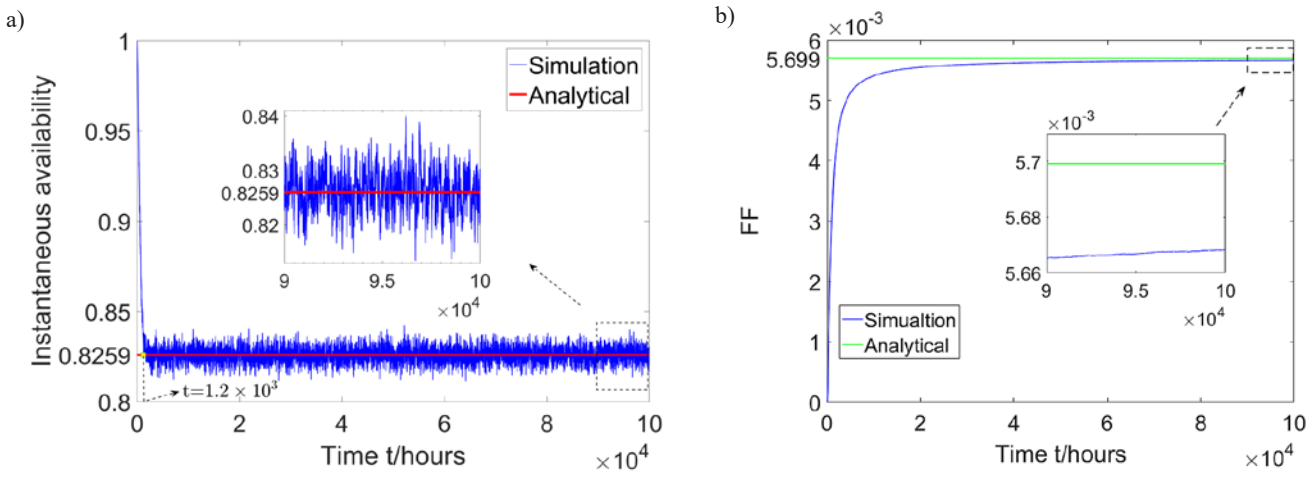


Fig. 4. Comparison between the analytical method and simulation in terms of the output. In Fig. 4(a), the simulation curve covers the output of the analytical method from $t = 1.2 \times 10^3$ hours to $t = 10^5$ hours. In Fig. 4(b), the simulation curve converges to the analytical output

respectively. We denote the stationary availabilities of the system and subsystem i based on SA method as A_S^{SA} and A_i^{SA}, respectively. According to the reference [16], the formulas of A_i^{NSA} and A_S^{NSA} are expressed as:

$$A_i^{NSA} = \left(1 + \sum_{r_i=0}^{n_i-k_i} \frac{\mu_i}{k_i\lambda_i + (n_i - k_i - r_i)\lambda_i^-}\right)^{-1}, \tag{32}$$

and:

$$A_S^{NSA} = \prod_{i=1}^{m}\left(1 + \sum_{r_i=0}^{n_i-k_i} \frac{\mu_i}{k_i\lambda_i + (n_i - k_i - r_i)\lambda_i^-}\right)^{-1}. \tag{33}$$

Table 3. The results of A_i^{NSA} and A_i^{SA}

i	A_i^{NSA}	A_i^{SA}	Δ_i
1	0.985888	0.983201	0.002687
2	0.989895	0.987913	0.001982
3	0.981677	0.978297	0.003380
4	0.984071	0.981079	0.002992
5	0.972536	0.967818	0.004718
6	0.965503	0.959907	0.005596
7	0.991138	0.989385	0.001754
8	0.982432	0.979172	0.003259
9	0.979550	0.975838	0.003712
10	0.993247	0.991891	0.001357

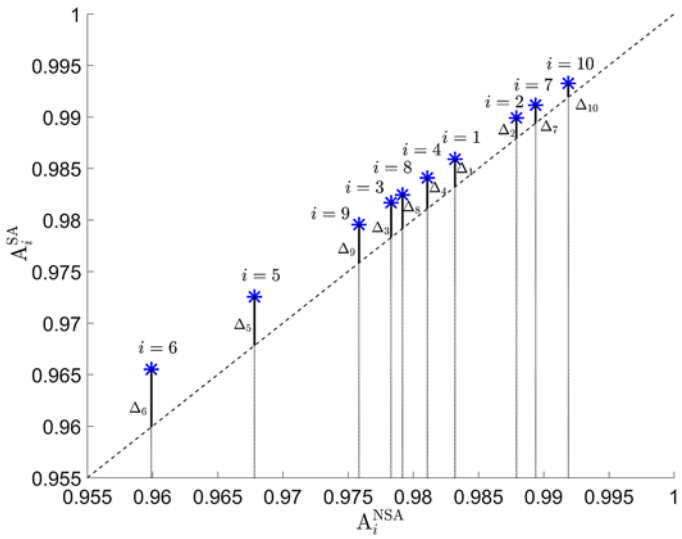


Fig. 5. Comparison of A_i^{NSA} and A_i^{SA}. The asterisks represent vectors (A_i^{NSA}, A_i^{SA}) and the length of the solid line represents Δ_i, the absolute differences between A_i^{NSA} and A_i^{SA}. The points on the dashed line indicate $A_i^{NSA} = A_i^{SA}$

We obtain $A_S^{SA} = 0.8259$ and $A_S^{NSA} = 0.8121$ based on the input parameters. A comparison of A_i^{SA} and A_i^{NSA} is shown in Fig. 5. The difference is directly apparent in Fig. 5, and the data in Fig. 5 are shown in Table 3. For $i \in 1,2,\cdots,10$, A_i^{SA} is always larger than A_i^{NSA}. If the SA is not considered, the stationary availability of the system and subsystems is always underestimated. From point $i = 10$ to $i = 6$ in Fig. 5, the difference Δ_i between A_i^{NSA} and A_i^{SA} increases as A_i^{NSA} decreases, which implies that SA states have a greater impact on the subsystem with lower availability.

In addition, that the subsystem with lower availability causes a longer SA time indicates stronger dependency being on other subsystems. This dependency can result in a larger difference associated with the availability of other subsystems.

4.4. Sensitivity analysis for the difference between A_S^{NSA} and A_S^{SA}

In this subsection, we perform a sensitivity analysis for the difference between A_S^{NSA} and A_S^{SA} with respect to different types of parameters. We choose two combinations of parameters, (n_1, k_1, λ_1) and (n_1, k_1, μ_1) to observe the variation of the

difference when changing these parameters. The other parameters are from Table 1.

(1) difference analysis for (n_1,k_1,λ_1)

We calculate A_S^{SA} and A_S^{NSA} with different values of (n_1,k_1,λ_1). λ_1 ranges from 5×10^{-3} to 1.5×10^{-1} with step 5×10^{-3}, while for (n_1,k_1), we choose 4 groups, $(5,2)$, $(8,2)$, $(5,4)$ and $(8,4)$. The results are plotted in Fig. 6. The difference between A_S^{SA} and A_S^{NSA}, $\Delta(n_1,k_1,\lambda_1)$, is plotted in Fig. 7. Each curve in Fig. 7 rises at first and then drops, which indicates that the SA effect on availability become weak when the failure rate is relatively smaller or greater. On the other hand, in Fig. 7, for low failure rate λ_1, $\Delta(n_1,k_1,\lambda_1)$ increases in the order of (8, 2), (5, 2), (8, 4), (5, 4), while $\Delta(n_1,k_1,\lambda_1)$ increases in the inverse order for high failure rate. The order of the (n_1,k_1) sequence is arranged in the descending order of the value of n_1/k_1. That is to say that, for the system with a high value of n/k in the subsystem, SA should be emphasized when failure rate is great. While for the system with a low value of that, SA should be emphasized when failure rate is small.

(2) Difference analysis for (n_1,k_1,μ_i)

We calculate A_S^{SA} and A_S^{NSA} for different value of (n_1,k_1,μ_1). μ_1 ranges from 3.33×10^{-2} to 6.67×10^{-1} with step 3.33×10^{-2} and (n_1,k_1) is chosen as $(5,2)$, $(8,2)$, $(5,4)$ and $(8,4)$. The results of A_S^{SA} and A_S^{NSA} are plotted in Fig. 8 and the difference between A_S^{SA} and A_S^{NSA}, $\Delta(n_1,k_1,\mu_1)$, is plotted in Fig. 9. For each μ_1 in this range, $\Delta(n_1,k_1,\mu_1)$ increases with the value of n_1/k_1 increasing. Therefore, the SA effect is relatively strong for the system with the relatively

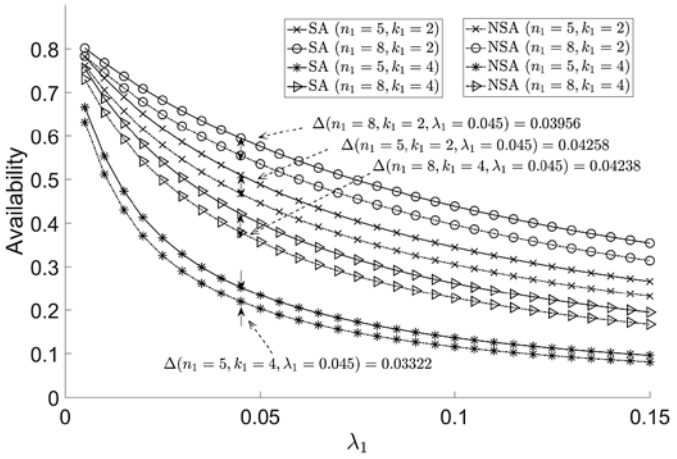


Fig. 6. A_S^{SA} and A_S^{NSA} for (n_1,k_1,λ_1), where λ_1 ranges from 5×10^{-3} to 1.5×10^{-1} with step 5×10^{-3} and (n_1,k_1) is set as $(5,2)$, $(8,2)$, $(5,4)$ and $(8,4)$

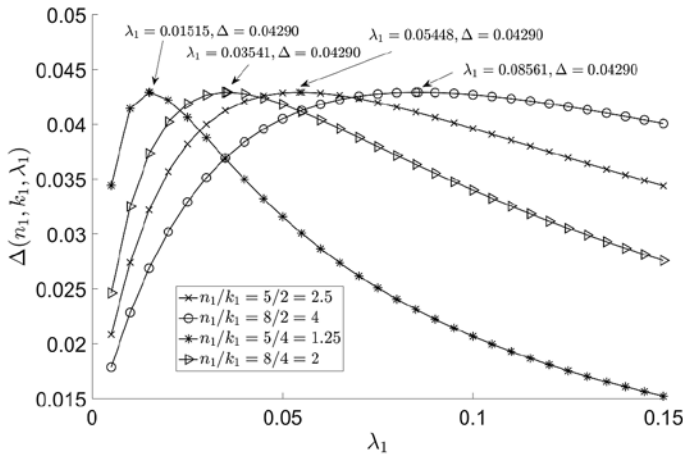


Fig. 7. Difference between A_S^{SA} and A_S^{NSA} for (n_1,k_1,λ_1), $\Delta(n_1,k_1,\lambda_1)$, where λ_1 ranges from 5×10^{-3} to 1.5×10^{-1} with step 5×10^{-3} and (n_1,k_1) is set as $(5,2)$, $(8,2)$, $(5,4)$ and $(8,4)$. The peak point of each curve is shown in the figure. The maximum of $\Delta(n_1,k_1,\lambda_1)$ in each curve is equal to 0.04290

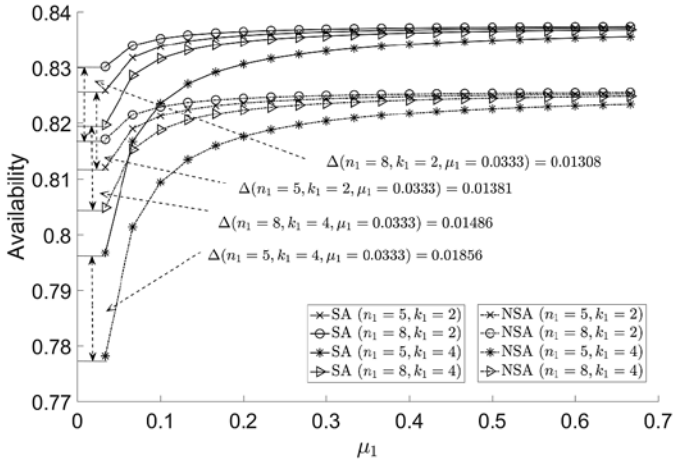


Fig. 8. A_S^{SA} and A_S^{NSA} for (n_1,k_1,μ_1), where μ_1 ranges from 3.33×10^{-2} to 6.67×10^{-1} with step 3.33×10^{-2} and (n_1,k_1) is set as $(5,2)$, $(8,2)$, $(5,4)$ and $(8,4)$

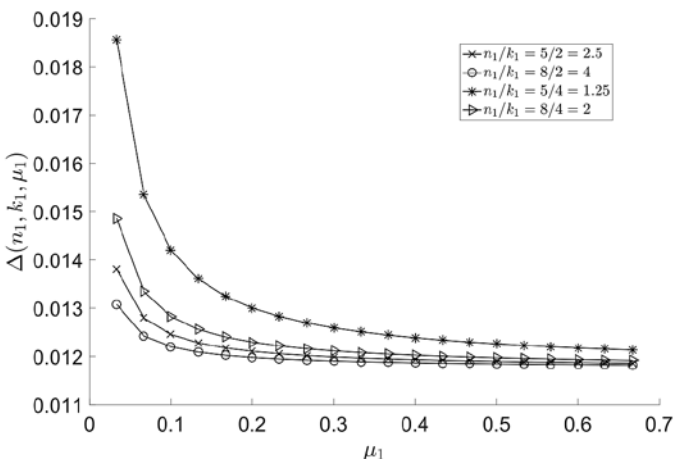


Fig. 9. Difference between A_S^{SA} and A_S^{NSA} for (n_1,k_1,μ_1), $\Delta(n_1,k_1,\mu_1)$, where μ_1 ranges from 3.33×10^{-2} to 6.67×10^{-1} with step 3.33×10^{-2} and (n_1,k_1) is set as $(5,2)$, $(8,2)$, $(5,4)$ and $(8,4)$. The minimum of curve exists for each curve

large value of n/k in a k-out-of-n subsystem. In addition, for each curve, decreases along with the increase of but the difference has a minimum. That is to say, the SA effect should be emphasized if both the value of n/k in a k-out-of-n subsystem and the repair rate of the subsystem are relatively small.

5. Conclusion

In this paper, we focus on SA of subsystems in a multi-component series system with different k-out-of-n:G warm standby subsystems. We relax the assumption that the k-out-of-n:G subsystems are independent for the system availability analysis. A transition process focusing on one subsystem with SA is constructed to analyze its failure rate. Then, a CTMC modeling the system failure and repair transition is constructed to derive the system stationary availability. We discuss the monotonicity of system stationary availability function based on the obtained expression. In numerical examples, we first verify our method by a Monte Carlo simulation. All relative errors are at or below10^{-3} level. Then, we make a comparison between the method assuming subsystem independence and the proposed method. We also perform a sensitivity analysis for the difference of the two in term of the system stationary availability. Findings in this paper can be drawn as follows. 1) The failure rates of subsystems with SA and their limits are derived. 2) The closed-form expressions for A_S, A_i, MTTF, MTTR, and SFF, considering SA are obtained. 3) The system stationary availability is a monotone function for its parameters. 4) The SA effect on the stationary availability should be emphasized in two cases, one is that the value of n/k and the failure rate of active components in a k-out-of-n subsystem are both relatively large or small, the other is the value of n/k and the repair rate are both relatively small. In future work, the arbitrary distributed failure times and repair times will be studied for practical application. Different repair strategies for the k-out-of-n:G subsystems will be modeled for availability analysis.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [No. 61573041 and No. 61573043]. The authors appreciate the comments and suggestions from the editors and reviewers.

References

1. Amari S V, Pham H, Misra R B. Reliability Characteristics of k-out-of-n Warm Standby Systems. Ieee Transactions on Reliability 2012; 61(4): 1007-1018, https://doi.org/10.1109/tr.2012.2220891.
2. Amiri M, Ghassemi-Tari F. A methodology for analyzing the transient availability and survivability of a system with repairable components. Applied Mathematics and Computation 2007; 184(2): 300-307, https://doi.org/10.1016/j.amc.2006.05.177.
3. Angus J E. On computing MTBF for a k-out-of-n:G repairable system. IEEE Transactions on Reliability 1988; 37(3): 312-313.
4. Barlow R E, Proschan F. Availability Theory for Multicomponent Systems A2 - KRISHNAIAH, PARUCHURI R, in Multivariate Analysis–III. 1973, Academic Press. 319-335, https://doi.org/10.1016/B978-0-12-426653-7.50027-2.
5. Cekyay B, Ozekici S. Reliability, MTIT and steady-state availability analysis of systems with exponential lifetimes. Applied Mathematical Modelling 2015; 39(1): 284-296, https://doi.org/10.1016/j.apm.2014.05.029.
6. Eryilmaz S. Reliability of a k-Out-of-n System Equipped With a Single Warm Standby Component. Ieee Transactions on Reliability 2013; 62(2): 499-503, https://doi.org/10.1109/tr.2013.2259202.
7. Habib M, et al. Design optimization and redundant dependency study of series k - out - of - n: G repairable systems. Ifac Papersonline 2016; 49(28): 126-131, https://doi.org/10.1016/j.ifacol.2016.11.022.
8. Habib M, et al. Multi-objective design optimisation of repairable k-out-of-n subsystems in series with redundant dependency. International Journal of Production Research 2017; 55(23): 7000-7021, https://doi.org/10.1080/00207543.2017.1346319.
9. Hevesh A H. Comments on: Steady-state Availability of k-out-of-n:G System with Single Repair. IEEE Transactions on Reliability 1984; 4(4): 324-324.
10. Huffman D, et al. Availability Analysis of Systems with Suspended Animation, in Annual Reliability and Maintainability Symposium, 2008 Proceedings. 2008. 284-+.
11. Kayedpour F, et al. Multi-objective redundancy allocation problem for a system with repairable components considering instantaneous availability and strategy selection. Reliability Engineering & System Safety 2017; 160: 11-20, https://doi.org/10.1016/j.ress.2016.10.009.
12. Kenyon R L, Newell R J. Steady-State Availability of k-out-of-n:G System with Single Repair. IEEE Transactions on Reliability 1983(2): 188-190.
13. Khalil Z S. Availability of Series Systems with Various Shut-off Rules. IEEE Transactions on Reliability 1985; R-34(2): 187-189, https://doi.org/10.1109/TR.1985.5221992.
14. Kim H. Optimal reliability design of a system with k-out-of-n subsystems considering redundancy strategies. Reliability Engineering & System Safety 2017; 167: 572-582, https://doi.org/10.1016/j.ress.2017.07.004.
15. Kuo C C, et al. Reliability-based measures for a retrial system with mixed standby components. Applied Mathematical Modelling 2014; 38(19-20): 4640-4651, https://doi.org/10.1016/j.apm.2014.03.005.
16. Kuo W, Zuo M J, Optimal reliability modeling: principles and applications. 2003: John Wiley & Sons.
17. Levitin G, Xing L, Dai Y. Sequencing Optimization in k-out-of-n Cold-Standby Systems Considering Mission Cost. International Journal of General Systems 2013; 42(8): 870-882, https://doi.org/10.1080/03081079.2013.819861.
18. Li X, Zuo M J, Yam R C M. Reliability analysis of a repairable k-out-of-n system with some components being suspended when the system is down. Reliability Engineering & System Safety 2006; 91(3): 305-310, https://doi.org/10.1016/j.ress.2005.01.010.
19. Moghaddass R, Zuo M J. Optimal design of a repairable k-out-of-n system considering maintenance. in reliability and maintainability symposium. 2011.
20. Moghaddass R, Zuo M J, Qu J. Reliability and Availability Analysis of a Repairable k-out-of-n: G System With R Repairmen Subject to Shut-Off Rules. Ieee Transactions on Reliability 2011; 60(3): 658-666, https://doi.org/10.1109/tr.2011.2161703.
21. Moghaddass R, Zuo M J, Wang W. Availability of a general k-out-of-n:G system with non-identical components considering shut-off rules using quasi-birth death process. Reliability Engineering & System Safety 2011; 96(4): 489-496, https://doi.org/10.1016/j.ress.2010.12.001.
22. Pham H. Commentary: Steady-state series-system availability. IEEE Transactions on Reliability 2003; 52(2): 146-147, https://doi.org/10.1109/tr.2003.811164.

23. She J, Pecht M G. Reliability of a k-out-of-n warm-standby system. IEEE Transactions on Reliability 1992; 41(1): 72-75, https://doi.org/10.1109/24.126674.
24. Sherwin D J. Steady-state series availability. Ieee Transactions on Reliability 2000; 49(2): 131-132, https://doi.org/10.1109/24.877327.
25. Shim J, Ryu H, Lee Y. Availability analysis of series redundancy models with imperfect switchover and interrupted repairs. Eksploatacja i Niezawodnosc-Maintenance and Reliability 2017; 19(4): 640-649, https://doi.org/10.17531/ein.2017.4.19.
26. Wang H, Pham H. Availability and maintenance of series systems subject to imperfect repair and correlated failure and repair. European Journal of Operational Research 2006; 174(3): 1706-1722, https://doi.org/10.1016/j.ejor.2005.03.030.
27. Wang K H, Ke J C. Probabilistic analysis of a repairable system with warm standbys plus balking and reneging. Applied Mathematical Modelling 2003; 27(4): 327-336, https://doi.org/10.1016/s0307-904x(02)00133-6.
28. Wang K H, Sivazlian B D. Reliability of a system with warm standbys and repairmen. Microelectronics Reliability 1989; 29(5): 849-860, https://doi.org/10.1016/0026-2714(89)90184-4.
29. Wu X Y, Hillston J, Feng C. Availability Modeling of Generalized k-Out-of-n:G Warm Standby Systems With PEPA. Ieee Transactions on Systems Man Cybernetics-Systems 2017; 47(12): 3177-3188, https://doi.org/10.1109/tsmc.2016.2563407.
30. Xie W, Liao H T, Jin T D. Maximizing system availability through joint decision on component redundancy and spares inventory. European Journal of Operational Research 2014; 237(1): 164-176, https://doi.org/10.1016/j.ejor.2014.02.031.
31. Yuan L. Reliability analysis for a k-out-of-n:G system with redundant dependency and repairmen having multiple vacations. Applied Mathematics and Computation 2012; 218(24): 11959-11969, https://doi.org/10.1016/j.amc.2012.06.006.
32. Zhang T L, Xie M, Horigome M. Availability and reliability of k-out-of-(M plus N): G warm standby systems. Reliability Engineering & System Safety 2006; 91(4): 381-387, https://doi.org/10.1016/j.ress.2005.02.003.

Yu WANG
Linhan GUO
Meilin WEN
Yi YANG
School of Reliability and Systems Engineering, Science and Technology on Reliability and Environmental Engineering Laboratory
Beihang University, Beijing, PR China

Emails: sy1714133@buaa,edu,cn, linhanguo@buaa.edu.cn, wenmeilin@buaa.edu.cn, chasesunny@163.com

Article citation info:
MIECZKOWSKI G. Criterion for crack initiation from notch located at the interface of bi-material structure. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 301–310, http://dx.doi.org/10.17531/ein.2019.2.15.

Grzegorz MIECZKOWSKI

CRITERION FOR CRACK INITIATION FROM NOTCH LOCATED AT THE INTERFACE OF BI-MATERIAL STRUCTURE

KRYTERIUM PĘKANIA STRUKTURY BI-MATERIAŁOWEJ Z OSTRYM KARBEM USYTUOWANYM NA INTERFEJSIE*

The fracture process of bi-material structure with the notch was analysed in this work. For fracture prediction, a criterion based on the Theory of Critical Distances was used. Under analysis were elements made of aluminium alloy and polymer combination (with a various structural notch-tip angle), which then were subjected to the three-point bending test. Values of critical loads resulting from the used hypothesis were compared with values obtained from the experiment. Validation of the selected criterion required defining a qualitative and quantitative description of singular stress fields present around the structural notch-tip area. Therefore, such solutions were obtained and methodology of their determining was discussed.

***Keywords**: crack initiation, interface fracture, bi-materials, singular stress fields, stress intensity factors.*

W pracy przedstawiono wyniki badań dotyczących pękania struktury bi-materiałowej z karbem usytuowanym na interfejsie. Do prognozowania inicjacji procesu pękania zastosowano kryterium oparte na punktowej teorii krytycznych dystansów. Analizowano elementy wykonane ze stopu aluminium i polimerów (PC, PMMA), które poddane były trójpunktowemu zginaniu. Wartości obciążeń krytycznych wynikających z wykorzystanej hipotezy porównano z wartościami uzyskanymi z eksperymentu. Walidacja wybranego kryterium wymagała określania jakościowego i ilościowego opisu osobliwych pól naprężeń, występujących w okolicy wierzchołkowej karbu strukturalnego. W związku z tym, uzyskano takie rozwiązania i omówiono metodykę ich otrzymywania.

***Słowa kluczowe**: inicjacja procesu pękania, bi-materiały, osobliwe pola naprężeń, współczynniki intensywności naprężeń.*

1. Introduction

Predicting durability of mechanical structures is a complex process that requires many factors to be taken into account. One of them is a structural complexity of structures. The structure is a physical object consisting of many elements, frequently made of materials with different mechanical and physical properties. Additionally, structures are usually subject to complex external loads, frequently of variable nature to various environmental factors. This results in various damage mechanisms – wear (corrosive, cavitational, frictional), fracture – often eliminating a device from further operation.

A point of initiation of cracks, in most cases, are voids and other material defects, and structural notches, the presence of which in the structure most often results from the necessity to join individual components to each other, providing the structure with the required feature characteristics, or requested material structure.

Notches can be classified into many different categories, depending on their shapes, location within a structure, material recurrence or material manufacturing technology. They all have one common characteristic – they cause a local increase of stress in the structure under load, and thus influence its durability and strength.

On the basis of observation and experiments, scientists have been trying for so long to define some critical conditions, at which material is damaged. Reference can be made to e.g. ideas by Galileo, Tresca, Beltrami, Coulomb, Mohr, Mises or Huber. They adopted some hypotheses, defining material strength, that until now are the basis for engineering calculations. Assuming homogeneity and continuity of a medium, of which a structural element has been made, they formulated limit values for a function, the exceeding of which resulted in material damage. These hypotheses did not take into account the significant influence of stress and strain field gradients on the strength. Theoretical development in solid state mechanics, and in theory of elasticity, in particular, has allowed for obtaining many analytical solutions describing local stress fields in the surrounding of both soft concentrators (e.g. in Kirsch's solution [8]), and sharp concentrators, generating singular stress fields (Sneddon's [31], and Williams' solutions [38]). Knowledge of new mathematical solutions has enabled formulation of consequent strength criteria including the influence of the presence of stress concentrators in homogeneous structural materials (e.g. Griffith [6], Sih [29], McClintock [16] and others).

Over the recent years, there has been a noticeable development of composite materials with pre-designed mechanical properties. As a rule, these are anisotropic materials or composites of a complex periodic structure.

Typical stress concentrator, present in layered composites, is a crack [30] or sharp notch located in the plane of bonding particular layers which form a composite [3, 5, 26]. Such concentrators are frequently present also in structural elements made by bonding two different materials with adhesive (such element can be considered a specific composite material). It is therefore necessary to define strength and resistance to cracking of composites (including mechanical properties of a bonding layer), where structural notches generate large stress gradients. A solution for this problem is an adequately formulated strength criterion. This criterion should include accurately determined equation with defined material constants, on basis of which it is possible to predict the moment of cracking process initiation. Predicting durability of elements with structural notches has not been the field of study taken by many scientists. In paper [12]

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

strength of two-phase elements with the structural notch was analysed experimentally. Material components were adhesive bonded. The authors performed three-point bending tests and determined values of crack initiation forces. A possibility of applying Leguillon's criterion for this type of elements (adhesive bonded uniform component with notch) was verified positively in paper [36].

As regards criteria for bi-material structures there are no available results published in the literature. The criterion that is frequently applied to homogeneous materials is the MClintock criterion. Thus, the main objective of this paper is to experimentally verify the possibility of applying this criterion (with appropriate modifications) for bi-material structures, where sharp structural notches generate singular stress fields. An idea for the criterion is described below.

Notations and Nomenclatures

a - Notch height

b - Gradient of combined stress

E - Young's modulus

f_{ik}^{I}, f_{ik}^{II} -Influence coefficient for stress

F - Load force at which stress intensity factors were calculated

F_k - Predicted critical force

g - Specimen thickness

h - Specimen height

H_o, H_1, H_2 - Influence coefficient for the characteristic equation

i - Material index (=1,2)

j - Generalised stress intensity factors/ Combined stresses index (=I,II)

K_E - Equivalent stress intensity factor

K_{Ec} - Critical value of the equivalent stress intensity factor

K_I, K_{II} -Generalised stress intensity factors

K_{Ic} - Fracture toughness

L - Spacing between supports set in the three-point bending test

Lc -Total specimen length

n - Nodes index

r, φ -Polar coordinates

u_r, u_φ - Displacements in polar coordinates

u_y - Vertical displacement applied as load condition in FEM simulations

α - Angle between the edge of material 1 and interface

β - Notch-tip angle

γ - Angle between the edge of material 2 and interface

Γ - Shear modulus' ratio

δ - Imaginary part of eigenvalue λ

λ - Eigenvalue

λ_r - Real part of eigenvalue λ

μ - Shear modulus

ν - Poisson's ratio

$\sigma_\varphi, \sigma_r, \tau_{r\varphi}$ -Stresses in polar coordinates

φ_0 - Cracking propagation angle

ψ -Mode mixity ratio

2. Fracture Criterion

With criterion proposed in paper [16] it is assumed that cracking will follow, if normal strain ε_φ in some small distance from the crack tip ρ_c reaches a critical value which can be noted as follows:

$$\varepsilon_\varphi(\rho_c) = \varepsilon_c \tag{1}$$

However, such stress form of this criterion, where the strain was replaced by an adequate component of normal stresses, had broader application.

In the approach as proposed in paper [23] it was assumed that propagation of cracking will follow only if circumferential stresses σ_φ at some finite distance $r=\rho_c$ reach critical value σ_c (2). Use of the condition formulated in such way, for elements with notches present in homogeneous materials, has been positively verified in many papers, e.g. [27]:

$$\max_{\varphi} \sigma_\varphi(\rho_c) = \sigma_c \tag{2}$$

Cracking propagation angle φ_0 is determined by maximizing σ_φ in relation to angle φ. For the elements with the notch located on adhesive bonding of uniform or bi-material structure, the cracking initially propagates along the interface. Thus, it can be predicted that $\varphi_0 = 0$ (Fig. 1) and the condition (2) will be written as:

$$\sigma_\varphi(\rho_c, 0) = \sigma_c \tag{3}$$

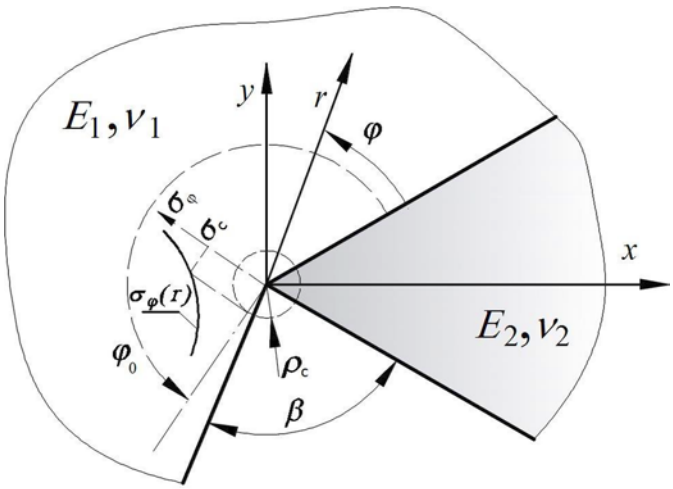


Fig. 1. Graphic interpretation McClintock criterion.

The parameter ρ_c, is considered as material constant and can be determined from experiment. For example, in condition (3), by taking advantage of dependence on circumferential stresses at crack tip and Griffith – Irwin criterion, the following is obtained:

$$\frac{K_{Ic}}{\sqrt{2\pi\rho_c}} = \sigma_c \tag{4}$$

and thus this allows for calculating the characteristic constant:

$$\rho_c = \frac{1}{2\pi}\left(\frac{K_{Ic}}{\sigma_c}\right)^2, \tag{5}$$

where K_{Ic} - fracture toughness, σ_c - tensile strength.

One advantage of the proposed method which is based on the so-called Theory of Critical Distances (TDC), is that in most cases in order to predict the fracture process it is not required to know the asymptotics description of stress fields - numerical solutions are used in the predicting process.

In the case when the fracture process occurs in a plane in which complex state of stresses is present, the use of numerical solutions may result in an erroneous prediction [13]. Most frequently in such situation, on the basis of analytical description of local stress fields, a global fracture criterion (using local parameter ρ_c) is formulated and it is based on an equivalent stress intensity factor [1, 13, 27] or minimum strain-energy density [9].

In this paper, on the basis of McClintock criterion, two concepts of prediction of fracture process initiation are proposed. First of them is based on an equivalent stress intensity factor, and the other on a dependence of condition (2) of critical stresses on proportions of shear and normal stresses occurring in the cracking plane. Detailed description of the concept is presented in section 6.

As it can be seen, when using the McClintock criterion for elements simultaneously loaded with shear and normal loads, it is required to know the distribution of stress fields occurring in the immediate vicinity to a singular point. Therefore, in the next section herein, forms of functions describing such stress fields and methodology of their determining will be discussed.

3. Analytical relations describing stress fields present in structural notch-tip area

A solution of the case of bi-material with the structural notch located on the interface (Fig. 2) was obtained using the approach applied by the authors of paper [21] for sharp corner in a uniform material.

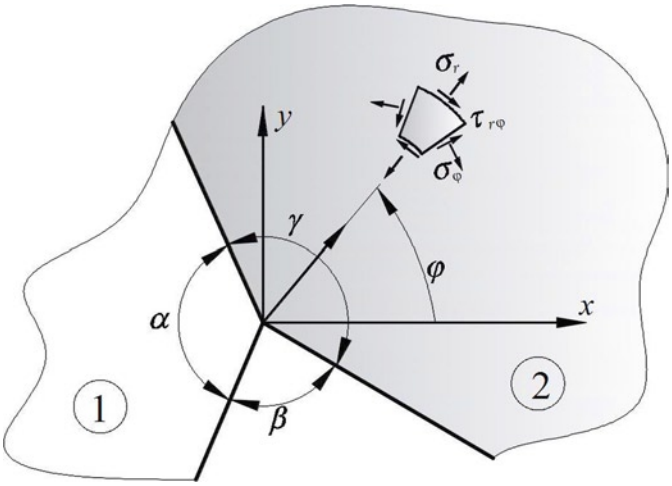


Fig. 2. Bi-material with structural notch located on interface

In the used method, the detailed description of which can be found in paper [19], by applying Airy's stress function it is possible to obtain general asymptotic solutions describing individual components of stress fields and displacements. For the analyzed bi-material structure the general asymptotic solutions are given by [18]:

$$u_{ri} = r^{\lambda}\left(A_i\cos\left((1+\lambda)\varphi\right)+B_i\sin\left((1+\lambda)\varphi\right)+C_i\cos\left((1-\lambda)\varphi\right)+D_i\sin\left((1-\lambda)\varphi\right)\right)$$

$$u_{\varphi i} = r^{\lambda}\left(-A_i\sin\left((1+\lambda)\varphi\right)+B_i\cos\left((1+\lambda)\varphi\right)-C_i\frac{\kappa+\lambda}{\kappa-\lambda}\sin\left((1-\lambda)\varphi\right)+D_i\frac{\kappa+\lambda}{\kappa-\lambda}\cos\left((1-\lambda)\varphi\right)\right)$$

$$\sigma_{ri} = r^{\lambda-1}\mu\left(A_i 2\lambda\cos\left((1+\lambda)\varphi\right)+B_i 2\lambda\sin\left((1+\lambda)\varphi\right)+C_i(3-\lambda)\frac{2\lambda}{\kappa-\lambda}\cos\left((1-\lambda)\varphi\right)+D_i(3-\lambda)\frac{2\lambda}{\kappa-\lambda}\sin\left((1-\lambda)\varphi\right)\right)$$

$$\sigma_{\varphi i} = r^{\lambda-1}\mu\left(-A_i 2\lambda\cos\left((1+\lambda)\varphi\right)-B_i 2\lambda\sin\left((1+\lambda)\varphi\right)+C_i(1+\lambda)\frac{2\lambda}{\kappa-\lambda}\cos\left((1-\lambda)\varphi\right)+D_i(1+\lambda)\frac{2\lambda}{\kappa-\lambda}\sin\left((1-\lambda)\varphi\right)\right)$$

$$\tau_{r\varphi i} = r^{\lambda-1}\mu\left(-A_i 2\lambda\sin\left((1+\lambda)\varphi\right)+B_i\,2\lambda\cos\left((1+\lambda)\varphi\right)+C_i(1-\lambda)\frac{2\lambda}{\kappa-\lambda}\sin\left((1-\lambda)\varphi\right)-D_i(1-\lambda)\frac{2\lambda}{\kappa-\lambda}\cos\left((1-\lambda)\varphi\right)\right). \quad (6)$$

where: $\mu_i = \dfrac{E_i}{2(1+\nu_i)}$ -shear modulus $\kappa_i = (3-\nu_i)/(1+\nu_i)$ - a plane stress, $\kappa_i = (3-4\nu_i)$ - a plane strain, ν_i -Poisson's ratio, i=1.2.

Particular solution is obtained by determining exponent λ and constants A_i, B_i, C_i, D_i. Constants are determined on basis of the following boundary conditions [18]:

1. of the left side surface of V-notch, for $\varphi = \alpha$;
$$\sigma_{\varphi 1} = \tau_{r\varphi 1} = 0$$
2. of the left side surface of V-notch, for $\varphi = -\gamma$;
$$\sigma_{\varphi 2} = \tau_{r\varphi 2} = 0$$
3. along the interface, for $\varphi = 0$;
$$u_{r1} = u_{r2}; u_{\varphi 1} = u_{\varphi 2};\ \sigma_{\varphi 1} = \sigma_{\varphi 2}; \tau_{r\varphi 1} = \tau_{r\varphi 2},$$

Moreover on basis of condition of zeroing matrix determinant of matrix boundary conditions the characteristic equation (7) can be determined, the individual roots of which determine the value of exponent λ in obtained asymptotic solutions (6). The characteristic equation takes the form [18]:

$$H_0 + \Gamma H_1 + \Gamma^2 H_2 = 0 \quad (7)$$

where:

$$H_o = \left(1-2\lambda^2+2\lambda^2\cos[2\alpha]\right)\left(1-\lambda^2+\lambda^2\cos[2\gamma]-\cos[2\gamma\lambda]\right)-$$
$$\kappa_1\left\{\cos\left[2(-\alpha+\gamma)\lambda\right]+\cos\left[2(\alpha+\gamma)\lambda\right]+\cos[2\alpha\lambda]\left(-2+4\lambda^2\sin[\gamma]^2\right)+\right.$$
$$\left.2\left(\lambda^2\sin[\gamma]^2-\sin[\gamma\lambda]^2\right)\kappa_1\right\}$$

$$H_1 = 5\lambda^2+\cos[2\alpha\lambda]+\cos[2\gamma\lambda]-2\cos\left[(-\alpha+\gamma)\lambda\right]^2-$$
$$\lambda^2\left(3\cos[2\alpha]-\cos[2(\alpha-\gamma)]+3\cos[2\gamma]+4\cos[2\gamma\lambda]\sin[\alpha]^2+4\left(\cos[2\alpha\lambda]+4\lambda^2\sin[\alpha]^2\right)\sin[\gamma]^2\right)+$$
$$\kappa_2\left(\cos\left[2(\alpha+\gamma)\lambda\right]-\cos[2\alpha\lambda]+2\sin[\gamma\lambda]^2-\lambda^2\left(1+\left(\cos[2\alpha]+4\cos[2\gamma\lambda]\sin[\alpha]^2-2\sin[\alpha]\sin[\alpha-2\gamma]\right)\right)\right)+$$
$$\kappa_1\left\{\cos\left[2(\alpha+\gamma)\lambda\right]+4\lambda^2\cos[\alpha]\sin[\alpha-\gamma]\sin[\gamma]+\cos[2\alpha\lambda]\left(-1+4\lambda^2\sin[\gamma]^2\right)+2\sin[\gamma\lambda]^2+\right.$$
$$\left.\left(\cos[2\alpha\lambda]+\cos[2\gamma\lambda]-2\cos\left[(-\alpha+\gamma)\lambda\right]^2+4\lambda^2\cos[\alpha-\gamma]\sin[\alpha]\sin[\gamma]\right)\kappa_2\right\},$$

$$H_2 = \left(1-2\lambda^2+2\lambda^2\cos[2\gamma]\right)\left(1-\lambda^2+\lambda^2\cos[2\alpha]-\cos[2\alpha\lambda]\right)-$$
$$\kappa_2\left\{\cos\left[2(-\alpha+\gamma)\lambda\right]+\cos\left[2(\alpha+\gamma)\lambda\right]+\cos[2\gamma\lambda]\left(-2+4\lambda^2\sin[\alpha]^2\right)+\right.$$
$$\left.2\left(\lambda^2\sin[\alpha]^2-\sin[\alpha\lambda]^2\right)\kappa_2\right\},$$

$$\Gamma = \frac{\mu_1}{\mu_2}.$$

On the basis of the characteristic equation (7) it can be inferred that eigenvalue λ depends on material constants and notch-tip angle. The roots of equation (7) cannot be determined analytically. They were calculated numerically. To this end, a special program was created in the Mathematica software.

Fig. 3 graphically illustrates solution of equation (7), where thick line was used to mark real eigenvalues λ (λ= λ_r, Im[λ]=0), while thin line – to mark real parts of complex eigenvalues λ (λ_r=Re[λ],

Im[λ]=δ), and dotted line – to mark imaginary parts of complex eigenvalues λ (δ = Im[λ]).

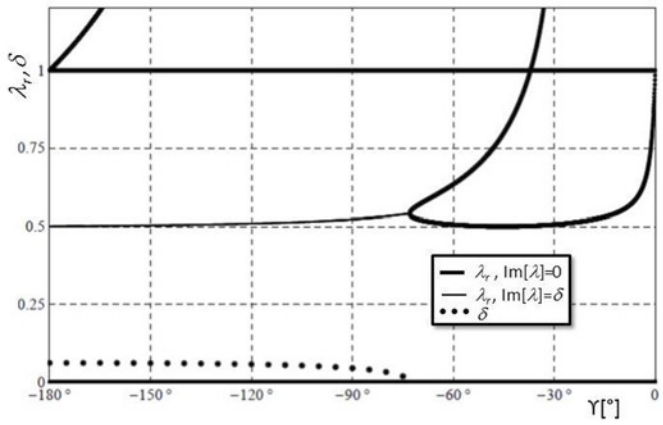


Fig. 3. Solution of the characteristic equation (7) for Γ = 0.033, α=180°, ν_1=0.37, ν_2=0.35, (plane strain)

The solution obtained implies that depending on material constants and notch geometry, there can be one or more singularities with a real or complex exponent λ. What is more, it is worth noting that for the notch problem located in the multi-material, the characteristic equations cannot be obtained independently for mode I and mode II [3, 5]

Since stresses can be described with complex exponent λ, generalised stress intensity factors were defined alike the authors of paper [32] (for the interfacial crack):

$$\left(\sigma_\varphi + \mathrm{i}\tau_{r\varphi}\right)_{\varphi=0} = \frac{K_I + iK_{II}}{\sqrt{2\pi}r^{1-\lambda_r}}\left(\frac{r}{2a}\right)^{\mathrm{i}\delta}\cosh\left[\pi\delta\right], \tag{8}$$

where dimension *a* can be considered as e.g. notch height.

Using an equation (6), boundary conditions and the adopted generalized stress intensity factors definition (8), an analytical description of the stress fields occurring in the notch tip area can be obtained [18]:

$$\sigma_{ik} = \frac{\cosh\left[\pi\delta\right]\sqrt{K_I{}^2 + K_{II}{}^2}}{\sqrt{2\pi}r^{1-\lambda_r}}\left(\sin\left[\arctan\left[\frac{K_{II}}{K_I}\right] + \delta\log\left[\frac{r}{2a}\right]\right]\mathrm{Re}\left[f_{ik}^{I}\right] + \cos\left[\frac{K_{II}}{K_I} + \delta\log\left[\frac{r}{2a}\right]\right]\mathrm{Re}\left[f_{ik}^{II}\right]\right), \tag{9}$$

where:

$$f_{rr}^{I} = M^{-1}\begin{pmatrix}(\lambda-1)\cos\left[(1+\lambda)\varphi\right]\left(\lambda\sin\left[2\epsilon\right]-\sin\left[2\epsilon\lambda\right]\right)-(1+\lambda)\left(\lambda-\lambda\cos\left[2\epsilon\right]+\cos\left[2\epsilon\lambda\right]-1\right)\sin\left[(1+\lambda)\varphi\right]\\+(\lambda-3)\left(\lambda\sin\left[2\epsilon+(\lambda-1)\varphi\right]+(1+\lambda)\sin\left[\varphi-\lambda\varphi\right]-\sin\left[2\epsilon\lambda+\varphi-\lambda\varphi\right]\right)\end{pmatrix},$$

$$f_{rr}^{II} = -M^{-1}\begin{pmatrix}(1-\lambda)\left(\lambda\cos\left[2\epsilon\right]+\cos\left[2\epsilon\lambda\right]-1-\lambda\right)\cos\left[(1+\lambda)\varphi\right]+\\(\lambda-3)\left(-\lambda\cos\left[2\epsilon+(\lambda-1)\varphi\right]+(\lambda-1)\cos\left[\varphi-\lambda\varphi\right]+\cos\left[2\epsilon\lambda+\varphi-\lambda\varphi\right]\right)-\\+(\lambda-1)\left(\lambda\sin\left[2\epsilon\right]+\sin\left[2\epsilon\lambda\right]\right)\sin\left[(1+\lambda)\varphi\right]\end{pmatrix},$$

$$f_{\varphi\varphi}^{I} = M^{-1}(1+\lambda)\begin{pmatrix}(1-\lambda)\sin\left[(1+\lambda)\varphi\right]+\lambda\sin\left[2\epsilon+(\lambda-1)\varphi\right]+\\(1+\lambda)\sin\left[\varphi-\lambda\varphi\right]-\sin\left[2\epsilon\lambda+\varphi-\lambda\varphi\right]-\lambda\sin\left[2\epsilon-(1+\lambda)\varphi\right]+\sin\left[2\epsilon\lambda-(1+\lambda)\varphi\right]\end{pmatrix},$$

$$f_{\varphi\varphi}^{II} = -M^{-1}\begin{pmatrix}(\lambda-1)\left(-1-\lambda+\lambda\cos\left[2\epsilon\right]+\cos\left[2\epsilon\lambda\right]\right)\cos\left[(1+\lambda)\varphi\right]+(\lambda-1)\left(\lambda\sin\left[2\epsilon\right]+\sin\left[2\epsilon\lambda\right]\right)\sin\left[(1+\lambda)\varphi\right]\\+(1+\lambda)\left(\lambda\cos\left[2\epsilon+(\lambda-1)\varphi\right]-\lambda\cos\left[\varphi-\lambda\varphi\right]+2\sin\left[\epsilon\lambda\right]\sin\left[\epsilon\lambda+\varphi-\lambda\varphi\right]\right)\end{pmatrix},$$

$$f_{r\varphi}^{I} = M^{-1}\begin{pmatrix}(1+\lambda)\left(1-\lambda+\lambda\cos\left[2\epsilon\right]-\cos\left[2\epsilon\lambda\right]\right)\cos\left[(1+\lambda)\varphi\right]+\\(\lambda-1)\left(\lambda\cos\left[2\epsilon+(\lambda-1)\varphi\right]-(1+\lambda)\cos\left[\varphi-\lambda\varphi\right]+\cos\left[2\epsilon\lambda+\varphi-\lambda\varphi\right]\right)+\\+(1+\lambda)\left(\lambda\sin\left[2\epsilon\right]-\sin\left[2\epsilon\lambda\right]\right)\sin\left[(1+\lambda)\varphi\right]\end{pmatrix},$$

$$f_{r\varphi}^{II} = M^{-1}(\lambda-1)\begin{pmatrix}(1+\lambda)\sin\left[(1+\lambda)\varphi\right]-\lambda\sin\left[2\epsilon+(\lambda-1)\varphi\right]\\-(\lambda-1)\sin\left[\varphi-\lambda\varphi\right]-\sin\left[2\epsilon\lambda+\varphi-\lambda\varphi\right]+\lambda\sin\left[2\epsilon-(1+\lambda)\varphi\right]+\sin\left[2\epsilon\lambda-(1+\lambda)\varphi\right]\end{pmatrix},$$

$M = 2\left(\lambda^2 - \lambda^2\cos\left[2\epsilon\right] + \cos\left[2\epsilon\lambda\right] - 1\right)$, $\epsilon = \alpha$ -for the material 1 and $\epsilon = -\gamma$ for the material 2.

Below a particular form of stress fields for angle φ=0, i.e. along the interface line is presented [18]:

$$\sigma_{\varphi 1,2_{\varphi=0}} = \frac{1}{\sqrt{2\pi}}\sqrt{K_I{}^2 + K_{II}{}^2}\,r^{\lambda_r-1}\cos\left[\arctan\left[\frac{K_{II}}{K_I}\right] + \delta\ln\left[\frac{r}{2a}\right]\right]\cosh\left[\pi\delta\right],$$

$$\tau_{r\varphi 1,2_{\varphi=0}} = \frac{1}{\sqrt{2\pi}}\sqrt{K_I{}^2 + K_{II}{}^2}\,r^{\lambda_r-1}\sin\left[\arctan\left[\frac{K_{II}}{K_I}\right] + \delta\ln\left[\frac{r}{2a}\right]\right]\cosh\left[\pi\delta\right]. \tag{10}$$

When the exponent is a real value ($\delta = 0$) dependence (10) will be simplified to the following form:

$$\sigma_{\varphi 1,2_{\varphi=0}} = \frac{1}{\sqrt{2\pi}}K_I\,r^{\lambda_r-1},\ \tau_{r\varphi 1,2_{\varphi=0}} = \frac{1}{\sqrt{2\pi}}K_{II}\,r^{\lambda_r-1}. \tag{11}$$

For a quantitative description of stresses, it is necessary to determine values of K_j. They were determined based on comparing the obtained analytical solutions with stresses obtained from *FEM* solution. Due to the fact that the main purpose of the presented paper was experimental verification of possibilities of using the McClintock criterion, *FEM* models of specimens were prepared, the strength of which was tested by means of experiments in [11, 12].

4. Testing specimens and FEM model

Specimens were modelled in numerical simulations, the geometry and material properties of which were identical as in specimens used in experimental testing (prepared by authors of papers [11, 12]).

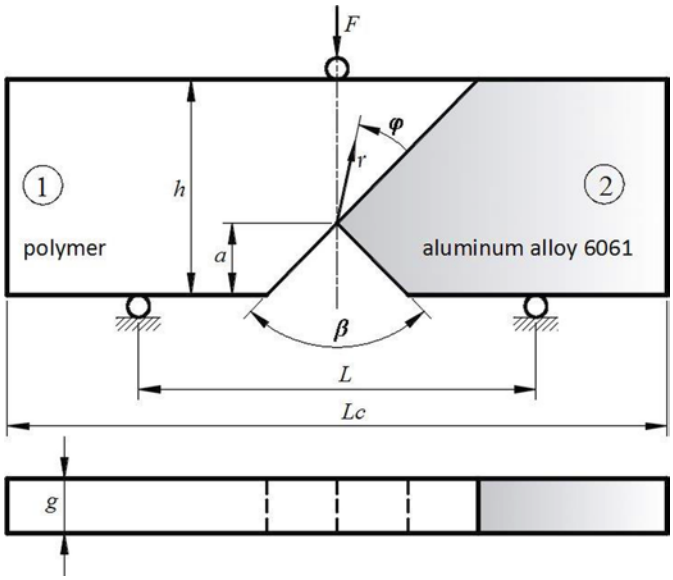


Fig. 4. Geometry and load of specimens with structural notch

Two types of specimens were analysed:
– component 1 made of PC (Polycarbonate) while component 2 of aluminum alloy 6061;

– component 1 made of PMMA (Poly(methyl methacrylate)) while component 2 of aluminum alloy 6061.

Components were bonded using Weld-on® 10 adhesive. The adhesive has been selected so that its stiffness properties were close to those of polymers. Therefore, it was possible to confirm the assumption that it is the bi-material that is modelled, not a three-layer composite. In this case, the adhesive interface could be treated as a layer without any thickness, but with different strength and fracture toughness than the polymers or the aluminum alloys separately. During the bonding process, achieving the minimum adhesive thickness was one of the goals. To obtain the minimum thickness of the adhesive, the specimens were joint under high pressure using a special holder. They were allowed to cure for a period of 24 hours to achieve the required bonding strengths. The thickness of the adhesive layer was not measured. Such information was not needed for the prediction of fracture toughness, assuming that the analyzed structure is the bi-material.

It is worth noting that using the selected adhesive in case where two different metal alloys are being combined, e.g. steel and aluminum alloy, ignoring the adhesive layer thickness and treating such a structure as a bi-material would be unacceptable. In that case, the approach based on the Traction-Separation criterion [25] can be used to predict the critical load. Namely, the adhesive layer should be modeled using special finite elements (cohesive elements). To define such elements it is necessary to know such parameters as maximum normal traction at the interface, normal separation across the interface where the maximum normal traction is attained and the limit value of separation. These parameters can be determined, knowing the thickness of the adhesive layer, based on experimental tests.

The carried out analyses were aimed at determining the applicability of the proposed fracture criterion for bi-material structures in a situation where a complex state of stress occurs on the interface, and the structural notch generates singular stress fields. Whereby the stress fields, depending on the geometrical and material features of the structure, can be described using real or complex λ exponents. Specimens with different notch-tip β angle were used in the study. β angles were selected in a way to obtain instances when stresses are described both in real and complex λ exponent. What's more, the variation of the notch-tip angles allowed for obtaining different proportions of tangential and normal stresses occurring in the plane of the connection. In all specimens, equal notch height a (measured from the lower surface of the sample), overall dimensions and the position of the support and loading points in the three-point bending test were assumed. This was to ensure the same boundary conditions (fixing and loading) for all examined specimens. Overall dimensions were chosen arbitrarily considering the capabilities of both the research stand and the device used for specimen preparation.

Specimen dimensions were as follows: total length Lc=254 mm, a spacing between supports L=90 mm, notch height a=19.1 mm, sample height h=50.8 mm, thickness g=5.4 mm, respectively. As regards the notch-tip angle β, three cases were considered: β= 30°, β= 90° and β= 120°. The specimens used in the tests are presented in Table 1. Material specification for individual components is given in Table 2.

Table 2. Mechanical properties of individual components of specimens [12]

	Young's modulus E [GPa]	Poisson's ratio ν
aluminum alloy 6061	70	0.35
PC	2.38	0.37
PMMA	3.79	0.37

Table 1. Tested bi-material structure

No	Tested bi-material structure	Notch-tip angle β[°]
1	PC/ aluminum alloy 6061	30
2	PC/ aluminum alloy 6061	90
3	PC/ aluminum alloy 6061	120
4	PMMA/ aluminum alloy 6061	30
5	PMMA/ aluminum alloy 6061	90
6	PMMA/ aluminum alloy 6061	120

Tested specimens (Fig. 4, Tab. 1) were modelled with *FEM*, using ANSYS application. Fig. 5 shows, for a given specimen, division to finite elements and boundary conditions.

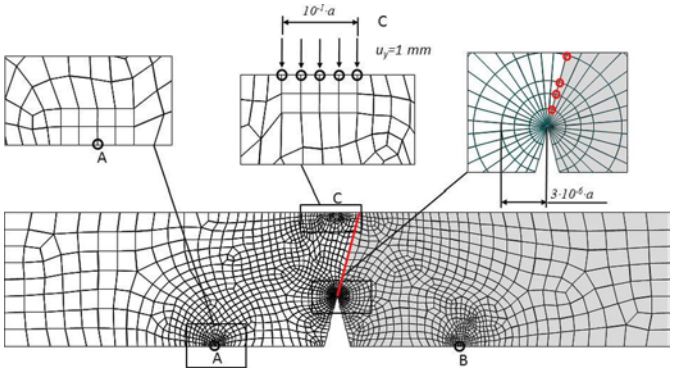


Fig. 5. Division to finite elements and conditions of mounting and loading of the specimen with notch with tip angle β= 300, the red colour was used to mark nodes from which the stresses were read for stress intensity factors calculation

Plane specimens were described with quadrangle, eight-node finite elements with increased refinement (according to the arithmetic series) in the tip area, with triangular special elements [35] surrounding singular point (Fig. 5). The total length of the lateral edges of the last three elements depended on the notch height a and were assumed to be $3\ 10^{-6}\ a$ for all specimens. Due to the high density of the finite element mesh in the notch tip area, the prepared models contained approximately 10000 finite elements. As already mentioned before, the specimens can be treated as a bi-material structure. Therefore, the adhesive layer was not included in the prepared numerical models. As for the connection conditions of individual components, the nodes laying on the interface were shared between both materials. There was not allowed for slip between components on the interface.

Due to the fact that it is difficult to determine real frictional conditions of the contact between supports and material of the specimen, which are present during the three-point bending test, certain simplifications should be adopted. Thus, two methods of mounting were tested:

I sliding supports (disabled possibility of moving vertically in nodes located in support points A and B);

II non-sliding supports (disabled possibility of moving vertically and horizontally in nodes located in support points A and B).

As regards the load conditions, the specimens were loaded with a constant vertical displacement u_y = 1mm, applied at selected nodes (point C). Loading force F was determined on the basis of stresses in the nodes to which the displacement u_y was applied.

Numerical calculations were carried out for plane stress and plane strain cases.

As already mentioned before, in the FEM models, the adhesive, as a separate material layer, was not included. Despite this, it was possible to include the interface's strength properties in the used fracture criterion. On the basis of numerical simulations, stress intensity factors were determined. They were used to calculate the value of the predicted failure function (described further in the paper), which were then compared with the critical values. The interface's strength properties in the tested specimens was taken into account in that the critical values of the failure function and the ρ_c parameter were determined based on the bonding tensile strength and bonding fracture toughness (Tab. 3). Both of parameters were determined experimentally for the bi-materials made of PMMA and aluminum alloy as well as PC and aluminum alloy.

5. Description of method applied to determine generalised stress intensity factors

To determine values of generalised stress intensity factors K_j extrapolation method was applied. This method, unlike e.g. energy methods [37], or methods based on the application of special finite elements [4], is less complex. The disadvantage of this method, though, is a necessity to use a high density of mesh of division to finite elements in the tip area of stress concentrator. Additionally, accuracy of the results is influenced by the selection of area, where the numerical solution is compared with the analytical solution. This inconvenience can be eliminated by using terms of higher order [17, 24, 33] in the analytical description or determining an adequate criterion for the selection of nodes, for which values of stresses obtained from FEM modelling are compared with the known analytical solution. Such criterion was specified in papers [20] (for the case of interfacial crack) and [18] (for the problem of structural notch). As is well known, if the stress chart of the type $\sigma = Ar^{-b}$ in logarithmic system is linear, the line gradient equals -*b*.

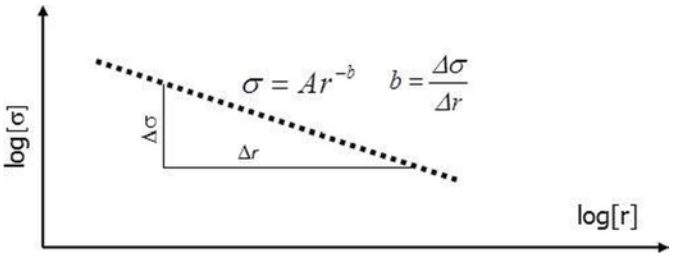


Fig. 6. Graphical interpretation of singular stress fields with theoretical gradient b [18]

Thus, when determining stress intensity factors, numerical and analytical solutions are compared only for nodes with the gradient of *b*.

For the case of the structural notch, stress components are always simultaneously dependent on K_I and K_{II}, and therefore, in order to use the above criterion, it is necessary to determine the so-called "combined stresses" [18]:

$$\sigma_I(r,0) = Sech[\pi\delta]\left(\sigma\cos\left[\delta\ln\left[\frac{r}{2a}\right]\right] + \tau\sin\left[\delta\ln\left[\frac{r}{2a}\right]\right]\right) = \frac{K_I}{\sqrt{2\pi}r^{1-\lambda_r}}, \tag{12}$$

$$\sigma_{II}(r,0) = \left(\tau\cos\left[\delta\log\left[\frac{r}{2a}\right]\right] - \sigma\sin\left[\delta\log\left[\frac{r}{2a}\right]\right]\right) Sech[\pi\delta] = \frac{K_{II}}{\sqrt{2\pi}r^{1-\lambda_r}}, \tag{13}$$

where - σ,τ are circumferential and shear stresses, respectively, obtained from *FEM* modelling (obtained from nodes located on the interface).

According to the adopted criterion when determining the sought generalised stress intensity factors, pairs of nodes with the gradient of $b = (\lambda_r - 1) \pm 0.01$ were taken into account.

The "combined stresses" $\sigma_j(r,0)$ in the distance of r_n and r_{n+1} from the notch tip can be noted as follows:

$$\sigma_j(r_n,0) = \frac{K_j}{\sqrt{2\pi}r^{1-\lambda_r}}(1+cr_n), \sigma_j(r_{n+1},0) = \frac{K_j}{\sqrt{2\pi}r^{1-\lambda_r}}(1+cr_{n+1}). \tag{14}$$

Using equations (12) – (14) after simple mathematical transformations, formulas (15) are obtained: (enabling to determine factors K_j at some distance from the notch tip):

$$\left.\begin{aligned} K_I &= \frac{\sqrt{2\pi}(r_n r_{n+1})^{1-\lambda_r}}{r_n - r_{n+1}} sech[\pi\delta]\left\{r_n^{\lambda_r}\left(\sigma_{(r_{n+1})}\cos\left[\delta\ln\left[\frac{r_{n+1}}{2a}\right]\right] + \tau_{(r_{n+1})}\sin\left[\delta\ln\left[\frac{r_{n+1}}{2a}\right]\right]\right) - \right. \\ &\left. r_{n+1}^{\lambda_r}\left(\sigma_{(r_n)}\cos\left[\delta\ln\left[\frac{r_n}{2a}\right]\right] + \tau_{(r_n)}\sin\left[\delta\ln\left[\frac{r_n}{2a}\right]\right]\right)\right\} \\ K_{II} &= \frac{\sqrt{2\pi}(r_n r_{n+1})^{1-\lambda_r}}{r_n - r_{n+1}} sech[\pi\delta]\left\{r_n^{\lambda_r}\left(\tau_{(r_{n+1})}\cos\left[\delta\ln\left[\frac{r_{n+1}}{2a}\right]\right] - \sigma_{(r_{n+1})}\sin\left[\delta\ln\left[\frac{r_{n+1}}{2a}\right]\right]\right) - \right. \\ &\left. r_{n+1}^{\lambda_r}\left(\tau_{(r_n)}\cos\left[\delta\ln\left[\frac{r_n}{2a}\right]\right] - \sigma_{(r_n)}\sin\left[\delta\ln\left[\frac{r_n}{2a}\right]\right]\right)\right\} \end{aligned}\right\} \tag{15}$$

Calculated stress intensity factor, for selected nodes (with the gradient of $b = (\lambda_r - 1) \pm 0.01$), is approximated with a straight line and this way generalised stress intensity factors K_j are determined.

It is worth noting that if exponent λ is a real number ($\delta = 0$), the dependence (15) is simplified to the form as given in paper [14]:

$$K_I = \frac{\sqrt{2\pi}(r_n r_{n+1})^{1-\lambda_r}\left(r_{n+1}^{\lambda_r}\sigma_{(r_n)} - r_n^{\lambda_r}\sigma_{(r_{n+1})}\right)}{r_{n+1} - r_n}. \tag{16}$$

6. Test results and discussion

As it was already discussed, in order to verify the McClintock stress criterion, it is necessary to know the qualitative and quantitative description of stress fields which occurred in the cracking plane, critical parameters and experimental data (damage loads). Damage loads were taken from paper [11, 12]. Since after the cracking process was initiated, the crack propagated along the interface, in the tested criterion critical parameters characterising properties of adhesive layer / interface [10] were used, for which critical distance σ_c was determined, according to formula (5). The methods of determining critical parameters for the adhesive layer are discussed in [7, 10, 22, 25].

In order to determine the quantitative description of mechanical fields, generalised stress intensity factors K_j were calculated. They were determined numerically with extrapolation method using data obtained from *FEM* modelling and analytical solutions. The extrapolation method, *FEM* modelling and analytical solution were presented and discussed in the previous sections herein.

Values of the calculated generalised stress intensity factors K_j, for all types of samples (Fig. 4), are given in Tables 4-5.

Exponents λ, obtained from equation (7), for material constants given in Table 2, are listed in Table 6.

Table 3. Strength properties of adhesive - Weld-on® 10 [10]

	Tensile strength σ_c [MPa]	Fracture toughness K_{Ic} [MPa m $^{0.5}$]	Critical distance ρ_c [mm]
PC/ aluminum alloy 6061	11.35	0.24	0.071
PMMA/ aluminum alloy 6061	12.85	0.28	0.075

Table 4. Values of generalised stress intensity factors Kj and applied force F PC /aluminum alloy 6061

	mounting conditions					
	sliding supports			non-sliding supports		
β [°]	K_I [Pa m$^{1-\lambda r}$]	K_{II} [Pa m$^{1-\lambda r}$]	F [N]	K_I [Pa m$^{1-\lambda r}$]	K_{II} [Pa m$^{1-\lambda r}$]	F [N]
30	6621327.4* 5654415.8**	16207.8* 565904.7**	4139.0* 3508.3**	2670172.3* 2337935.8**	334892.9* 576351.2**	5598.9* 4718.3**
90	6953117.8* 7929593.6**	-2675922.6* -1851771.4**	3532.6* 3004.2**	2654647.7* 3089615.1**	-1281221.0* -622169.7**	5071.6* 4276.7**
120	3224964.6* 3206237.2**	-4631416.6* -6451442.2**	3345.1* 2853.9**	1156739.9* 1185145.1**	-1821056.5* -2538667.0**	5001.9* 4225.0**
*-plane strain, **-plane stress						

Table 5. Values of generalised stress intensity factors Kj and applied force F PMMA/aluminum alloy 6061

	mounting conditions					
	sliding supports			non-sliding supports		
β [°]	K_I [Pa m$^{1-\lambda r}$]	K_{II} [Pa m$^{1-\lambda r}$]	F [N]	K_I [Pa m$^{1-\lambda r}$]	K_{II} [Pa m$^{1-\lambda r}$]	F [N]
30	10590286.9* 9092521.1**	88797.8* 449519.4**	6510.0* 5522.1**	4304874.73* 3774299.6**	466048.7* 845462.9**	8818.2* 7435.5**
90	15655266.1* 13838295.9**	-8259321.6* -4731359.1**	5572.9* 4742.3**	6032440.86* 5427305.2**	-1915526.6* -578767.5**	8002.9* 6753.4**
120	4796910.9* 4040648.9**	-6494991.4* -6507052.4**	5279.8* 4507.1**	1798019.9* 1525983.9**	-2641592.9* -2594161.8**	7892.7* 6671.4**
*-plane strain, **-plane stress						

Table 6. Values of exponents λ

β [°]	type of specimen			
	PC/ aluminum alloy 6061		PMMA/ aluminum alloy 6061	
	λ_r	δ	λ_r	δ
30	0.5032* 0.5033**	0.0611* 0.0958**	0.5051* 0.5052**	0.0579* 0.0913**
90	0.5222* 0.5231**	0.0450* 0.0810**	0.5339* 0.5352**	0.0235* 0.0646**
120	0.5058* 0.5324**	0* 0**	0.5003* 0.5071**	0* 0**
*-plane strain, **-plane stress				

As already mentioned, if both shear and normal stresses are present in the cracking plane, in order to predict the cracking process the equivalent stress intensity factor K_E can be used. For the analysed case it is defined (on the basis of analytical description of local stress fields) using formula (17):

$$K_E = \cosh(\pi\delta)\sqrt{{K_I}^2 + {K_{II}}^2}, \tag{17}$$

Critical value of the factor K_{Ec} can be determined by solving the below system of equations (18):

$$\frac{\sigma_\varphi(\rho_c,0)}{\sigma_c} = 1, \left(\frac{\partial\sigma_\varphi}{\partial\varphi}\right)_{\varphi=0,r=\rho_c} = 0. \tag{18}$$

After the system of equations (18) is solved - using formulas (5), (9) and (10) a dependence is obtained, allowing to determine the value of critical equivalent stress intensity factor (19):

$$K_{Ec} = (2\pi)^{\lambda_r - \frac{1}{2}} \left(\frac{{K_{Ic}}^2}{\sigma_c^2}\right)^{1-\lambda_r} \sigma_c. \tag{19}$$

It is worth noting that for the tension element with a crack ($\lambda_r = 0.5$) or a notch, with tip angle equal to π ($\lambda_r = 1$), the dependence (19) is simplified accordingly to the following forms: $K_{Ec} = K_{Ic}; K_{Ec} = \sqrt{2\pi}\sigma_c$, which is consistent with the literature data.

Assuming that the fracture process will be initiated when:

$$K_E = K_{Ec}, \tag{20}$$

predicted critical force can be calculated from the following condition:

$$F_k = \frac{K_{Ec}F}{K_E}, \tag{21}$$

where F is a force at which K_E (17) were calculated.

As already mentioned it is difficult to model the actual frictional and contact conditions [2, 15] occurring in the support area with numerical modelling of three-point bending test. Therefore Figure 7 (for plane stress) and Table 7 (for plane strain) provide values for the predicted force F_k (21) determined by using two variants of specimen fixing in *FEM* models. It is obvious that the real critical force will take values from a range limited by forces estimated when using sliding supports and non-sliding supports in numerical models.

The values of critical forces determined using the formula (21) were compared with the experimental data, which is shown in Figure 7 and Table 6. The experimental crack initiation load is an average obtained from at least three experiments performed for each type of specimen.

On the basis of the obtained results it can be stated that distribution of the estimated critical forces is consistent with experimental data. Better convergence of results, both for plane stress and plane strain condition, was obtained by using non-sliding supports in the *FEM* model:

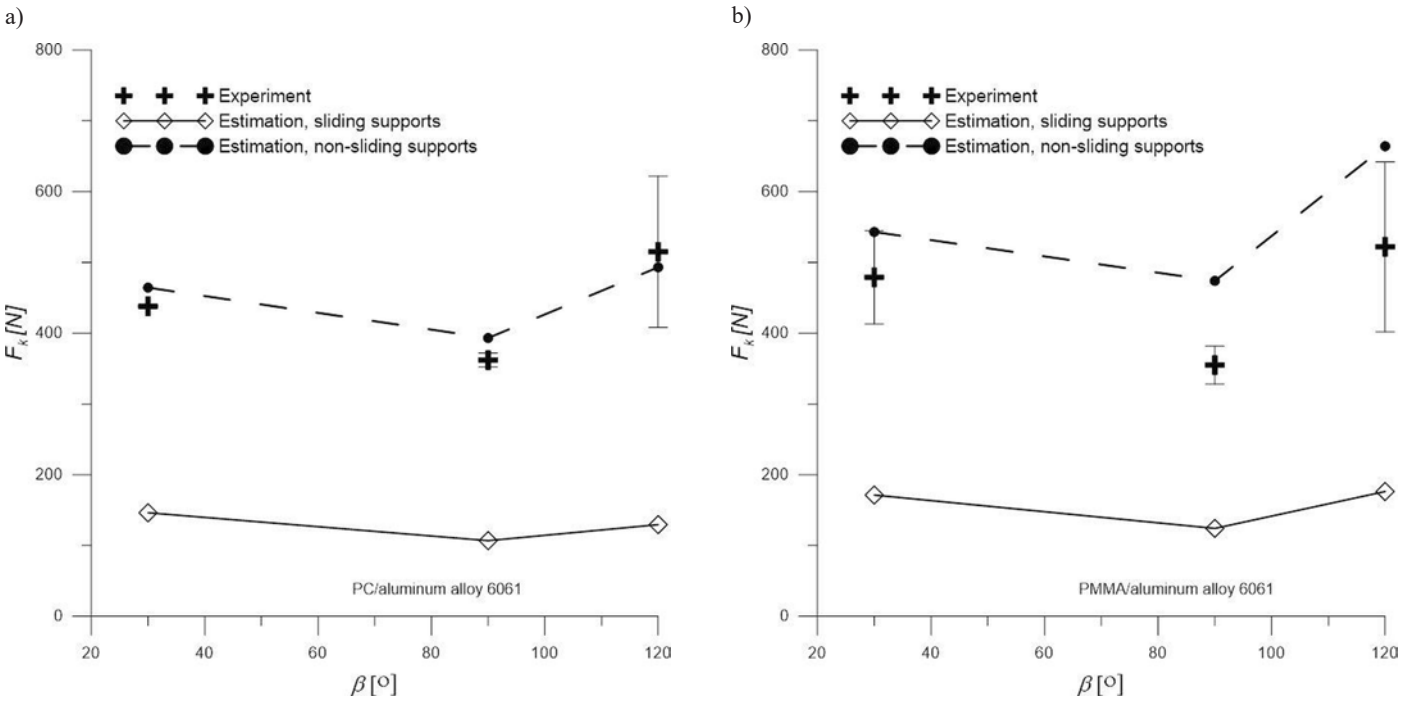


Fig. 7. Comparison of critical stress values, obtained from tested criterion with values obtained from experimental testing [12], bonded polycarbonate/ aluminum alloy 6061specimens, b) bonded PMMA/aluminum alloy 6061 specimens, plane stress

Table 7. Values of critical forces Fk (21) calculated for various types of specimens, plane strain

β [°]	PC/ aluminum alloy 6061		PMMA/ aluminum alloy 6061	
	experiment [12]	estimation (21)	experiment [12]	estimation (21)
30	438±1	151.9* 505.7**	479±66	177.7* 588.6**
90	362±10	139.2* 403.3**	355±27	121.3* 487.2**
120	515±107	150.3* 587.9**	522±120	183.6* 693.6**
*- sliding supports, **- non-sliding supports				

When determining critical forces in the analytical description first singular term was used only. Use of terms of a higher order was not necessary due to a fact that analytical (when using the first singular term only) and numerical solutions matched one another in an area larger than the critical distance ρ_c, which is shown in Figure 8.

It is worth noting that prediction of fracture toughness with use of the concept of equivalent stress intensity factor is quite complicated. Since it is necessary to determine exponents λ and generalized stress intensity factors K_j. For uniform materials it is not a major problem, as approximate formulas are available which allow for calculation of generalized stress intensity factors, and exponent λ depends only on tip angle of the notch and can be easily determined (exponent values λ can be found, e.g. in paper [28]). For a bi-material with a notch situated on the interface, both K_j and λ depend on geometrical and material features of the structure and should be determined individually for each analysed case. Due to the above inconveniences, the paper attempts to develop a procedure of predicting fracture toughness which might be more practical from the engineering point of view.

Authors of many papers, e.g. [30] indicate that critical value of energy release factor depends on the ratio of shear and normal stresses which are present in the cracking plane. Therefore, critical stresses

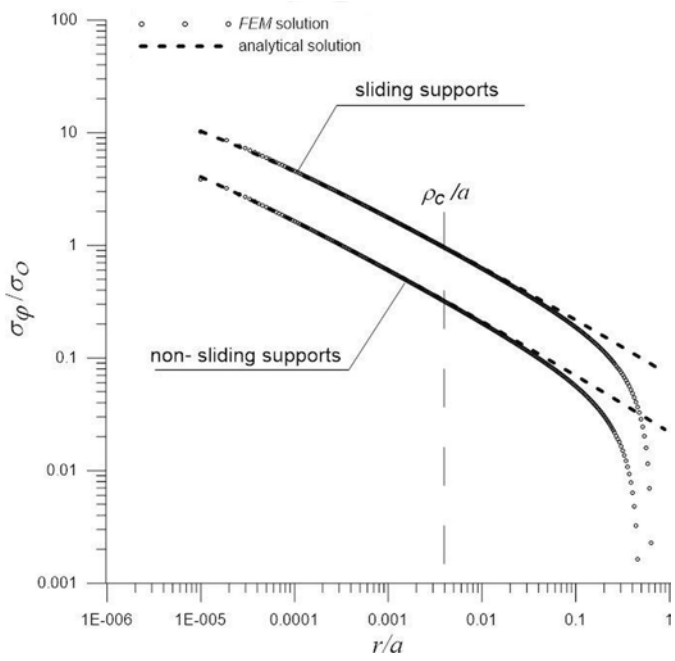


Fig. 8. The normalized circumferential stress versus r/a for PC/aluminum alloy 6061 specimen with notch-tip angle β= 300, $\sigma_0 = 3FL/\left(2gh^2\right)$, plane stress

also must be dependent on such factor. Accordingly, condition (3) can be written as follows (22):

$$\sigma_\varphi(\rho_c, 0) = \sigma_c(\psi) \tag{22}$$

Taking into account inter-relationships between cracking mechanics parameters, the following form of function depending on critical stress to mode mixity ratio ψ was proposed:

$$\sigma_c(\psi) = \sigma_c / \sqrt{1+\psi^2}, \tag{23}$$

where: $\psi = \dfrac{\tau_{r\varphi}(\rho_c, 0)}{\sigma_\varphi(\rho_c, 0)}$.

It is worth noting that the proposed modification of a criterion of cracking (23) is concurrent with a concept provided in paper [34]. Author of paper [34] suggests the use of an additional factor including impact of quantitative portion of shear and normal stresses in the process of prediction of crack initiation. This factor is not constant and depends on geometry of the object and load conditions.

By using criterion (23) critical force value can be estimated on the basis of dependence (24):

$$F_k = \frac{\sigma_c(\psi) F}{\sigma_\varphi(\rho_c, 0)}. \tag{24}$$

Critical force F_k (24) can be determined in two ways:
- using analytical (qualitative description (9, 10)) and numerical (quantitative description (15)) solutions;
- using only numerical solution (stresses obtained, e.g. using *FEM*).

Of course the latter, from practical point of view, is less complicated and recommended for use in engineering calculations.

However in this paper the former is used. The reason for that was the necessity to check to what extent solutions obtained from formulas (21) and (24) match each other. Because critical forces (21) were estimated using analytical and numerical solutions, the same approach should be used in formula (24). Thus using analytical and numerical solutions, critical forces F_k (24) for all analysed specimens were determined. The obtained results were exactly the same as results obtained on the basis of formula (21) and therefore there was no need to include them in the paper.

7. Summary and conclusions

This paper dealt with the analysis of interface crack initiation at sharp notches along adhesive bonding in bonded bi-material structure. The possibility of using McClintock criterion was analysed to predict the loads at the crack initiation for such type of construction elements. In order to use this criterion, usually, it is necessary to know the qualitative and quantitative description of stress fields around concentrator's tip. Thus, the analytical description was obtained, and methodology of its obtaining was presented. Moreover, the method for determining generalised stress intensity factors was discussed, including qualitative nature of singularity of stresses fields, and for the selected cases, their values were determined.

The obtained analytical and numerical solutions allowed for formulation of the form of damage criterion and critical parameters.

Two forms of cracking criterion were developed, based on:
- equivalent stress intensity factor;
- modification of McClintock criterion involving dependence of critical stresses on proportion of shear and normal stresses occurring in the cracking plane.

The carried out analyses showed that from both forms of the cracking criterion the same results of prediction of critical forces are obtained. However from the practical point of view the second form of the cracking criterion is more favourable. This is due to the fact that in the prediction of cracking stresses determined by using, e.g. FEM can be used only with no need of determining stress intensity factors.

Values of critical loads resulting from the hypothesis were compared with values obtained from the experiment. Since actual frictional and contact conditions that are present in the specimen support area cannot be reflected with numerical modelling of three-point bending test, a range in which the predicted critical forces occur, was determined only. In the vast majority of analysed cases critical forces determined experimentally occurred in the range defined by means of the tested criterion. What is more, tendency of variability of the predicted forces was consistent with experimental data. This suggests that the analysed criterion can be used to predict initiation of the process of cracking of elements bonded using an adhesive with a notch situated on the interface. However in order to state such fact clearly it would be necessary to carry out additional experimental tests. Such tests should be planned so that actual fixing and loading conditions for the specimens are reflected in numerical modelling. The author shall try to carry out such tests and verify suitability of McClintock's hypothesis again in his next paper.

Acknowledgement

Studies were carried out within work no. S/WM/3/17 and funded from under development of science by Ministry of Science and Higher Education.

References

1. Ayatollahi M, Torabi A R. A criterion for brittle fracture in U-notched components under mixed mode loading. Engineering Fracture Mechanics 2009; 39: 1883–1896, https://doi.org/10.1016/j.engfracmech.2009.04.008.
2. Baranowski P, Damaziak K, Małachowski J. Brake system studies using numerical methods, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2013; 15 (4): 337–342.
3. Bogy D B, Wang K C. Stress singularities at interface corners in bonded dissimilar isotropic elastic materials. International Journal of Solids and Structures 1971; 1: 993-1005, https://doi.org/10.1016/0020-7683(71)90077-1.
4. Byskov E. Calculation of stress intensity factors using finite element method with cracked elements. International Journal of Fracture Mechanics 1970; 6(2): 59-167, https://doi.org/10.1007/BF00189823.
5. Carpinteri A, Paggi M. Analytical study of the singularities arising at multi-material interfaces in 2D linear elastic problems. Engineering Fracture Mechanics 2007; 74: 59–74, https://doi.org/10.1016/j.engfracmech.2006.01.030.
6. Griffith A A. The phenomena of rupture and flow in solids. Philosophical Transactions series A 1920; 221: 163-198, https://doi.org/10.1098/rsta.1921.0006.
7. Kinloch A J. Adhesion and adhesives, Science and Technology. London: Springer, 1987, https://doi.org/10.1007/978-94-015-7764-9.
8. Kirsch G. Die theorie der elastizität und die bedürfnisse der festigkeitslehre. Verein deutscher Ingenieure Zeitschrift 1898; 29: 797-807.
9. Knesl Z, Klusak J, Nahlik L. Crack initiation criteria for singular stress concentrations, Part I: A Universal assessment of singular stress concentrations, Engineering Mechanics 2007; 14(6): 399–408.
10. Krishnan A, Xu L R. Systematic evaluation of bonding strengths and fracture toughnessess of adhesive joints. The Journal of Adhesion 2011; 87(1): 53–71, https://doi.org/10.1080/00218464.2011.538322.

11. Krishnan A, Xu LR. Experimental studies on the interaction among cracks, notches and interfaces of bonded polymers. International Journal of Solids and Structures 2013, 50: 1583–1596, https://doi.org/10.1016/j.ijsolstr.2013.01.024.
12. Krishnan A, Roy Xu L. An experimental study on the crack initiation from notches connected to interfaces of bonded bi-materials. Engineering Fracture Mechanics 2013; 111: 65–76, https://doi.org/10.1016/j.engfracmech.2013.08.010.
13. Leguillon D. A criterion for crack nucleation at a notch in homogeneous materials. Comptes Rendus de l'Académie des Sciences - Series IIB – Mechanics 2001; 329(2): 97–102, https://doi.org/10.1016/S1620-7742(01)01302-2.
14. Li Y, Song M. Method to calculate stress intensity factor of V-notch in bi-materials. Acta Mechanica Solida Sinica 2008; 21(4): 337–346, https://doi.org/10.1007/s10338-008-0840-3.
15. Łukaszewicz A. Nonlinear numerical model of heat generation in the rotary friction welding, Journal of Friction and Wear 2018; 39 (6): 612-619, https://doi.org/10.3103/S1068366618060089.
16. McClintock F A. Ductile fracture instability in shear. Journal of Applied Mechanics 1958; 25: 582-588.
17. Mieczkowski G. Description of stress fields and displacements at the tip of a rigid, flat inclusion located at interface using modified stress intensity factors. Mechanika 2015; 21(2): 91-98, https://doi.org/10.5755/j01.mech.21.2.8726.
18. Mieczkowski G. Stress fields and fracture prediction for adhesively bonded bi-material structure with sharp notch located on the interface. Mechanics of Composite Materials 2017; 53(3): 305-320, https://doi.org/10.1007/s11029-017-9663-y.
19. Mieczkowski G. Stress fields at the tip of a sharp inclusion on the interface. Mechanics of Composite Materials 2016; 52(5):601-610, https://doi.org/10.1007/s11029-016-9610-3.
20. Naik R A, Crews J H. Determination of stress intensity factors for interface cracks under mixed-mode loading. Paper presented at the ASTM National Symposium on Fracture Mechanics, June 30-July 2,1992, Gatlinburg, TN.
21. Parton V ., Perlin P I. Mathematical methods of the theory of elasticity. Moscow: Mir Publishers, 1984.
22. Pirondi A, Nicoletto G. Fatigue crack growth in bonded DCB specimens. Engineering Fracture Mechanics 2004; 71(4–6): 859–871, https://doi.org/10.1016/S0013-7944(03)00046-8.
23. Ritchie R O, Knott J F, Rice J R. On the relation between critical tensile stress and fracture toughness in mild steel. Journal of the Mechanics and Physics of Solids 1973; 21: 395-410, https://doi.org/10.1016/0022-5096(73)90008-2.
24. Rogowski G, Molski K L. The T-stress effect on the plastic zone size in a thin ductile material layer sandwiched between two elastic adherents. Engineering Fracture Mechanics 2016; 168 (A): 260-270.
25. Rudawska A, Dębski H. Experimental and numerical analysis of adhesively bonded aluminium alloy sheets joints. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2011; 1: 4–10.
26. Savruk M P, Shkarayev S, Madenci E. Stress near apex of dissimilar material with bilinear behavior. Journal of Applied Fracture Mechanics 1999; 31: 203-212, https://doi.org/10.1016/S0167-8442(99)00014-2.
27. Seweryn A, Łukaszewicz A. Verification of fracture criteria of elements with V-shaped notches. Eksploatacja i Niezawodnosc — Maintenance and Reliability 2001; 5: 6–8.
28. Seweryn A, Molski K. Elastic stress singularities and corresponding generalized stress intensity factors for angular corners under various boundary conditions. Engineering Fracture Mechanics 1996; 55: 529-556, https://doi.org/10.1016/S0013-7944(96)00035-5.
29. Sih G C. Strain-energy-density factor applied to mixed mode crack problems. International Journal of Fracture 1974; 10: 305-321, https://doi.org/10.1007/BF00035493.
30. Sih G C, Chen E P. Cracks in composite materials, Ch.3 (Mechanics of Fracture VI) ed. G. C. Sih. Hague: Martinus Nijhoff Publishers, 1981.
31. Sneddon I N. The distribution of stress in the neighbourhood of a crack in an elastic solid. Proceedings of the Royal Society of London A 1946;187(1009): 229-260, https://doi.org/10.1098/rspa.1946.0077.
32. Sun C T, Jih C J. On strain energy release rates for interfacial cracks in bi-material media. Engineering Fracture Mechanics 1987; 28: 13-20, https://doi.org/10.1016/0013-7944(87)90115-9.
33. Sun C T, Qian H. Brittle fracture beyond the stress intensity factor. Journal of Mechanics of Materials and Structures 2009; 4(4): 743-753, https://doi.org/10.2140/jomms.2009.4.743.
34. Taylor, D. The Theory of Critical Distances: A new perspective in fracture mechanics. Oxford: Elsevier, 2007.
35. Tracey D M. Finite elements for determination of crack tip elastic stress intensity factors. Engineering Fracture Mechanics 1971; 3(3): 255-265, https://doi.org/10.1016/0013-7944(71)90036-1.
36. Tran V-X, Leguillon D, Krishnan A. Interface crack initiation at V-notches along adhesive bonding in weakly bonded polymers subjected to mixed-mode loading. International Journal of Fracture Mechanics 2012; 176: 65–79, https://doi.org/10.1007/s10704-012-9727-x.
37. Treifi M, Oyadiji S O. Strain energy approach to compute stress intensity factors for isotropic homogeneous and bi-material V-notches. International Journal of Solids and Structures 2013; 50: 2196–2212, https://doi.org/10.1016/j.ijsolstr.2013.03.011.
38. Williams M L. Stress singularities resulting from various boundary conditions in angular corners of plate in extension. Journal of Applied Mechanics 1952; 9: 526-528

Grzegorz MIECZKOWSKI
Faculty of Mechanical Engineering
Bialystok University of Technology
ul. Wiejska 45C, Poland

E-mail: g.mieczkowski@pb.edu.pl

Article citation info:
AL-GARNI A, ABDELRAHMAN W, ABDALLAH A. ANN-based failure modeling of classes of aircraft engine components using radial basis functions. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 311–317, http://dx.doi.org/10.17531/ein.2019.2.16.

Ahmed AL-GARNI
Wael ABDELRAHMAN
Ayman ABDALLAH

ANN-BASED FAILURE MODELING OF CLASSES OF AIRCRAFT ENGINE COMPONENTS USING RADIAL BASIS FUNCTIONS

MODELOWANIE USZKODZEŃ ELEMENTÓW SILNIKA SAMOLOTOWEGO W OPARCIU O SZTUCZNE SIECI NEURONOWE O RADIALNYCH FUNKCJACH BAZOWYCH

The objective of this research is to present a model to predict failure of two categories of critical aircraft engine components; non-rotating components such as valves and gearboxes, and rotating components such as engine turbines. The work utilizes Weibull regression and artificial neural networks employing Back Propagation (BP) as well as Radial Basis Functions (RBF). The model utilizes training failure data collected from operators of turboprop aircraft working in harsh desert conditions, where sand erosion is a detrimental factor in reducing turbine life. Accordingly, the model is more suited for accurate prediction of life of critical components of such engines. The algorithm, which uses Radial Basis Function (RBF) NN, uses a closest point specifier. The activation is based on the deviation of the earlier prototype from the input vector. Two earlier models are used for comparison purposes; namely Weibull regression modeling and Feed-Forward BP network. Comparison results show that the failure times represented by RBF are in better compromise with actual failure data than both earlier modeling methods. Moreover, the technique has comparatively higher efficiency as the neuron's number in each layer of ANN is reduced, to decrease computation time, with minimum effect on the accuracy of results.

Keywords: *neural network, radial basis function, Reliability, engine components.*

Celem pracy jest przedstawienie modelu służącego do predykcji uszkodzeń dwóch kategorii krytycznych elementów silnika samolotowego: elementów nieobrotowych, takich jak zawory i skrzynie biegów oraz elementów obrotowych, takich jak turbiny silnika. W pracy wykorzystano regresję Weibulla i sztuczne sieci neuronowe oparte na propagacji wstecznej oraz radialnych funkcjach bazowych (RBF). Model wykorzystuje dane o błędach zebrane od operatorów samolotów turbośmigłowych pracujących w trudnych warunkach pustynnych, gdzie erozja powodowana przez piasek stanowi szkodliwy czynnik ograniczający żywotność turbin. Prezentowany model jest więc szczególnie przydatny do trafnego prognozowania żywotności krytycznych elementów takich silników. Algorytm, który wykorzystuje sieci neuronowe o radialnych funkcjach bazowych, używa specyfikatora najbliższego punktu. Aktywacja bazuje na odchyleniu wcześniejszego prototypu od wektora wejściowego. Dwa wcześniejsze modele oparte na regresji Weibulla (Weibull regression modeling) oraz sieciach typu Feed-Forward Backpropagation wykorzystano do badań porównawczych. Wyniki porównania pokazują, że czasy uszkodzeń odwzorowane przez RBF pozostają w większej zgodzie z rzeczywistymi danymi o uszkodzeniach niż w przypadku obu wcześniejszych metod modelowania. Co więcej, technika ta ma porównywalnie większą efektywność, ponieważ liczba neuronów w każdej warstwie sieci neuronowej została zredukowana tak aby zmniejszyć czas obliczeń, przy minimalnym wpływie na dokładność wyników.

Słowa kluczowe: *sieć neuronowa, radialna funkcja bazowa, niezawodność, elementy silnika.*

1. Introduction

Reliability of modern aircraft engines is highly affected by failure prediction methods of their critical parts. When these engines are operated in severe desert conditions, the high temperature, pressure, and velocity of the intake air may increase the effect of sand erosion on the rotating engine critical components, such as high-pressure turbine blades, and gearboxes. Sand also contributes to blockage of failure of nonrotating parts like engine valves that get blocked by sand particles. To better schedule cost-effective preventive maintenance, more accurate modeling and prediction of component life is important. This also helps to enhance both aircraft safety and reliability.

The type of engine system which is currently under consideration comprises a 4-stage turbine subjected to extreme high pressure and temperature as it draws air energy from the combustion chamber. The maximum Turbine Inlet Temperature (TIT) can reach up to 1077ºC while it has a maximum generated power of 11,000 hp. Both turboprop and turbofan engines contain also no rotating parts that are critical to engine operations. These include turbine bleed air valves, high pressure valves, stator vanes, casing and support, support for rear bearing, and thermocouples. The failure of any of these components can lead to the failure of the entire turbine system. Hence, it is important to address this type of failure for aircraft safety and this research present a failure model for the parts that were found to fail the most in engines operated in the desert climate of Saudi Arabia. The work will proceed to compare predictions made by Weibull regression and Artificial neural networks algorithms that utilize either Back Propagation or Radial Basis Functions with the actual failure data from available data from the aviation industry.

2. Literature Review

There is a large body of research on modeling of reliability of aircraft system components for different types of failure. Most of this research was based on a Weibull distribution to predict the component life, utilizing discrete-stressed volume approach [21]. Later, Sheikh et al. [16] and Al-Garni et al. [2] used Feed-Forward, Back-Propagation ANN to better model failure, as compared to the more common two-parameter and three-parameter This advantage of FF-BP models was extended to both component and system levels [3, 4, 5, 18]. The work enabled customizing maintenance plans to suit different operator conditions. Later, Qattan [15] used also Radial Basis Functions for modeling of failure time of rotating components.

Artificial Neural Networks were introduced as a means of modeling. A great body of literature was dedicated to explaining these methods, see for example Kutusurelis [10] and Paul et al. [14]. The basic convergence criterion, known as the delta rule, in all of these algorithms is based on reducing some mean square error to its least value. The convergence criterion is generally utilized for single layer networks and also forms the foundation for a back propagation used for multi-layer networks. Additionally, Hopfield [9] used a "spin glass" model for storing data in dynamically stable type of networks. By that time, the well-established back propagation network used a basic delta rule to propagate the output error back to the hidden units, see for example Parker [13]. In his work, he used BP to train a multilayer feed forward network.

Another type of net, known as Radial Based Functions (RBF), was formulated by Broomhead and Lowe [8]. In RBF networks, the accuracy of modeling depends mainly on the number of neurons in the input and output layers. Most of the research work aimed at selecting the most suitable structure and parameter values of the network that increase the failure prediction reliability. Several variations of PB and RBF networks were used to study failure of aircraft components with different training methods, including deep belief networks of Lin et al. [11].

In a more recent work, Abdelrahman et al. [1] investigated the training of BP ANN to model failure of engine components. The work, which was expanded by Al-Wadiee [6], showed significant improvement over the traditional Weibull model results. Both work, however, did not deal with utilization of the predicted results.

Later, Tian [17] used ANN based method to predict the remaining life of equipment that are fitted with health monitoring. The monitoring of age and multiple condition was used to train the NN. Also, Bin et al. [7] utilized two decomposition methods to obtain the fault feature frequency in rotating machinery. Then, they used an artificial neural network for the early fault diagnosis of these components. The fault pattern of the machinery was identified using a three layers BP NN model taking the fault feature frequency as the target input. The spectral bandwidth energy of vibration signal spectrum was taken as a characteristic parameter, whereas 10 types of representative rotor fault were considered the output.

Wang and Wang [20] used a nested extreme response surface to tackle time dependency issues in reliability analysis. In their work, Nieto et al. [12] combined support vector machines with particle swarm optimization technique. This significantly improved regression accuracy. Vanini et al. [19] also used multiple dynamic neural networks to learn the nonlinear dynamics of jet engines. Residuals obtained by comparing operating modes of healthy and faulty conditions of engines were used to develop a reliable criterion for detecting and isolating faults.

3. Modeling of component failure

As this work aims at failure modeling of aircraft critical components using Weibull and ANN (Back propagation/Radial Basis Function). Since the Weibull regression method is well documented in literature [21], more emphasis is given in the present work to the back propagation and Radial Basis Function ANN modeling.

3.1. Back propagation artificial neural network

The back propagation ANN modeling utilizes some criterion function to maximize the performance by continuously varying the network weights through gradient descent algorithm. The goal of this model is to train the network in such a manner that a compromise can be achieved between the ability to respond accurately after receiving training input patterns and the capability to produce accurate responses for similar inputs. The network has two main segments; the forward-feed and the back-propagation. The input signals are transformed into output signals using a suitable activation function. Log-sigmoid function is an activation function that proved to give good results for component failure time modeling. The algorithm is detailed in works of [8] and [1], and is only summarized here.

Let m and n be the number of inputs and outputs of the network; respectively, and let X_d be the input signal to the ANN, whereas $f(net_k)$ is the activation function. This function is normally a nonlinear log-sigmoid function, whose parameter values depend on the preferred output data range. Let N be the number of intermediate neurons. In this case, we can define a normalized set of inputs as follows:

$$x_j = \text{normalized values of } X_d \ , 1 < d \leq m \tag{1}$$

Here the rank of neuron activation is defined as:

$$x_k = f(\text{net}_k), \quad m < k \leq N + n \tag{2}$$

where:

$$\text{net}_k = \sum_{j=1}^{k-1} W_{kj} x_j \ , \quad m \leq k \leq N + n \tag{3}$$

Here, W_{kj} are the elements of the weight matrix. The number of neurons in the corresponding neighboring layers determines the size of this weight matrix W.

The activation log-sigmoid function is given by:

$$f(net_k) = \frac{1}{1 + e^{-netk}} \tag{4}$$

The network output is calculated from the output layer neuron signals as:

$$O_s = X_{N+s} \ \ 1 \leq s \leq n \tag{5}$$

In the above equations, N is always less than the number of inputs of the ANN. Moreover, If we consider the inputs as neurons of the input layer, then the sum $(N + m)$ is the total number of neurons in the network. Figure 1 shows a typical 3 layer ANN model with inputs p_R and weights $w^1_{s,R}$.

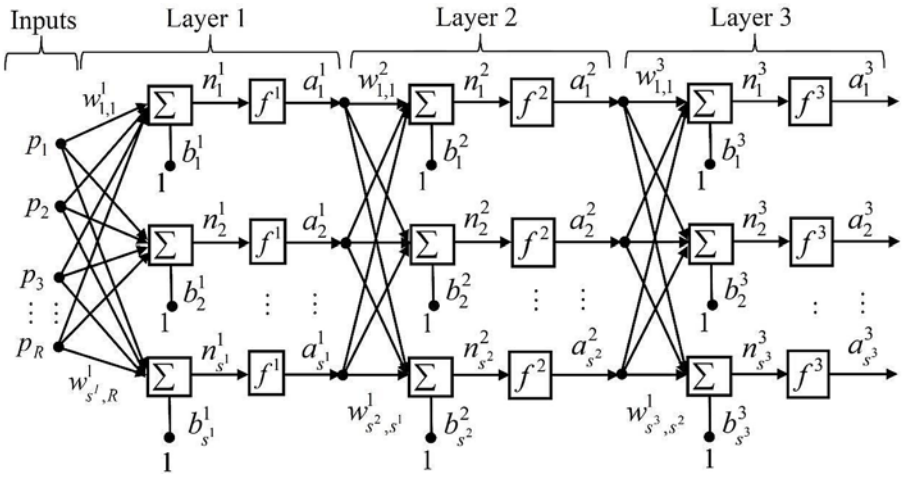


Fig. 1. A schematic of multilayer ANN with a general activation function

3.2. Radial basis function artificial neural network

In RBF neural network modeling, the hidden unit activation is based upon the spacing between the input prototype and the input vector. This RBF model employs a closest point classifier type to detect anomalies. An extended version of the model utilizes nominal data. Here, as soon as the data is added to the system, it is compared to the RBF model, and is classified as nominal if it falls within the boundaries of the model. If it does not fall within these boundaries, then it is classified as anomalous. This approach has many applications in complex systems and subsystems, including those of military aircraft. The accuracy of the model in this case depends on the number of elements in each layer, as well as the appropriate choice of the radial basis function. The selection of the radial basis function for the detection of an anomaly is done considering the condition:

$$f(x) = exp\left(-\frac{1}{\theta_\alpha}\sum_k |x_k - \mu|^\alpha\right). \tag{6}$$

where $\alpha \in (0,\infty)$ θ_α is the α central moment, given by:

$$\theta_\alpha = \sum_{k=1}^{N} |x_k - \mu|^\alpha$$

and μ is the center of the data set defined as:

$$\mu = \frac{1}{N}\sum_{k=1}^{N} x_k \,. \tag{7-b}$$

The mean μ and the central moment θ_α as given by equations (7), are evaluated for the nominal condition from the data of a sampled time series. In this case, the distance from the center μ to any vector x is given by:

$$x - \mu_{l_\alpha} = \left(\sum_{k=1}^{N} |x_k - \mu|^\alpha\right)^{1/\alpha}. \tag{8}$$

Therefore, the radial basis function is defined by $f_{nom} = f(x)$ at the nominal condition. The mean and central moment are not changed for the anomalous conditions, whereas the radial basis function is calculated using the possibly anomalous data set under the condition at a slow time scale. Anomaly can be measured at the k epoch by the distance from the neuron's center, given by:

$$M_k \equiv d(f_{nom}, fk) \tag{9}$$

This distance is inversely proportional to the outputs of the Gaussian transfer functions of the hidden layer. Fig. 2 depicts the general architecture of a RBF network. The network consists of three layers. The first layer has one neuron for each predictor variable. It is pivotal to normalize the range of the input values prior to the input layer. The second layer, the hidden layers, receive the resulting values from the first input layer. Since the neuron's number in this layer varies, an optimum value is obtained through the process of training. In this layer, every neuron is modeled as RBF which is centered on a point whose dimensions are the same as the number of the predictor variables. The RBF function's radius represents the spread which may vary for each dimension. The spreads and the centers are optimally evaluated again in this process of training.

The Euclidean distance is evaluated from the neuron's center point after a hidden neuron is given some input values in the form of vector x from the input layer. Next, for this distance, the model uses RBF type kernel function by utilizing the spread values that were calculated in the previous step. The output of the value is then transferred to the summation layer which is also the last layer. In the summation layer, first the neuron value produced by the hidden layer is multiplied by a weight (W_i) which is related to the neuron and afterwards the output of the network is obtained by summation of all the weighted values. Accordingly, training is carried out to determine the neuron's number in the hidden layer, the RBF function's radius in each dimension, the center's coordinates for each hidden-layer RBF function, and the weights that are applied to the outputs of the RBF function as they are transferred to the last layer (summation layer).

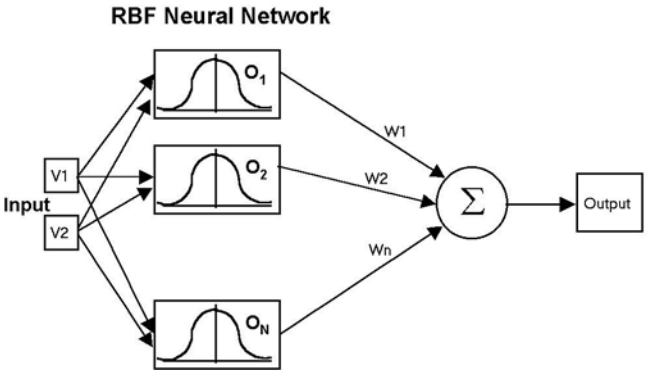


Fig. 2. Model for the RBF Network

The error function is calculated as:

$$\text{error} = \Sigma\left[F(t) - O(t)\right]^2 \tag{10}$$

4. Results and discussion

Actual failure data over a period of thirty years was provided by local aviation operator working within the Arabian Gulf Area contained both the Time Since Overhaul (T.S.O.). A subset of this data is used to train the ANN network with 2, 4, and 1 neurons in the in-

put, hidden, and output layers; respectively. The prediction results are compared to Weibull predictions, and to the actual data in Fig.3. Fig. 4 presents a similar comparison when the ANN model utilizes 4, 20, and 1 neurons in the three layers.

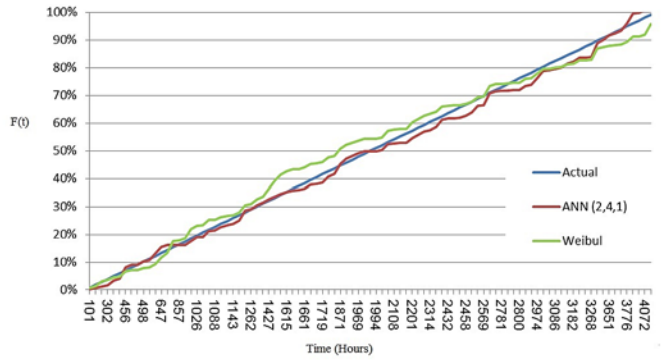


Fig. 3. Comparison of results of ann (2, 4, 1), Weibull model, and actual failure rate for turbine requiring overhaul maintenance (T.S.O)

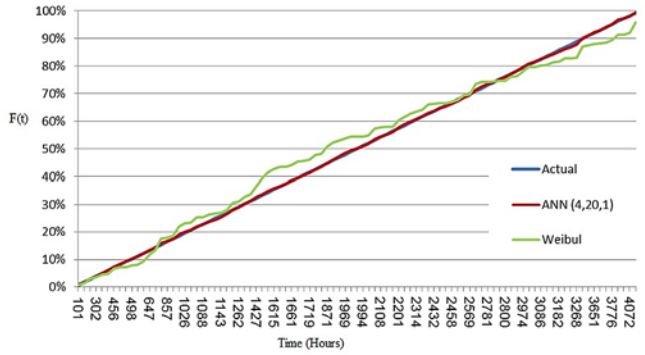


Fig. 4. Comparison of results of ANN (4, 20, 1), Weibull model, and actual failure rate for turbine requiring overhaul maintenance (T.S.O)

It can be noted from Figures 3 and 4 that the network training significantly improves with this change in the network configuration. It was also found that there is only negligible effect of changing other parameters such as learning rate and momentum constant.

A comparison was carried out between the maximum error percent in predictions using several ANN-BP configurations and the results of these configurations are summarized below in Table 1. The error in this table is calculated based on comparison with the actual collected field data. Again, it is clear from the table that ANN (4,20,1) configuration has the least mean error among all the other ANN-BP configurations and hence this configuration most closely matches with the actual data. Further increase in neuron numbers in any of the layers does not significantly reduce error, but increases run-time considerably.

Table 1. Percentage error for failure of Turbine requiring overhaul maintenance (T.S.O.)

Network Configuration	Mean Error (%)
BP - ANN (4, 20, 1)	0.84
BP - ANN (4, 10, 1)	1.00
BP - ANN (4, 8, 1)	1.51
BP - ANN (3, 6, 1)	4.51
BP - ANN (2, 4, 1)	6.85

Table 2. Mean percentage error for failure rate of turbine requiring overhaul maintenance (T.S.O) for Weibull, BP ANN (4, 20, 1), RBF ANN

Method	Mean Error (%)
Weibull	16.55
ANN- BP (4,20,1)	0.84
ANN-RBF (best config.)	1.09E-15

It is worth noting that a similar approach can be implemented for predicting turbine failure requiring overhaul maintenance (T.S.O), which is sometimes used in the aviation industry. Table 2 presents the MATLAB results for comparison between BP ANN (4, 20, 1), Weibull regression model, and RBF neural network model for this case. It is clear from the table that there is negligible mean percentage error of around 1.09E-15 % when Radial based function neural network was used for prediction of failure rate for turbines that requires overhaul maintenance. Hence, the comparison in Table 2 clearly establishes that ANN-RBF model presents more accurate modeling of turbine failure rate than both Weibull regression and ANN-BP models.

Figure 5 presents a comparison of the three methods' results over a range of more than 4,000 hours. From this figure, it can be deduced from figure that it is more advantageous to use RBF neural network for predicting failure rate as compared to Weibull or BP ANN model. Moreover, the BP ANN (4, 20, 1) configuration also depicts far better results than Weibull regression model.

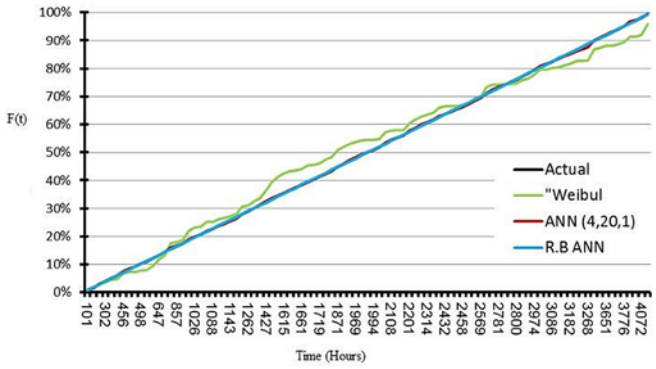


Fig. 5. Comparison of BP ANN model, RBF ANN and Weibull regression with actual failure rate for turbine requiring overhaul maintenance (T.S.O)

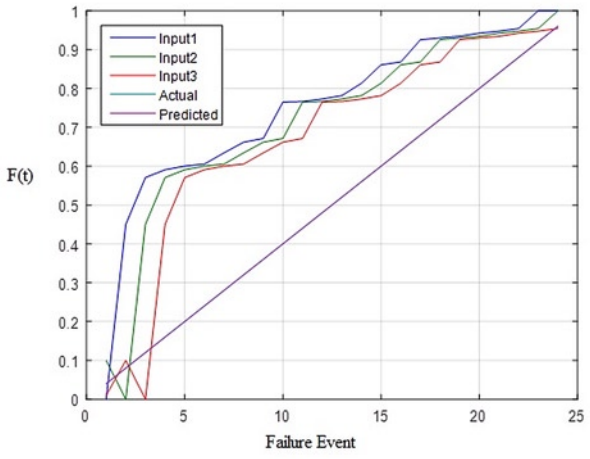


Fig. 6. RBF ANN analysis of Gear boxes failure training data without denormalization

Table 3. General failure data for gear boxes using Weibull analysis- Part I

t_i	$t_i - t_0$ $t_0 = 506.6$	i	$F = i/N + 1$ $N = 44$	$R = 1 - F$	$\ln (t_i - t_0) = x_i$	$\ln[\ln(1/R)] = y_i$
4838.6	4332.00	1	0.022222	0.977778	8.373785	-3.795447
5183.7	4677.10	2	0.044444	0.955556	8.450434	-3.090870
5246.5	4739.90	3	0.066667	0.933333	8.463771	-2.673752
5345.2	4838.60	4	0.088889	0.911111	8.484381	-2.374184
5437.9	4931.30	5	0.111111	0.888889	8.503358	-2.138911
5663	5156.40	6	0.133333	0.866667	8.547994	-1.944206
5680.6	5174.00	7	0.155556	0.844444	8.551401	-1.777405
5753.2	5246.60	8	0.177778	0.822222	8.565336	-1.630945
5773.3	5266.70	9	0.200000	0.800000	8.569159	-1.499940
5868	5361.40	10	0.222222	0.777778	8.586980	-1.381050
5900	5393.40	11	0.244444	0.755556	8.592931	-1.271888
5926.2	5419.60	12	0.266667	0.733333	8.597777	-1.170683
6031.5	5524.90	13	0.288889	0.711111	8.617020	-1.076088
6752.5	6245.90	14	0.311111	0.688889	8.739681	-0.987048
6915	6408.40	15	0.333333	0.666667	8.765365	-0.902720
7385.9	6879.30	16	0.355556	0.644444	8.836272	-0.822421
7395	6888.40	17	0.377778	0.622222	8.837594	-0.745582
7555	7048.40	18	0.400000	0.600000	8.860556	-0.671727
8173.2	7666.60	19	0.422222	0.577778	8.944629	-0.600448
8293	7786.40	20	0.444444	0.555556	8.960134	-0.531391
8540.9	8034.30	21	0.466667	0.533333	8.991475	-0.464246
8586.3	8079.70	22	0.488889	0.511111	8.997110	-0.398735
8953	8446.40	23	0.511111	0.488889	9.041496	-0.334606
9046.1	8539.50	24	0.533333	0.466667	9.052458	-0.271625
9523	9016.40	25	0.555556	0.444444	9.106800	-0.209573
9542.9	9036.30	26	0.577778	0.422222	9.109005	-0.148241
9542.9	9036.30	27	0.600000	0.400000	9.109005	-0.087422
9684.41	9177.80	28	0.622222	0.377778	9.124543	-0.026910
9777.5	9270.90	29	0.644444	0.355556	9.134636	0.033506
9786.3	9279.70	30	0.666667	0.333333	9.135584	0.094048
9848	9341.40	31	0.688889	0.311111	9.142211	0.154955
9925	9418.40	32	0.711111	0.288889	9.150421	0.216492
9946.7	9440.10	33	0.733333	0.266667	9.152722	0.278961
10249	9742.60	34	0.755556	0.244444	9.184263	0.342715
10563	10056.20	36	0.800000	0.200000	9.215945	0.475885
11044	10537.40	37	0.822222	0.177778	9.262686	0.546514
11462	10954.90	38	0.844444	0.155556	9.301542	0.620981
11545.8	11039.20	39	0.866667	0.133333	9.309208	0.700571
11674	11167.00	40	0.888889	0.111111	9.320718	0.787195
11714	11207.30	41	0.911111	0.088889	9.324321	0.883920
12376	11869.10	42	0.933333	0.066667	9.381694	0.996229

A question that need to be investigated is whether these conclusions can be extended to failure prediction of nonrotating components, or components not directly affected by the airflow. Accordingly, a similar study was carried out for gearbox failure data. The failure data for gear boxes was also collected from local aviation company. Tables 3 and 4 present Weibull analysis for Gear box failure for two sets of Weibull parameters, comprising failure data of 42 gearboxes.

Afterwards, RBF-ANN analysis was carried out, with different neurons numbers, and is compared to Weibull analysis results. This is presented in Figure 6. In this figure Input 1, 2, and 3 refer to the

Table 4. General failure data for gear boxes using Weibull analysis – Part II

(1)	(2)	(3)	(4)	(5)	(6)	(7)
t_i	$t_i - t_0$	i	F = i/N + 1	R = 1- F	ln ($t_i - t_0$)	ln[ln(1/R)]
9046.1	8539.50	24	0.533333	0.466667	9.052458	-0.271625
9523	9016.40	25	0.555556	0.444444	9.106800	-0.209573
9542.9	9036.30	26	0.577778	0.422222	9.109005	-0.148241
9542.9	9036.30	27	0.600000	0.400000	9.109005	-0.087422
9684.41	9177.80	28	0.622222	0.377778	9.124543	-0.026910
9777.5	9270.90	29	0.644444	0.355556	9.134636	0.033506
9786.3	9279.70	30	0.666667	0.333333	9.135584	0.094048
9848	9341.40	31	0.688889	0.311111	9.142211	0.154955
9925	9418.40	32	0.711111	0.288889	9.150421	0.216492
9946.7	9440.10	33	0.733333	0.266667	9.152722	0.278961
10249	9742.60	34	0.755556	0.244444	9.184263	0.342715
10331	9824.50	35	0.777778	0.222222	9.192635	0.408180
10563	10056.20	36	0.800000	0.200000	9.215945	0.475885
11044	10537.40	37	0.822222	0.177778	9.262686	0.546514
11462	10954.90	38	0.844444	0.155556	9.301542	0.620981
11545.8	11039.20	39	0.866667	0.133333	9.309208	0.700571
11674	11167.00	40	0.888889	0.111111	9.320718	0.787195
11714	11207.30	41	0.911111	0.088889	9.324321	0.883920
12376	11869.10	42	0.933333	0.066667	9.381694	0.996229

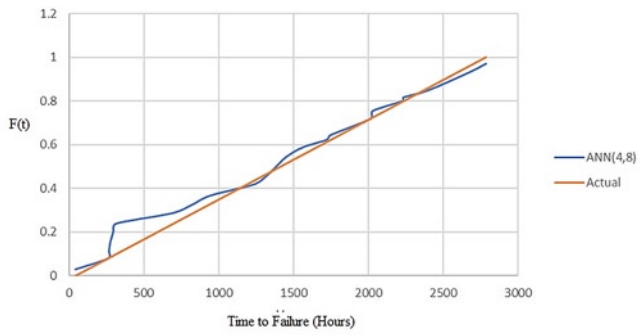


Fig. 7. Comparison of failure prediction results for RBF ANN with the actual data of bleed valves.

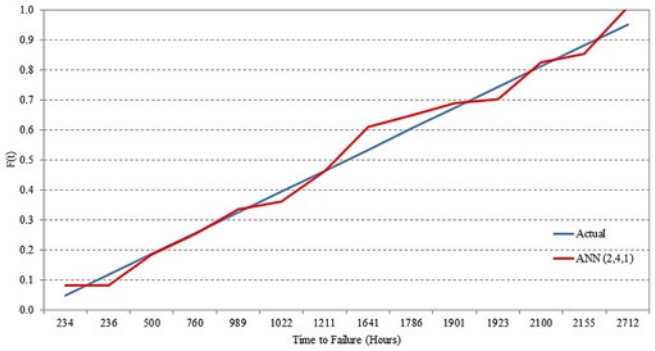


Fig. 8. Comparison of results of ANN (2, 4, 1) for High stage valve with the actual data

different ANN configurations, with increasing number of neurons in the input and hidden layers. It can be seen from the figure that increasing number of neurons results in more accurate predictions, but that prediction accuracy for gearbox failures is less than that of turbine failures. Figure 6: RBF ANN analysis of Gear boxes failure training data without denormalization

Finally, the failure prediction of nonrotating components; namely engine valves was carried out. Collected failure data of high pressure valves and engine bleed valves was used to train the ANN. Calculations of failure prediction of bleed valves are presented in Table 5, whereas calculations for the high pressure valves are presented in Table 6. For the bleed valves, Table 5, a network configuration of 4 neurons in the input layer and 8 in the hidden layer was selected. However, for the case of high pressure valves, Table 6, the results of several network configurations were presented and compared.

Figures 7, 8, and 9 illustrate the comparison between selected network configuration and actual results for bleed valves (Figure 7), and for high pressure valves (Figure 8 and 9). From these figures it is clear that prediction results are better than predictions done for gearboxes, and the accuracy is comparable to predictions done for turbine components.

5. Conclusion

In this work, a RBF NN model is developed for the modeling of critical rotating and nonrotating engine parts operating in desert conditions. The accuracy of the trained model is determined by a comparison to earlier methods of BP NN and Weibull distribution. Testing several values for the number of neurons shows that the percentage deviation from actual failure data is small for networks with 4 or more neurons in the input layer. Further increase in number of neurons does not improve accuracy much and considerably increases run-time. Moreover, it is found that an intermediate neuron number of 20 produces reasonable accuracy for predicting failure of both rotating and nonrotating components, without much increase in run time. The RBF modeling of failure data provided by local operator proved to be superior than the other

two methods. RBF NN showed much higher accuracy, as represented by the much smaller value of the sum of error squares. This modeling, therefore, can be utilized to formulate replacement and overhaul guidelines of both rotating and nonrotating engine components that correspond to a certain optimum level of reliability.

Acknowledgement

This paper contains the studies and results of a research work generously funded by King Abdul-Aziz City of Science and Technology, through Project No. AT-35-106.

References

1 Abdelrahman W G, Al-Garni A Z, Al-Wadiee W. Application of back propagation neural network algorithms on modeling failure of B-737 bleed air system valves in desert conditions. Applied Mechanics and Materials 2012; 225: 505-510, https://doi.org/10.4028/www.scientific.net/AMM.225.505.
2. Al-Garni A Z, Ahmed S, Siddiqui M. Modeling failure rate for Fokker F-27 tires using neural network. Transactions – Japan Society for Aeronautical and Space Sciences 1998; 41: 29-37.
3. Al-Garni A Z, Jamal A, Ahmad A, Al-Garni A, Tozan M. Failure-rate prediction for De Havilland Dash-8 tires employing neural network technique. AIAA Journal of Aircraft 2006; 43(2): 537-543, https://doi.org/10.2514/1.16609.
4. Al-Garni A Z, Tozan M, Al-Garni A, Jamal A. Failure forecasting of aircraft air conditioning/cooling pack with field data. Journal of Aircraft 2007; 44(3): 996-1002, https://doi.org/10.2514/1.26561.
5. Al-Qutub A, Al-Garni A Z. Comparison between neural network and Weibull models for failure of Boeing 737 engines. Transactions of the Japan Society for Aeronautical and Space Sciences 1999; 42(137):128-134.
6. Al-Wadiee W. Back propagation neural network algorithms on modeling failure of B-737 bleed air system valves in desert conditions. MS Thesis: King Fahd University of Petroleum and Minerals, 2011.
7. Bin G, Gao J, Li X, Dhillon B. Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network. Mechanical Systems and Signal Processing 2012; 27: 696-711, https://doi.org/10.1016/j.ymssp.2011.08.002.
8. Broomhead D, Lowe D. Multivariable function interpolation and adaptive networks. Computer Systems 1988; 2: 321–355.
9. Hopfield J. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences 1982; 79(8): 2554-2558, https://doi.org/10.1073/pnas.79.8.2554.
10. Kutsurelis J. Forecasting financial markets using neural networks: An analysis of methods and accuracy. PhD dissertation: Naval Postgraduate School, 1998.
11. Lin X S, Li B W, Yang X. Engine components fault diagnosis using an improved method of deep belief networks. 7th International Conference on Mechanical and Aerospace Engineering (ICMAE) 2016; 18-20 July 2016, https://doi.org/10.1109/ICMAE.2016.7549583.
12. Nieto P G, García-Gonzalo E, Lasheras F S, de Cos Juez, F J. Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability. Reliability Engineering & System Safety 2015; 138: 219-231, https://doi.org/10.1016/j.ress.2015.02.001.
13. Parker D. Learning logic. Technical Report TR-87. Cambridge, MA: Center for Computational Research in Economics and Management Science MIT, 1985.
14. Paul S, Kapoor K, Jasani, D, Dudhwewala R, Gowda V B, Nair T R. Application of artificial neural networks in aircraft maintenance, repair and overhaul solutions. arXiv preprint arXiv:1001.3741, 2010.
15. Qattan N A. Reliability analysis of C-130 turboprop engine components using artificial neural network. PhD dissertation: King Fahd University of Petroleum and Minerals, 2013.
16. Sheikh A K, Al-Garni, A Z, Affan Badar M. Reliability analysis of aeroplane tyres. International Journal of Quality & Reliability Management 1996; 13(8): 28-38, https://doi.org/10.1108/02656719610128475.
17. Tian Z. An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring. Journal of Intelligent Manufacturing 2012; 23(2): 227-237, https://doi.org/10.1007/s10845-009-0356-9.
18. Tozan M, Al-Garni A Z, Jamal A. Failure Distribution Modeling for Planned Replacement of Aircraft Auxiliary Power Unit Oil Pumps. Maintenance Journal 2006; 19(1): 60-69.
19. Vanini Z S, Khorasani K, Meskin N. Fault detection and isolation of a dual spool gas turbine engine using dynamic neural networks and multiple model approach. Information Sciences 2014; 259: 234-251, https://doi.org/10.1016/j.ins.2013.05.032.
20. Wang Z, Wang P. A new approach for reliability analysis with time-variant performance characteristics. Reliability Engineering & System Safety 2013; 115:70-81, https://doi.org/10.1016/j.ress.2013.02.017.
21. Zaretsky E Y. Fatigue criterion to system design, life, and reliability. Journal of Propulsion and Power 1987; 3(1): 76-83, https://doi.org/10.2514/3.22955.

Ahmed AL-GARNI
Wael ABDELRAHMAN
Ayman ABDALLAH
Department of Aerospace Engineering
King Fahd University of Petroleum and Minerals
Dhahran, 31261, Saudi Arabia

E-mails: algarni@kfupm.edu.sa, wgamal@kfupm.edu.sa, aymanma@kfupm.edu.sa

Article citation info:
TAN CM, NARULA U, LAI LA, PANDEY S, TUNG JH, LI CY. Optimal maintenance strategy on medical instruments used for haemodialysis process. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 318–328, http://dx.doi.org/10.17531/ein.2019.2.17.

Cher Ming TAN
Udit NARULA
Lu An LAI
Sumit PANDEY
Jung Hua TUNG
Chung Yi LI

OPTIMAL MAINTENANCE STRATEGY ON MEDICAL INSTRUMENTS USED FOR HAEMODIALYSIS PROCESS

OPTYMALNA STRATEGIA KONSERWACJI URZĄDZEŃ MEDYCZNYCH WYKORZYSTYWANYCH W PROCESIE HEMODIALIZY

Haemodialysis machines are one of the important medical equipment which is used to treat renal failures and minimum downtimes are thus essential. Uninterrupted and constant use of these machines in hospitals worldwide makes them vulnerable to failures if not maintained properly. Consequently, the maintenance cost for dialysis machine is high. A method to implement a cost effective maintenance strategy is demonstrated in this work. Root Cause Based Maintenance (RCBM) strategy is employed at the component level to optimize the Reliability Based Maintenance schedules derived from the existing maintenance and failure data. In order to minimize the average cost of maintenance for Haemodialysis machines and ensure their high operational availability, a Cost-Model is derived, and Genetic Algorithm is employed for optimization in this work. The application of RCBM strategy results in cost saving of about 60% of the cost incurred using current maintenance scheme. Statistical and optimization calculations are performed using Reliasoft's Weibull++ and MATLAB tools respectively.

Keywords: *reliability-based maintenance, dialysis machines, genetic algorithm, reliability analysis, statistical distributions, operational availability.*

Aparaty do hemodializy to ważne urządzenia medyczne wykorzystywane w leczeniu niewydolności nerek, dlatego ich przestoje muszą być jak najkrótsze. Ciągłe, nieprzerwane korzystanie z tych urządzeń w szpitalach na całym świecie sprawia, że, w przypadku braku właściwej konserwacji, są one podatne na awarie. W związku z tym koszty konserwacji aparatów do dializy są wysokie. W prezentowanej pracy przedstawiono metodę wdrażania ekonomicznej strategii konserwacji. Wykorzystano strategię konserwacji opartą na analizie przyczyn źródłowych uszkodzenia (RCBM). Zastosowano ją na poziomie części składowych w celu optymalizacji harmonogramów konserwacji opartej na niezawodności (RBM) tworzonych na podstawie istniejących danych dotyczących konserwacji i uszkodzeń. Aby móc zminimalizować średni koszt konserwacji aparatów do hemodializy i zapewnić ich wysoką gotowość operacyjną, opracowano model kosztowy, a optymalizację przeprowadzono za pomocą algorytmu genetycznego. Zastosowanie strategii RCBM daje około 60-procentową oszczędność kosztów, jakie ponosi się przy użyciu obecnie wykorzystywanego programu konserwacji. Obliczenia statystyczne i optymalizacyjne wykonano, odpowiednio, przy użyciu oprogramowania Weibull ++ i MATLAB firmy Reliasoft.

Słowa kluczowe: *konserwacja oparta na niezawodności, aparaty do dializy, algorytm genetyczny, analiza niezawodności, rozkłady statystyczne, gotowość operacyjna.*

Notations Used:

η	Scale parameter, or characteristic life (Weibull Distribution)
β	Shape parameter (or slope) (Weibull Distribution)
γ	Location parameter (or failure free life
z	$\ln(t)-\mu$
e^{μ}	Scale parameter (Gamma Distribution)
k	Shape parameter (Gamma Distribution)
t'	$\ln(t)\cdot$t values are the times-to-failure
μ'	Mean of the natural logarithms of the times-to-failure
σ'	Standard deviation of the natural logarithms of times-of-failure
P_i	Price of component i
$T_{r,i}$	Time of replacing work of component i
$T_{fc,i}$	Time of function check and rinse
C_{mp}	Manpower cost per hour (250 *NTD* / *hour*)
T_s	Time interval that we want to calculate for the cost
t_i	Replacement period of component i
k	Number of component
C_{Ri}	The cost of replacing component i^{th} time

t_{tk} Regular maintenance period (220 days)

P_{tk} Price of tool kit(2600ntd)

$\left(T_{r,tk}+T_{fc,tk}\right)$ Time of repairing work plus function check and rinse (2 hours).

C_{mp} Manpower cost per hour (250 $NTD/hour$)

$C_{ex,i}$ Expected cost of component i

$C_{uex,i}$ Unexpected cost of component i

C_{tk} Cost of tool kit maintenance

C_{rp} The repeat calculated function check and rinse cost that occur when more than one component was repaired in a repair event.

λ Failure rate.

Abbreviations Used:

RV	Relief Valve
TMP	Trans Membrane Pressure
NTB	Network Terminal Box
R Chamber	Rise Chamber
BLD	Blood Leak Detector
CCB	Carbon Cleaning Brush
PdM	Predictive Dialysis Maintenance
V39	Valve 39

1. Introduction

Medical technology has witnessed rapid strides in the modern world which is a result of sophisticated and yet fascinating advancement in the medical methods and procedures. Highly advanced medical equipment has become an indispensable tool for modern medical procedures by virtue of which, most of the diagnosis and treatment are highly dependent on this equipment. The annual revenue for medical technology industry is worth half a trillion US dollars [13, 33] which shows the extensive use of medical equipment in modern medical facilities. Amongst few life-supporting medical equipment which are used in quick succession in hospitals and clinics, Haemodialysis is an important machine which works as an artificial kidney. In critical situations like renal failures, it is this Haemodialysis machine that removes the toxic waste products and restores the normal levels of body fluid volume and composition [3,18]. Hence, the availability requirement of this machine is high in order to treat the patients when need arises, and thus the number of such machine in a given hospital tend to be high.

However, the hospitals and medical facilities all around the world are presented with a challenge to maintain the Haemodialysis machine effectively, given the fact that it is being used continuously, one patient after another. The process of maintenance must be performed efficiently and effectively because improper maintenance and repairs can lead to unsafe conditions and reduced system performance and availability. Moreover, the cost-factor makes the maintenance strategy challenging because it is believed that 1/3rd of the overall maintenance cost is wasted either due to unnecessary or ineffective maintenance [24, 27], and given the large number of such machine, the waste can be significant. Therefore, there is an urge to develop a cost effective maintenance strategy.

Ever since the conception of Minimal repair model in 1960 [5], many optimal Maintenance strategies have been laid down in order to have minimum failures and maximum efficiency. The most common maintenance technique used is either preventive [5, 17] or corrective [7] in nature. Both these techniques have time and again been deemed as ineffective because Corrective Maintenance (CM)[2] inculcates large and unpredictable downtimes while Preventive Maintenance (PM) [21, 29] swings between extreme cases of 'more than necessary' and 'less than necessary' repairing frequency during infancy and aging periods respectively [34]. In particular, PM is carried out periodically on the basis of experience [41] or from the recommendation of the equipment manufacturers, and it does not account for unexpected failures because their failure mechanisms are time based [8]. PM and CM has been employed for general medical equipment [16, 19, 23, 25] but their implementation lacks simplicity and cost-effectiveness [21, 24].

Contrary to these maintenance strategies, Predictive Maintenance (PdM) [4, 40] is known to be more viable and effective method [6,26] which has been applied to various applications such as railway network [6], diode lasers [30], aircraft [36, 37], high yield etching process [10, 16, 38] etc. PdM technique is basically a condition-driven preventive maintenance scheme which takes into account the operational condition, efficacy and other health indicators that determine actual time needed for maintenance [22, 24, 25, 35, 36].

PdM technique has been applied for the evaluation of Haemodialysis machine's performance where their focusses were on the dialysis session performance and adequacy [10, 18]. However, these dialysis session performance and adequacy are dependent on the patients' condition rather than Haemodialysis machine, and thus their techniques cannot be applied to the machine maintenance.

To be more specific, the most common performance metrics for Haemodialysis machines is 'Kt/V' (dialyzer urea clearance K and dialysing time t per unit urea distribution volume V) or 'URR' (Urea Reduction Ratio) values [8, 5, 11] and they were considered as the health index for the application of PdM [19]. However, post and pre-dialytic urea concentration are required to evaluate URR, and urea clearance, dialysing time and distribution volume are required to calculate Kt/V. All these are dependent more on the patients' conditions rather than the Haemodialysis machine itself. Hence, it is not possible to determine the actual machine's health index or metrics for developing the maintenance strategy of the equipment. Therefore, the standard PdM strategy is not appropriate for the Haemodialysis system.

As the machine degradation depends on its components' degradation, reliability of the components will determine the reliability of the machine and its degradation, hence maintenance policy can also be developed at the components' level as well. The reliability of all the components can be calculated and thus optimal maintenance schedule can be planned accordingly. Such methodology is termed as the RCBM (Root Cause Based Maintenance) policy and has been developed by one of the author [35].

This work therefore implements RCBM strategy on Haemodialysis machines in Linkou Chang Gung Memorial Hospital, Taiwan. In particular, a method to evaluate specific time intervals to replace or maintain the individual components is demonstrated, in order to avoid unexpected failures and minimize the maintenance cost. The reliability data of each component is evaluated using Reliasoft software while the optimal maintenance time intervals are deduced using Genetic Algorithm which is implemented using MATLAB software. The maintenance cost reduced to approximately 40% of the original with the implementation of the RCBM strategy.

2. Haemodialysis description

4008S Haemodialysis Systems by Fresenius Medical Care stationed at one of the most reputed hospitals of Taiwan- Linkou Chang Gung Memorial Hospital are used for this work. There are 82 Haemodialysis machines in total installed at the hospital branch. Fig. 1 depicts the functional classification of the Haemodialysis system which comprises of several sub-parts or components under regular maintenance.

Each of the functions defined in Fig. 1 is performed by several sub-parts or components out of which 19 components are of utmost interest due to regular failures observed in a span of about 8 years,

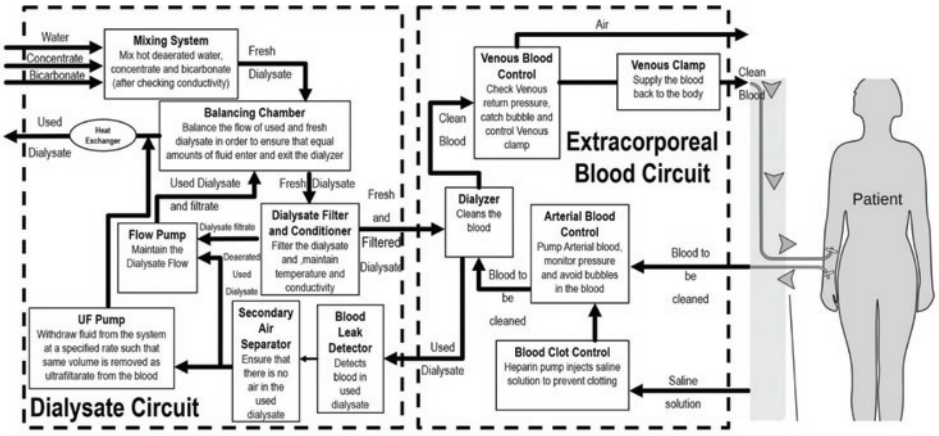


Fig. 1. Functional classification of 4008S Haemodialysis machine used for PdM strategy [9].

even though they are regularly maintained by the hospital. Table I shows a list of these 19 components.

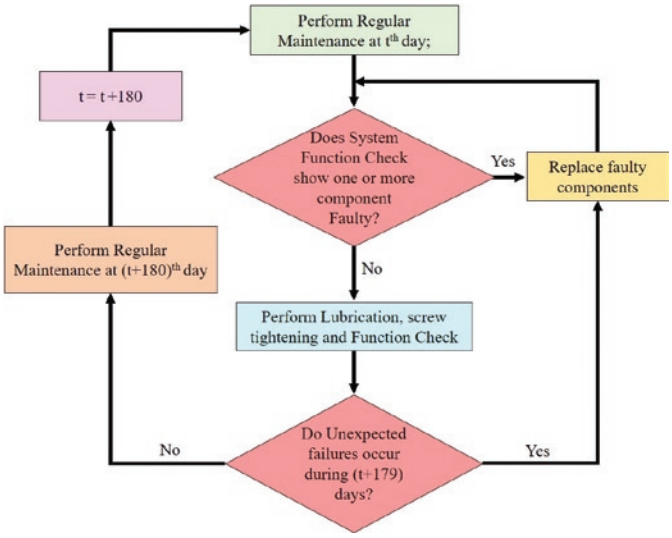


Fig. 2. Current practice of 180-day regular maintenance for the Haemodialysis machines at the Hospital.

The maintenance strategy used at the Hospital has been preventive in nature wherein the maintenance staff carries out regular maintenance every 180 days suggested by the manufacturer as shown in Fig. 2. If any faulty components are detected during the regular maintenance, they are replaced by a new component. Any failure in between regular maintenance cycle of 180 days will cause a suspension of that particular machine for a day or two, and a spare machine will be used in replacement while the faulty machine is diagnosed with necessary remedial action taken.

Such a maintenance scheme has the following drawbacks:

1. Not all the components have similar failure rates, and they also vary differently over the age of the components and machine. Hence, there are possibilities where not all components require maintenance every 180 days, some may be less and some may be more.
2. The downtime and maintenance cost is high particularly for unexpected failures which are not taken into consideration with the current practices with regards to maintenance at the Hospital.

In a need for improvement in the maintenance strategy where unexpected failures and various components' failure rates are to be taken into consideration, methodology for the application of PdM strategy on the 4008S Haemodialysis machine is developed in this work as will be discussed in the subsequent sections.

3. Methodology of RCBM strategy

The methodology for the application of RCBM strategy (a version of PdM) consists of formulation a 'Maintenance Policy' based on the Reliability analysis of components from maintenance records and the evaluation of a 'Cost Model' (described in subsequent sub-sections), in order to achieve a cost effective maintenance. This is followed by optimization of maintenance schedules for each component, in order to minimize the overall cost of maintenance.

The aforementioned methodology incorporates following assumptions:

Table I. Components of interest for application of PdM

S. No	Name of Components	No. of failures in a span of 3060 days	Abbreviation used for the component	S. No	Name of Components	No. of failures in a span of 3060 days	Abbreviation used for the component
1	Motor 29	141	M29	11	Carbon Cleaning Brush	5	CCB
2	Valve 39	20	V39	12	Silicone Tube	10	ST
3	Relief Valve	38	RV	13	AK O-Ring	14	AKOR
4	Trans Membrane Pressure	33	TMP	14	#65 Regulator	12	65R
5	Network Terminal Box	6	NTB	15	Flow Current	7	FC
6	Blood Leak Detector	61	BLD	16	Power Supply	23	PS
7	Gear 29	71	G29	17	CAL Pressure	4	CP
8	Filter 210	28	F210	18	Heparin Pump	8	HP
9	Motor 21	76	M21	19	AK Tube	8	AKT
10	Rise Chamber	7	RCh				

a) If a component is replaced during regular maintenance or unexpected failures, the reliability of the component will be restored to 100% after replacement.
b) The time window for cost calculation of the developed maintenance strategy is based on the existing life cycle of the machine (which is 3060 days) for fair comparison.
c) The failure mechanism of one component does not interact or influence that of other components for the sake of simplicity in analysis.

3.1. Reliability based 'Maintenance Policy'

In order to develop a new maintenance policy, the maintenance records from the hospital, for a duration of about 8 years is studied. The current maintenance policy is a fixed period preventive maintenance of 180 days wherein faults are diagnosed during regular maintenance or unexpected failures and rectified by replacing the faulty component. Fig. 3 shows the stick-diagram for the existing maintenance strategy of a component presented as a Network Terminal Box (NTB) for illustration. The solid thick sticks represent 180-day regular maintenance and the thin-smaller sticks represent the actual failures that happened in-between regular maintenance cycles.

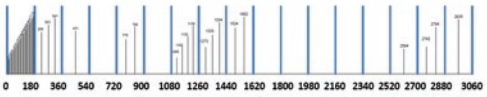


Fig. 3. Stick-diagram for the existing maintenance strategy of a component presented as Network Terminal Box, for illustration

Fig. 3 highlights two major issues associated with the existing maintenance strategy. *Firstly*, the initial failures (within the first 180 days) are large in number which was not detected earlier before our analysis, and *secondly*, unexpected failures occur quite often, in-between regular maintenance period. All these lower the operational availability of the equipment and increases the maintenance cost due to unexpected maintenance and the need of spare machines in order to maintain high availability. The first issue can be detected through standard data logging. However, the second issue requires the study of the reliability of each component.

In order to improve the existing maintenance strategy, the maintenance record for each of the 19 components mentioned in Table I is analysed using Weibull++ v10 by Reliasoft. The failure times and reliability of each component (excluding initial failures) are estimated

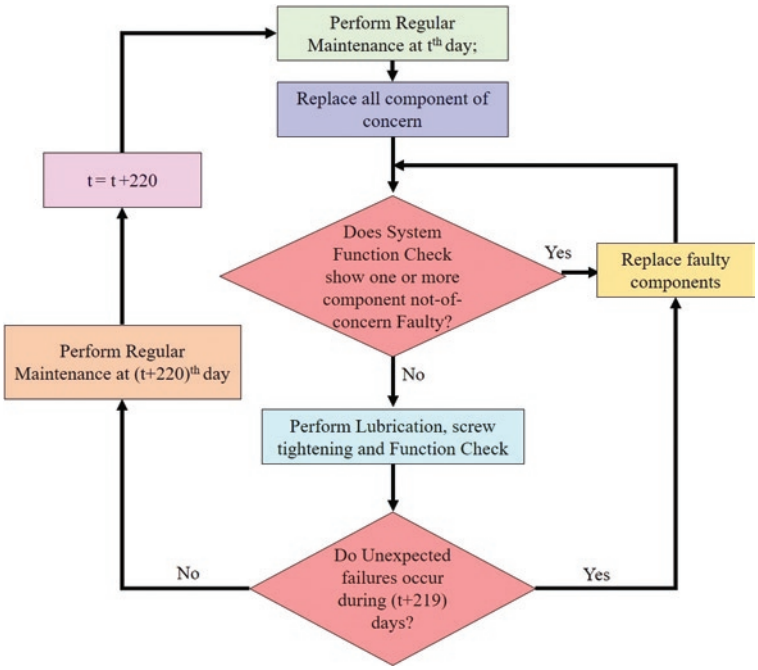


Fig. 4. Suggested 220-day maintenance policy

Table II. 220-Day Maintenance Scheduling based on Reliability Data (without considering the initial failures)

S. No	Component	Associated Statistical Distribution	Distribution Parameters
1	M29	2P-Weibull	Beta: 2.49, Eta (Day): 971.02
2	V39	3P-Weibull	Beta: 1.32, Eta (Day): 5345.53, Gamma (Day): 89.06
3	RV	2P-Weibull	Beta: 2.30, Eta (Day): 2612.57
4	TMP	2P-Weibull	Beta: 2.10, Eta (Day): 3736.00
5	NTB	2P-Weibull	Beta: 2.08, Eta (Day): 2743.23
6	BLD	3P-Weibull	Beta: 0.74, Eta (Day): 4050.97, Gamma (Day): 196.61
7	G29	Lognormal	Log-Mean (Day): 7.34, Log-Std: 0.70
8	F210	Gamma	Mu (Day): 7.77, K: 1.78
9	M21	3P-Weibull	Beta: 1.93, Eta (Day): 1833.00, Gamma (Day): 4.19
10	RCh	2P-Weibull	Beta: 1.55, Eta (Day): 9503.80
11	CCB	3P-Weibull	Beta: 0.89, Eta (Day): 21327.93, Gamma (Day): 351.63
12	ST	Lognormal	Log-Mean (Day): 7.99, Log-Std: 0.15
13	AKOR	Lognormal	Log-Mean (Day): 8.46, Log-Std: 0.76
14	65R	3P-Weibull	Beta: 2.24, Eta (Day): 3711.45, Gamma (Day): 375.00
15	FC	Gamma	Mu (Day): 8.24, Gamma (Day): 1.82
16	PS	2P-Weibull	Beta: 2.08, Eta (Day): 3579.83
17	CP	2P-Weibull	Beta: 1.47, Eta (Day): 12556.33
18	HP	Lognormal	Log-Mean (Day): 8.70, Log-Std: 0.70
19	AKT	Lognormal	Log-Mean (Day): 7.96, Log-Std: 0.37

Table III. 220-Day Maintenance Scheduling based on Reliability Data

Maintenance Cycle	Components of-concern to be maintained in			Total No. of Components to be maintained
	220 days	440 days	1100 days	
1st cycle on 220th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	-	-	10
2nd cycle on 440th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	CCB, ST, AKOR, 65R, FC, PS, CP	-	17
3rd cycle on 660th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	-	-	10
4th cycle on 880th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	CCB, ST, AKOR, 65R, FC, PS, CP	-	17
5th cycle on 1100th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	-	HP, AKT	12
6th cycle on 1320th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	CCB, ST, AKOR, 65R, FC, PS, CP	-	17
7th cycle on 1540th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	-	-	10
8th cycle on 1760th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	CCB, ST, AKOR, 65R, FC, PS, CP	-	17
9th cycle on 1980th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	-	-	10
10th cycle on 2200th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	CCB, ST, AKOR, 65R, FC, PS, CP	HP, AKT	19
11th cycle on 2420th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	-	-	10
12th cycle on 2640th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	CCB, ST, AKOR, 65R, FC, PS, CP	-	17
13th cycle on 2860th day	M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh	-	-	10
14th cycle on 3080th day	M21, RCh	-	-	2

using the Cumulative Distribution Function (CDF) based on specific statistical distribution associated with the specific components as shown in Table II.

Assuming the reliability of each component must be above 0.97, which implies that the chance of the failure of each component is less than 0.03, we can derive a suitable maintenance cycle for each component as shown in Table III.

From our reliability analysis of the components, we found that their reliabilities are adequately high, apart from their early failure, and thus 180-days maintenance cycle is too conservative. Also, Table III shows that not all the 19 components need to be maintained during the regular maintenance. In a life cycle period of 3060 days, 10 components including M29, V39, RV, TMP, NTB, BLD, G29, F210, M21, RCh are to be maintained every 220 days, 7 components including CCB, ST, AKOR, 65R, FC, PS, CP are to be maintained every 440 days, 2 components including HP, AKT are to be maintained every 1100 days. The 14th cycle has an exception of only 2 components (M21 and RCh) instead of 10 in 220-day cycle. Fig. 4 shows the suggested 220-day maintenance strategy.

In comparing the current 180-days maintenance policy with the suggested 220-days maintenance policy, one can obtain the following major advantages:

- Number of maintenance cycles reduce to 14 cycles from 17 in a life cycle period of 3060 days.
- The regular maintenance cost reduces by 18.18%. This results in a total saving of 18,279 Unit Dollar (Unit$) per year per machine (excluding initial and unexpected failures' consideration)

The value of 0.97 reliability criteria is arbitrary, and it could be too stringent or too loss to assure cost effective maintenance with respect to lower unexpected failure and high availability. Thus, a systematic method is needed to determine the reliability criteria for

each component, and a Cost-Model for a RCBM maintenance policy is required. The development of the Cost-Model will be discussed in the next sub-section.

One needs to note that the initial failures are not included in the cost saving calculations because these can be avoided by considering better quality components or by considering a different vendor.

3.2. The Cost model

The 'Cost Model' is formulated in order to evaluate the cost-saving from the suggested maintenance strategy including both expected and unexpected failures as well as for optimization of maintenance strategy. Assuming there are k components and the replacement period of component i to be t_i, then the total cost of maintenance can be given by equation (1):

$$\text{Totalcost}\left(t_1, t_2, t_3, \ldots, t_k\right) = \sum_{i=1}^{k} C_{ex,i} + C_{tk} + \sum_{i=1}^{k} C_{uex,i} - C_{rp} \quad (1)$$

where $C_{ex,i}$ is expected cost of replacement for i^{th} component, $C_{uex,i}$ is expected cost of replacement for i^{th} component, C_{tk} is toolkit cost and C_{rp} is calculated function check and rinse cost (repeated).

With this equation, we can develop optimal RCBM policy to minimize the total cost of maintenance as follows, where Cm is the minimum total cost of maintenance:

$$Cm = min\left(Total\ cost\left(t_1,\ t_2,\ t_3, \ldots, t_k\right)\right) \quad (2)$$

The detailed calculations of the 4 terms in equation (1) are deduced below.

3.2.1. Expected cost of replacement ($\sum_{i=1}^{k} C_{ex,i}$)

As replacement of components are bound to happen, the expected cost of replacement ($\sum_{i=1}^{k} C_{ex,i}$) for k components is the summation of expected cost of replacement $C_{ex,i}$ for i^{th} component which can be computed as the sum of the price of the component P_i and the related manpower cost as given below:

$$C_{R,i} = P_i + \left(T_{r,i} + T_{fc,i}\right) \times C_{mp} \quad (3)$$

where C_{Ri} is the cost of replacing component i^{th} time, $T_{r,i}$ and $T_{fc,i}$ are the man-hour for replacing and performing function check-rinse respectively, and C_{mp} is the manpower cost per hour.

Hence, the expected cost of replacement for k components is given by equation (4):

$$\sum_{i=1}^{k} C_{ex,i} = \sum_{i=1}^{k} int\left[\frac{T_s}{t_i}\right] \times C_{R,i} \quad (4)$$

where $C_{ex,i}$ is expected cost of replacement for i^{th} component, $int\left[\frac{Ts}{ti}\right]$ represents the expected number of replacements within a specific time interval of T_s.

3.2.2. Regular maintenance cost (C_{tk})

The regular maintenance involves basic function check-rinse and repairing using a tool kit which costs P_{tk}. Thus, the regular maintenance cost is given by equation (5):

$$C_{tk} = int\left[\frac{T_s}{t_{tk}}\right] \times \left(P_{tk} + \left(T_{r,tk} + T_{fc,tk}\right) \times C_{mp}\right) \quad (5)$$

where C_{tk} is toollkit cost, t_{tk} is the regular maintenance period, $\left(T_{r,tk} + T_{fc,tk}\right)$ is the time of repairing work and function check and rinse (about 2 hours) and C_{mp} is the manpower cost per hour.

3.2.3. Unexpected cost of replacement ($\sum_{i=1}^{k} C_{uex,i}$)

Unexpected failure can happen in-between designated maintenance cycles. In order to calculate the unexpected cost of replacement, number of unexpected failures is to be determined using the Non-Homogeneous Poisson's (NHP) process [2, 15, 16, 31], as the failure rate of all the components are time-dependent.

To apply NHP process (Fig. 5), one needs to determine the time interval so that the failure rate can be treated as a constant within the interval. Too large the time interval will produce error as the variation in the failure rate will be too high to be considered as approximately

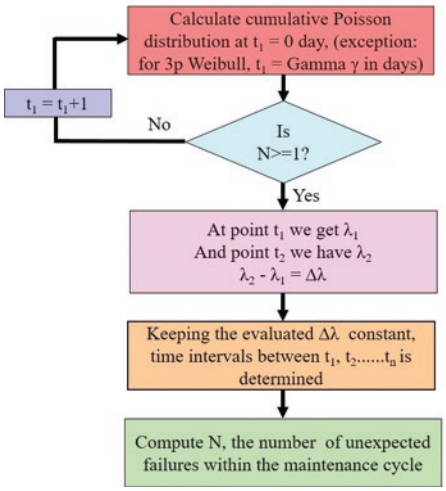


Fig. 5. Procedure to calculate number of unexpected failures

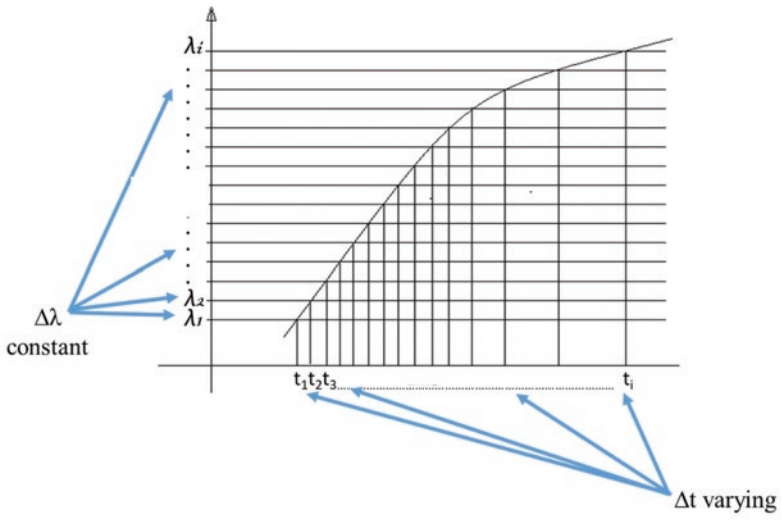


Fig. 6. 'Plot-Division' to determine Δt based on constant Δλ

constant. Too small the interval will also produce error as the number of failure will be zero.

Fig. 6 shows an example of the failure rate vs time curve as derived from the reliability function [12] of a component given as:

$$h(t) = (dF / dt) / (1 - F(t)) \quad (6)$$

where F(t) is the cumulative density function given as F(t)=1-R(t).

In Fig. 6, the failure curve is divided into time intervals in such a way that Δt will be varied according to the curve while Δλ will be constant. Table IV shows an example for the calculation of *N* in case of a component- *Motor 29*, for a time period of 25 days, which is its replacement period.

Table IV. An example of calculation of N in case of a component- Motor 29, for a time period of 25 days

N	Time	lambda	Reliability	Poisson Distribution, P_N	Cumulative Poisson Distribution	$N*P_N$
0	25	0.000619	0.984644123	0.993829119	0.9938291	0
1	25	0.000619	0.984644123	0.006151802	0.9999809	0.006151802
2	25	0.000619	0.984644123	1.90398E-05	1	~0
3	25	0.000619	0.984644123	3.92855E-08	1	~0

From Table IV, unexpected failure cost for Motor 29 for a time period of 25 days can be calculated as:

$$Component\ Cost \times (N \times P_N) = 9291.97 Unit\$ \times (1 \times 0.006151802) = 57.514319\ Unit\$$$

In general, the equation for Unexpected Cost calculation can be expressed as shown below:

$$\sum_{i=1}^{k} C_{uex,i} = \sum_{i=1}^{k} \left(\left(int\left[\frac{T_s}{t_i}\right] \times E\left[N_f(i,t_i)\right] + E\left[N_f(i,t_{irm})\right] \right) \times C_{R,i} \right) \quad (7)$$

where $E\left[N_f(i,t_i)\right]$ is the expected value of the number of unexpected failures within t_i for component i, $t_{irm} = T_s - int\left[\frac{T_s}{t_i}\right] \times t_i$ is the remaining time before the $T_s\ (here\ T_s = 3060\ days)$ ends and after the last replacement of component i, and $C_{R,i}$ is the replacement cost of the i^{th} component as shown in equation (3).

3.2.4. Repetitive cost of function check and rinse (C_{rp})

During a repair event, if there are more than one components that have to be replaced or, if this repair event coincides with the regular maintenance event, the function check and rinse will only be performed once. In order to compensate the repetitive calculation of the cost of function check and rinse (C_{rp}) as evident from equations (3) and (5), the term C_{rp} is subtracted in equation (1). Assume there are w repair event in an interval T_s, repetitive cost of function check and rinse is given by equation (8):

$$C_{rp} = \sum_{j=1}^{w} C_{rp,j} = \sum_{j=1}^{w} (m_j - 1) T_{fc} \times C_{mp} \quad (8)$$

where m_j is the number of the components that have to be repaired in the repair event j.

3.2.5. Optimal failure criteria for minimization of total maintenance cost

In our previous example, we assume the reliability criteria of each component to be 0.97. Such high reliability criteria can reduce the number of unexpected failures within a maintenance cycle, but it will also shorten the maintenance period and thus increases the expected cost of maintenance. Fig. 7 shows the plot of Total Maintenance Cost for one of the components- Motor 21 (M21) with respect to the varying failure criteria as computed using Equations (4)-(8).

From Fig. 7, one can see that the minimum total maintenance cost occurs when the failure probability is chosen to be 0.3, neither too high nor too low. However, the plot in Figure 7 is specific to only one component. Every other component will have similar curve with different failure criteria for minimum value of Total Maintenance Cost. Therefore, there is a need for optimization of the maintenance periods in order to cater for different failure criteria and minimization of Total Maintenance Cost.

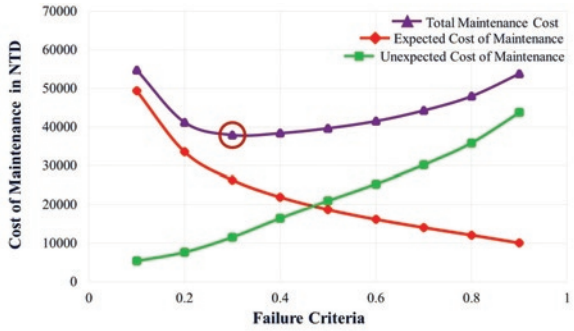


Fig. 7. Cost of Maintenance vs Failure Criteria chosen for replacement period and regular maintenance of Motor 21; red solid circle shows the point of lowest total cost of maintenance for Motor 21

4. Optimization of RCBM policy

In order to select an appropriate optimization technique for the proposed RCBM policy, we studied different optimization algorithms [20, 28] and Genetic Algorithm (GA)[32, 38] is selected because of the following reasons:

1. GA has a capability to be implemented as a 'universal optimizer' that could be used for optimizing any type of problem belonging to different fields [38,39].
2. Simplicity and ease of implementation of GA.

The basic principal of genetic algorithm is inspired by Charles Darwin's theory of natural evolution and Gene theory. Genetic algorithm optimizes the output in five phases[14] which in this case study can be understood as:

1. *Population Definition:* It defines target maintenance schedule (individuals in Gene theory) set.
2. *Fitness Function Analysis*: It determines how fit a maintenance schedule (individual in Gene theory) is by assigning a fitness score. The fitness function in this case is the cost model.

3. *Selection*: The idea is to select the fittest maintenance schedule (individuals in Gene theory) based on the fitness score.
4. *Crossover*: It is the most significant phase in a Genetic Algorithm. For each pair of maintenance schedules (parents in Gene theory) to be interacted (mated in Gene theory), a crossover point is chosen at random from within the prior set of information (Genes in Gene theory).
5. *Mutation*: In certain new maintenance schedules (offspring in Gene theory) formed, some of their information (Genes in Gene theory) can be subjected to a variation (Mutation in Gene theory) with a low random probability.

In this work, Genetic Algorithm function is implemented in MATLAB optimization toolbox using the 'Cost Model' derived in the previous section, so as to find the best periodic replacement schedule for each component. The key elements for implementation of GA is shown below:

1. In order to find the $min\left(Total\ cost\left(t_1, t_2, t_3, \ldots, t_k\right)\right)$, the replacement period of each component t_i is encoded as the GA's population. Here the replacement period t_i is considered in month. The population size is taken as 20 times the total number of components.
2. The constraints considered is with respect to the replacement period t_i.

$$Constraint => 1\left(month\right) \le t_i \le T_s$$

3. where T_s is 3060 days in this case-study
4. The fitness function is according to the deduced 'Cost Model'.
5. The stop criteria for the iterations in GA is as follows:
   - Limit of generation = Infinity
   - Stall generations = 300
   - Function tolerance = 1e-7

Genetic Algorithm often needs large execution times to obtain the approximate optimized solution. Therefore, a computer with 20 core parallel computing is employed for the execution of Genetic Algorithm on an Intel® Xeon® CPU E5-2650 v3 @2.30GHz, which takes about 36 hours for the solution to converge.

5. Results and discussions

5.1. Optimal maintenance schedules

The optimal maintenance schedule obtained from the Genetic Algorithm execution is summarized in Table V. It corresponds to minimized total maintenance cost value of 359,102 Unit$ for the entire life cycle of 3060 days for the Haemodialysis machine studied in this work.

5.2. Comparison of Average Cost of Maintenance

The average cost of maintenance of one machine (ACoM) is calculated as follows:

$$ACoM = \left(\sum_{j=1}^{r} \frac{TC_j}{T_j} \right) \div r, \quad \left(cost\ /\ day\right) \tag{9}$$

where r is the number of machines, TC_j is the total cost of maintenance of machine 'j' and T_j is the age of machine 'j' in days. All the machines are divided into four groups (on the basis of age) in order to draw a fair comparison between the actual maintenance cost and the corresponding optimized value.

Fig. 8 shows the comparison of the ACoM for all the four groups associated with the current and optimal maintenance strategy.

Using the optimal RCBM strategy, huge cost saving is exhibited with respect to the ACoM as shown in Fig. 8. The Average Cost of Maintenance of a machine which follows the current 180-day regular maintenance strategy is 128.91 Unit$/day. This value decreases to 78.83 Unit$/day in case of Reliability based 220-day maintenance strategy. The ACoM for the optimized RCBM strategy reduces further to 49.84 Unit$/day, accounting for almost 60% of the total cost saving.

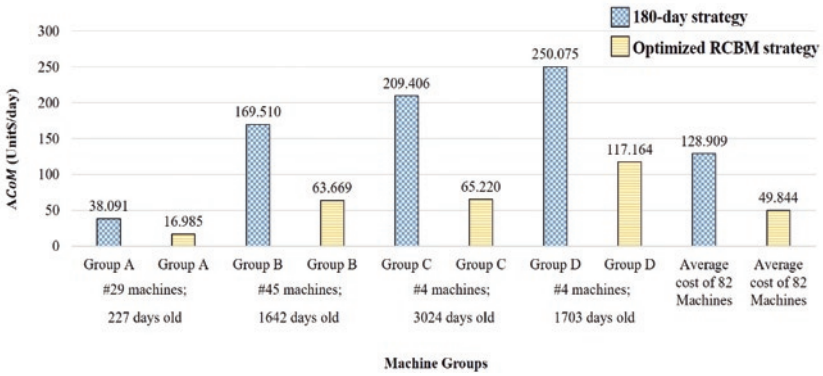


Fig. 8. Comparison of ACoM between the machine groups with existing 180-day policy and optimized RCBM strategy

Table V. Optimised Maintenance Schedules deduced from GA

S. No	Component	Optimal Maintenance Schedule (in Days)
1	Motor-29	780
2	Valve-39	1950
3	Relief-Valve	1830
4	Trans-Membrane Pressure	1890
5	Network-Terminal Box	1860
6	Blood-Leak Detector	3090
7	Gear-29	1860
8	Filter-210	2040
9	Motor-21	1830
10	Rise-Chamber	1890
11	Carbon-Cleaning Brush	2880
12	Silicone-Tube	2340
13	AK-O-Ring	780
14	#65-Regulator	1890
15	Flow-Current	630
16	Power-Supply	1710
17	CAL-Pressure	630
18	Heparin-Pump	3090
19	AK-Tube	780

Table VI. Comparison of Operational Availability for 180-day maintenance policy and optimized RCBM policy

Maintenance Policy	MTBF (in days)	MDT (in days)	OpAv
180-day Regular Maintenance	622.02	3.0200	0.99517
Optimized RCBM	742.82	0.0384	0.99995

5.3. Comparison of Operational Availability

The operational availability (OpAv) [1, 21,22] for a given system is defined as the fraction of average availability over a period of time and is expressed by equation (10):

$$OpAv = \frac{MTBF}{MTBF + MDT} \tag{10}$$

where MTBF is the Mean Time Between Failures and MDT is the Mean Down-time.

Table VI shows a comparison of MTBFs, MDTs and OpAVs of the current 180-day maintenance policy and optimal RCBM policy.

The operational availability (OpAv) calculated for the optimal RCBM is much higher than the same for 180-day regular maintenance practice.

6. Conclusion and future work

Haemodialysis machines at a reputed hospital are investigated in this work. The existing maintenance strategy is analyzed which is preventive in nature. Root Cause Based Maintenance (RCBM) strategy is implemented for reducing the cost of maintenance. Optimization of the maintenance schedules using Genetic Algorithm results in improving the operational availability and cost saving of about 60% of the current practice.

Spare part analysis and interaction of components degradation are important future works which can enable complex scenario to be included in the maintenance optimization.

Acknowledgement

The authors would like to thank the Linkou Chang Gung Memorial Hospital, Taiwan and Center for Reliability Sciences and Technologies (CReST), Chang Gung University, Taiwan for the support and smooth conduct of the project. The maintenance records were collected from Linkou Chang Gung Memorial Hospital, Taiwan. The statistical analysis and optimization were performed at the Centre of Reliability Sciences and Technologies in Chang Gung University.

Funding Source

This project has been funded by Linkou Chang Gung Memorial Hospital, Taiwan under the project: CZRPD2E0013.

References

1. Anderson D J, Brown T J, Carter C M. System of systems operational availability modeling. In NDIA SE Conference. San Diego, CA, USA, 2012; 1–20.
2. Andrzejczak K, Młyńczak M, Selech J. Poisson-distributed failures in the predicting of the cost of corrective maintenance. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 602–609, https://doi.org/10.17531/ein.2018.4.11.
3. Azar A T, Wahba K, Rafi M, Abdalla S C. System dynamics highlights the effect of maintenance on hemodialysis performance. In 25th International Conference of the System Dynamics Society, Boston, Massachusetts USA 2007.
4. Ballesteros A, Sanda R, Maqua M, Stephan J-L. Maintenance related events in nuclear power stations. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (1): 26–30, https://doi.org/10.17531/ein.2017.1.4.
5. Barlow R, Hunter L. Optimum preventive maintenance policies. Operations Research 1960; 8:90–100, https://doi.org/10.1287/opre.8.1.90.
6. Consilvio A, Di Febbraro A, Meo R, Sacco N. Risk-based optimal scheduling of maintenance activities in a railway network. EURO Journal on Transportation and Logistics 2018; 1–31, https://doi.org/10.1007/s13676-018-0117-z.
7. Daugirdas J T. Second generation logarithmic estimates of single-pool variable volume Kt/V: An analysis of error. Journal of the American Society of Nephrology 1993; 4: 1205–1213.
8. Dieulle L, Berenguer C, Grall A, Roussignol M. Continuous time predictive maintenance scheduling for a deteriorating system. In: Annual Reliability and Maintainability Symposium. Proceedings. International Symposium on Product Quality and Integrity (Cat. No.01CH37179) IEEE 2001; 150–155, https://doi.org/10.1109/RAMS.2001.902458.
9. Duc L M, Tan C M, Luo M, Leng I C H. Maintenance scheduling of plasma etching chamber in wafer fabrication for high-yield etching process. IEEE Transactions on Semiconductor Manufacturing 2014; 27: 204–211, https://doi.org/10.1109/TSM.2014.2304461.
10. European Best Practice Guidelines Expert Group on Hemodialysis. European Renal Association Section II. Haemodialysis adequacy. Nephrology Dialysis Transplantation 17 Suppl 2002; 7: 16–31.
11. Fresenius 4008E/4008B/4008H/4008S. Hemodialysis machine technical manual 2003. http://darman.umsha.ac.ir/uploads/15_2154_82_fresenius_4008_hemodialysis.pdf. Accessed 19 Nov 2018
12. Hadjila T, Ahmed A S. Estimation and simulation of conditional hazard function in the quasi-associated framework when the observations are linked via a functional single-index structure. Communications in Statistics - Theory and Methods 2018; 47: 816–838, https://doi.org/10.1080/03610926.2016.1213294.
13. Hemodialysis Adequacy 2006 Work Group. Clinical practice guidelines for hemodialysis adequacy, Update. American Journal of Kidney Diseases 2006; 48: S2–S90, https://doi.org/10.1053/j.ajkd.2006.03.051.
14. Holland J H. Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. Oxford: U Michigan Press, 1975.

15. Hong H P. Application of the stochastic process to pitting corrosion. Corrosion 1999; 55: 10–16, https://doi.org/10.5006/1.3283958.
16. Huang CY, Lyu M R, Kuo S Y. A unified scheme of some nonhomogenous poisson process models for software reliability estimation. IEEE Transactions on Software Engineering 2003; 29: 261–269, https://doi.org/10.1109/TSE.2003.1183936.
17. Jaśkowski P. Methodology for enhancing reliability of predictive project schedules in construction. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2015; 17 (3): 470–479, https://doi.org/10.17531/ein.2015.3.20.
18. Jonsson P. Safety and biological aspects of present techniques of haemodialysis. Umea, Sweden, 2006.
19. Khalaf AB, Hamam Y, Alayli Y, Djouani K. The effect of maintenance on the survival of medical equipment. Journal of Engineering, Design and Technology 2013;11: 142–57, https://doi.org/10.1108/JEDT-06-2011-0033.
20. Khan S, Asjad M, Ahmad A. Review of modern optimization techniques. International Journal of Engineering Research & Technology 2015; 4: 984–988, https://doi.org/10.17577/IJERTV4IS041129.
21. Legát V, Mošna F, Aleš Z, Jurča V. Preventive maintenance models – higher operational reliability. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (1): 134–141, https://doi.org/10.17531/ein.2017.1.19.
22. Lin C C, Tseng H Y. A neural network application for reliability modelling and condition-based predictive maintenance. International Journal of Advanced Manufacturing Technology 2005; 25: 174–179, https://doi.org/10.1007/s00170-003-1835-3.
23. Loganathan M K, Gandhi O P., Maintenance cost minimization of manufacturing systems using PSO under reliability constraint, International Journal of System Assurance Engineering and Management 2016; 7: 47-61, https://doi.org/10.1007/s13198-015-0374-2.
24. Mahfoud H, ElBarkany A, ElBiyaali A. Preventive Maintenance Optimization in Healthcare Domain: Status of Research and Perspective. Journal of Quality and Reliability Engineering 2016; 2016: 1–10, https://doi.org/10.1155/2016/5314312.
25. Mazidi P, Sanz Bobi M A, Shayesteh E, Hilber P. Impact of health indicators on maintenance management and operation of power systems. In Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 2017; 231: 716–731, https://doi.org/10.1177/1748006X17731901.
26. Mekki S, Wahed A, Wahba K K, Ouda B K. A system dynamics-based model for medical equipment maintenance procedure planning in developing countries. In Cairo International Biomedical Engineering Conference, Cairo, Egypt 2012:104–108, https://doi.org/10.1109/CIBEC.2012.6473299.
27. Mobley R K. An introduction to predictive maintenance. Oxford: Butterworth-Heinemann, 2002.
28. Mohamed A A, Leemis L M, Ravindran A. Optimization techniques for system reliability: a review. Reliability Engineering & System Safety 1992; 35: 137–146, https://doi.org/10.1016/0951-8320(92)90033-H.
29. Ridgway M G. Manufacturer recommended PM intervals: is it time for a change? Biomedical Instrumentation & Technology 2009; 43: 498–500, https://doi.org/10.2345/0899-8205-43.6.498.
30. Ried S, Rauch S, Irmler L, Rikels L, Killi A, Papastathopoulos E, Sarailou E, Zimer H. Next generation diode lasers with enhanced brightness. In Proceedings: High-Power Diode Laser Technology XVI (SPIE) 2018; 15, https://doi.org/10.1117/12.2289464.
31. Rutter C M, Yu O, Miglioretti D L. A hierarchical non-homogenous Poisson model for meta-analysis of adenoma counts. Statistics in Medicine 2007; 26: 98–109, https://doi.org/10.1002/sim.2460.
32. Songran Liu, Zhe Li, A modified genetic algorithm for community detection in complex networks, In International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), 16-18 Feb. 2017, https://doi.org/10.1109/ICAMMAET.2017.8186747.
33. Statista Medical Technology Industry - Statistics & amp; Facts | Statista. https://www.statista.com/topics/1702/medical-technology-industry/. Accessed 19 Nov 2018
34. Tan C M, Raghavan N. A framework to practical predictive maintenance modeling for multi-state systems. Reliability Engineering & System Safety 2008; 93:1138–1150, https://doi.org/10.1016/j.ress.2007.09.003.
35. Tan C M, Raghavan N. Root cause analysis based maintenance policy. International Journal of Quality & Reliability Management 2007; 24: 203–228, https://doi.org/10.1108/02656710710722293.
36. Venter G. Review of optimization techniques. In: Encyclopedia of Aerospace Engineering. Chichester, UK: John Wiley & Sons, Ltd, 2010, https://doi.org/10.1002/9780470686652.eae495.
37. Wang Y. Development of predictive structural maintenance strategies for aircraft using model-based prognostics. TEL Archives 2017. https://tel.archives-ouvertes.fr/tel-01715558/. Accessed 19 Nov 2018
38. Whitley D. A genetic algorithm tutorial. Statistics and Computing 1994; 4: 65–85, https://doi.org/10.1007/BF00175354.
39. Why genetic algorithms is popular than other heuristic algorithms? https://www.researchgate.net/post/Why_genetic_algorithms_is_popular_than_other_heuristic_algorithms. Accessed 19 Nov 2018
40. Yi Zheng, Yongsong Wei, Shaoyuan Li, Coupling Degree Clustering-Based Distributed Model Predictive Control Network Design, IEEE Transactions on Automation Science and Engineering 2018; 15(4): 1749–1758, https://doi.org/10.1109/TASE.2017.2780444.
41. Zhou X, Xi L, Lee J. Reliability-centered predictive maintenance scheduling for a continuously monitored system subject to degradation. Reliability Engineering & System Safety 2007; 92: 530–534, https://doi.org/10.1016/j.ress.2006.01.006.

Cher Ming TAN
Center for Reliability Sciences & Technologies, Chang Gung University
No. 259, Wen-Hua 1st Road, Guishan District, Taoyuan City, Taiwan, ROC 33302

Department of Electronic Engineering, Chang Gung University
No. 259, Wen-Hua 1st Road, Guishan District, Taoyuan City, Taiwan, ROC 33302

Department of Mechanical Engineering, Ming Chi University of Technology
No. 84, Gungjuan Road, Taishan Dist., New Taipei City, Taiwan, ROC 24301

Department of Urology, Chang Gung Memorial Hospital
No. 5, Fuxing Street, Guishan Dist., Taoyuan City, Taiwan, ROC 33305

Institute of Radiation Research, College of Medicine, Chang Gung University
No. 259, Wen-Hua 1st Road, Guishan District, Taoyuan City, Taiwan, ROC 33302

Udit NARULA
Center for Reliability Sciences & Technologies, Chang Gung University
No. 259, Wen-Hua 1st Road, Guishan District, Taoyuan City, Taiwan, ROC 33302

Lu An LAI
Sumit PANDEY
Center for Reliability Sciences & Technologies, Chang Gung University
Department of Electronic Engineering, Chang Gung University
No. 259, Wen-Hua 1st Road, Guishan District, Taoyuan City, Taiwan, ROC 33302

Jung Hua TUNG
Center for Reliability Sciences & Technologies, Chang Gung University
No. 259, Wen-Hua 1st Road, Guishan District, Taoyuan City, Taiwan, ROC 33302

Chung Yi LI
Center for Reliability Sciences & Technologies, Chang Gung University
Department of Electronic Engineering, Chang Gung University
No. 259, Wen-Hua 1st Road, Guishan District, Taoyuan City, Taiwan, ROC 33302

E-mails: cmtan@cgu.edu.tw, ud@cgu.edu.tw, laileon.main@gmail.com, sumit.pandey.tech@outlook.com, fred.tung@mail.cgu.edu.tw, chungyili@mail.cgu.edu.tw

Article citation info:
WAWRZOSEK J, IGNACIUK S, BOCHNIAK A. Selected problems of ambiguity of the dual price of water in the post-optimization analysis of the water supply system. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 329–340, http://dx.doi.org/10.17531/ein.2019.2.18.

Jacek WAWRZOSEK
Szymon IGNACIUK
Andrzej BOCHNIAK

SELECTED PROBLEMS OF AMBIGUITY OF THE DUAL PRICE OF WATER IN THE POST-OPTIMIZATION ANALYSIS OF THE WATER SUPPLY SYSTEM

WYBRANE PROBLEMY NIEJEDNOZNACZNOŚCI CENY DUALNEJ WODY W POSTOPTYMALIZACYJNEJ ANALIZIE SYSTEMU WODOCIĄGÓW*

In literature it is believed that the dual price of water is an objective premise for shaping the market price of water. However, the authors note that a single vector of dual prices in the distribution of water, when ambiguous, should not become the basis for making decisions both regulating the price of water and affecting the procedures for modernizing the water supply network. This work cautions water management engineers not to duplicate common software errors and indicates how, despite the complete lack of literature tips, the technical problems encountered could be practically solved. The linear dependence of the row vectors of the left-hand parameters of binding constraints in the linear programming model for water consumption is identified here as the reason for the ambiguity of dual price vectors. This ambiguity in the issues of water distribution requires shaping alternative technical scenarios allowing for a variant selection of the method for modifying the water abstraction system. Therefore, the principles for determining the proportionality of simultaneous changes in certain parameters of the right-hand conditions of constraint conditions are described. These principles for the optimal selection of the most productive vectors for the parametric linear programming method were formulated and indicated on a simplified model of water distribution. The methodology developed in the work enables, among others, generating alternative technical scenarios for saving varying amounts of water, resulting in various financial savings.

Keywords: *water distribution network, dual price of water, operation process, water management model, post-optimization.*

W literaturze uważa się, że cena dualna wody jest obiektywną przesłanką do kształtowania rynkowej ceny wody. Jednak autorzy zauważają, że pojedynczy wektor cen dualnych w dystrybucji wody, gdy jest niejednoznaczny, nie powinien stać się podstawą do podejmowania decyzji zarówno normującej cenę wody jak i wpływającej na procedury modernizujące sieć wodociągową. Praca uczula inżynierów gospodarki wodnej by nie powielali powszechnych błędów oprogramowania oraz wskazuje jak, pomimo kompletnego braku literaturowych wskazówek, praktycznie rozwiązywać napotykane problemy techniczne. Liniowa zależność wektorów wierszowych parametrów lewych stron wiążących warunków ograniczających w modelu programowania liniowego dla zużycia wody identyfikowana jest tu jako przyczyna niejednoznaczności wektorów cen dualnych. Ta niejednoznaczność w zagadnieniach dystrybucji wody wymaga kształtowania alternatywnych scenariuszy technicznych pozwalających na wariantowy wybór sposobu modyfikacji systemu poboru wody. Dlatego opisano zasady wyznaczania proporcjonalności jednoczesnych zmian niektórych parametrów prawych stron warunków ograniczających. Na uproszczonym modelu dystrybucji wody sformułowano i wskazano te zasady optymalnego doboru najbardziej produktywnych wektorów dla metody parametrycznego programowania liniowego. Opracowana w pracy metodyka umożliwia m.in. wygenerowanie alternatywnych scenariuszy technicznych oszczędzania różnej ilości wody, skutkującej różnymi oszczędnościami finansowymi.

Słowa kluczowe: *sieć dystrybucji wody, cena dualna wody, proces eksploatacji, model gospodarowania wodą, postoptymalizacja.*

1. Introduction

1.1. Fixing the price of water in order to balance water demand

An analysis conducted by the World Bank [64] indicates that water shortages in some regions may reduce GDP by up to 6% and lead to increased migration and in some cases to a greater risk of conflict. Droughts and periods of water scarcity have become a more common and more frequent phenomenon in Europe [17]. The drought experiences in Europe in 2011, 2012, 2015 and 2018 were the worst in a century and affected not only Southern and Western Europe, but also the countries in Northern Europe (including Great Britain, France, Germany, Sweden and Poland). The difference between water supply and its growing demand also determines the key limitations of China's economic development. It is estimated that before 2005 due to a shortage of water in production, China lost $ 28 511 million annually. Brown [7] concluded that the shortage of water in China will soon be a threat to global cereal demand. Therefore, water, alongside cereals and crude oil, is referred to as a strategic resource. Hence, it is suggested that public authorities should shape water pricing for users in order to reflect its true shortage or alternative costs [15].

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

1.2. Methods of optimization and post-optimization in the planned exploitation of water resources

Optimization and post-optimization methods can be used as a tool allowing for dynamic correction and improvement of the operation of complex systems. These methods require two approaches: mathematical and managerial [29-30]. The tools and principles of optimization enabled the development of normative models for optimal management of large-scale water systems, taking into account the ubiquitous uncertainty in forecasting natural processes and economic effects [13]. Numerous optimization models are used to ensure high parameters of network reliability, water quality reliability parameters, appropriate operational schemes with considerations for numerous hydraulic limitations in the form of: *hydraulic head*, leakage, changes in the energy consumption of a pump and sequential discreet pump operation which minimises operational costs. Most optimization models use mathematical techniques such as linear programming (*LP*), dynamic programming (*DP*) and nonlinear programming (*NLP*) or variations thereof [2, 10, 13-14, 19, 22, 25, 53]. Some models of problems with reliability or optimization of the cost and maintenance time of a water supply or sewage system are constructed on the statistical-stochastic basis [3, 38]. Romaniuk [44-46] presents numerous numerical experiments focusing, for example, on optimizing the value of expected costs and duration of exploiting a water supply network, when the decision-making parameter is unconditional exchange time, i.e. when it is better to replace a network fragment instead of performing its another repair in the future.

Models of reliability, readiness and safety in connection with linear programming are useful in the identification and prediction of reliability, readiness and safety of complex technical systems as well as in the optimization and analysis of the operating costs of these systems [28, 32-33, 66]. The uncertainty of the parameters of a typical linear programming model indicated here requires, in addition to an optimization procedure, easy access to post-optimization procedures [56, 58].

Freire-González et al. [18] review literature on the existing research related to input-output models to assess the economic impact of water scarcity during drought and linear input-output *(IO-LP)* methods in the approach to water resource planning in the context of drought and water shortage. Gibbons [22] and Liu et al. [36] point out that correct pricing for water resources, reflecting their real value, is very important for saving water and for mitigating water shortages. They further recognize the dual price of water as the one that should serve to quantify the actual value of water resources, which also reflects their shortage. They do so despite the fact that it is sometimes claimed that it is practically impossible to obtain a dual water price by solving a linear programming model. However, Liu et al. [36] combining the input-output analysis method with the *LP* method, developed a model with limitations imposed on final demand, total production, trade balance and water availability. This model was used to estimate the dual price of water. These results constitute a valuable reference for determining reasonable prices for industrial and productive water in the areas of the nine main Chinese river basins. A review of analogous studies on modelling the value of water in various sectors of the economy in South Africa is carried out by Nieuwoudt and Backeberg [42] and in the world by Conradie and Hoag [11]. These studies were carried out as a result of the emergence of demand for models measuring the willingness to pay for water used, e.g., for irrigating agricultural crops.

1.3. Problems of linear programming in the modelling of water resources

A typical system of water resources consists of water reservoirs, hydro power station, irrigated land, artificial and navigational channels, etc. being within the range of a river or basin. Therefore, optimal planning of a multi-purpose water resource system, i.e. designing the "best" system, which is to be built and used in the planning horizon, is subordinated to, among others, technical, economic, financial, social and political restrictions. These restrictions include seasonal fluctuations in water supply, geographical and geological conditions of selected locations, existence of capital, loans, labour and local services, interest rate (and its trends), regional development plans, etc. [25]. The quoted authors used the *LP* model for a very complex water resources system, taking into account a number of constraints related to the reservoir, irrigation, hydro power station, artificial sewage and navigation limitations. The cases of rivers in southern Argentina were investigated, and these typical problems were described by about 300 constraints and 300 variables. With such a high number of constraints, it is not difficult to find linearly dependent row vectors for left-hand parameters of constraint conditions, and this already generates problems in the post-optimization analysed here.

McKee et al. [41] developed a model of an aquifer exploitation process through more than 900 wellbores, mainly for the needs of industry, municipal supply and crop irrigation in Arkansas. They took into account three variants of the *LP* models to simulate optimized outflows of surface and ground water, while simultaneously retaining the stream flow rate and numerous hydraulic limitations. Also this complex issue poses the above-mentioned problems with linear dependence.

Techniques for multi-criteria optimization, e.g. for a contaminated aquifer, come down to a single-criterion optimization through the use of the weighted sum method or the method of constraints [16, 27, 40]. In the latter of these methods, one of the objective functions is optimized by using other objective functions as a constraint, i.e. including them into the constraint conditions of the model, thereby increasing the number of constraints. This usually leads to the problem of linear dependence, analysed in this work and associated with the excessive number of constraint conditions.

Abdy Sayyed et al. [1] optimize the water distribution network by minimising network costs under constraint conditions resulting from pressure requirements at all nodes. Due to the high number of constraint conditions, which complicates the issue, already in the optimal design some constraint conditions are replaced there by an additional penalty in the objective function. The penalty is applied for failing to meet pressure constraints. In the cited work, three methods of penal inference were applied. This denotes a tendency, which sometimes emerges in the literature, to escape from an excess of constraint conditions. But this requires proper selection of the penalty function.

Frizzone et al. [19] in order to maximise the net income for several crops subjected to water and crop area access constraints, perform linearization of the non-linear objective function. Linearization is a typical tool used to carry out the optimization process [37].

1.4. Critical evaluation of some of the results presented in the literature

Numerous studies [26, 29-30, 40, 48, 55-58] indicate the necessity of caution when using *LP* methods. This is mainly due to the ambiguity of sensitivity reports containing dual price vectors [55-58]. In this work, it is noted that the ambiguity of dual price vectors is a consequence of the linear dependence of the row vectors of the parameters of the left-hand sides binding the constraint conditions of the *LP* model. This dependence must take place when the number of m_0 of binding constraint conditions exceeds the number of decision variables n. But this dependence is not usually controlled by analysts. Koltai and Terlaki [29] indicate that it appears almost always and for a small number m of constraint conditions. If the problem of the ambiguity of the dual price vector does not occur in *dual model* (*DM*), then the cyclicality of post-optimization procedures will quickly lead to it. It is noted that in the literature cited, water management models

contain *m* and *n* values, very often numbering even hundreds or thousands. And yet, already for their small values, this should inspire great caution in analysts when operating with dual prices. A lack of necessary criticism of the authors of many works who use dual prices for economic and technical issues is noticeable. And so, a single ambiguous vector of dual prices in water distribution should not become the basis for making decisions both regulating the market price of water and affecting the procedures modernising the water supply network. In this work, the need to consider several alternative technical decision scenarios based on some nodal solutions in the dual model to the original *LP* model, is signalled. This is where the methodology for creating these alternative technical decision scenarios is formulated. The authors note that the dubiousness in the literature on the usefulness of dual prices that can be seen so far is probably related to both the lack of knowledge about their ambiguity and the lack of practical methods for utilising this fact. The works [24, 55-58] may be an exception here. In section 2, on a simple example of a water supply network with infinitely many dual price vectors, it is suggested how to propose the framework for alternative modernization scenarios of a previously optimized network.

1.5. Interpretation of dual prices. Problems in the sensitivity analysis

Sensitivity analysis reports for *LP* indicate the features determining the choice of the optimal decision variant:

1. they describe some of the simple effects caused by deviation from the optimal plan (i.e. the height of the marginal increment, that is the amount by which the optimal value of the objective function and the scope of the correction in this amount should be adjusted);
2. they indicate how long one should refrain from changing the variant of the optimal decision when changing individual parameters of the linear objective function;
3. they indicate, by means of dual prices, how the optimal value of the objective function will change when changing (not necessarily single) parameters of right-hand sides of particular constraint conditions in a certain scope;
4. they indicate whether it is profitable for the company to increase the availability of a certain resource by a certain number of units (analysis of "more for less" Arsham [4]).

In general, the dual price measures the change in the value of the objective function resulting from the increased availability of a specific resource by a unit, usually with a clearly understated and implied assumption that the remaining "deficit" resources will not change. Each limited resource is then accounted for by a separate dual price. Each time, the impact of the change in the amount of each single "deficit" factor is examined (i.e. the one for which the constraint is binding). The dual prices remain constant until the set of binding limitations of the optimal solution changes. Each of these prices, measures the value of benefits from expanding production capacity or losses resulting from their reduction. In other words, the dual price corresponding to the right-hand side of a particular constraint condition indicates how much the value of the objective function will change when the limitation is relaxed. If a specific production factor, i.e. a certain resource, is not fully utilized in the optimal solution (i.e. it is not „deficit", and it does not constitute a binding), then it has a dual price equal to zero. It may be partially used after increasing other resources, but it does not have to be deficient. However, as it has been shown in the present work, any other factor with a dual price equal to zero, once it has been fully consumed (i.e. it has become „deficit" because it constitutes a binding constraints), may be additionally required in the proportion subjected to estimation in order to increase the amount of certain resources with a non-zero dual price. The analysis of changes not only for a single parameter of the right-hand sides of constraint conditions is the subject of numerous studies. A broad review of the literature in this area is presented by Shahin et al. [48]. The classification performed in this work allows for distinguishing – apart from ordinary sensitivity analysis – 7 other types of post-optimization analyses: 1) the rule of 100% [6], 2) "symmetric tolerance" – [59-61], 3) "non-symmetric tolerance" which is an extension of the symmetric tolerance, and introduced by Arsham and Oblak [5], Wondolowski [63] and Wendell [62], 4) (*PLP*) parametric linear programming [21, 47], 5) multiparametric linear programming [54], 6) sensitivity analysis with the functional dependence of the parameters of right-hand sides of constraint conditions or the objective function coefficients [23], 7) sensitivity analysis with the correlation of the above mentioned parameters [48]. Arsham's [4] work compares most of the above-mentioned methods through the construction of the largest sensitivity region for the general *LP*. Thus, Arsham [4] indicates most of these types as special cases in his analysis. Nevertheless, the question of many special cases, especially degenerated ones, remains unresolved, as Arsham [4] clearly indicates. It should be noted here that the preservation of certain (preferably optimal) proportions when increasing resources is the basis of the (optimal) *PLP*. Otherwise, part of the increased resources may remain unused, i.e. unproductive. Thus, the need to formulate the principles of optimal selection of the most productive vectors for *PLP* is recognized, which is the subject of the present work. Another problem, unresolved in the literature, is the practical usefulness of ambiguous sensitivity analysis reports, or even the fact that they are shown by popular calculation packages [26]. The fact that there are infinitely many solutions to the dual model [55-56] makes a sensitivity report most often unhelpful for an average analyst due to problems with the interpretation of ambiguous reports obtained. So far, this last problem has been described only partially and only in individual cases of transport models [57-58]. The transport models presented there are a special case of sensitivity analysis with the functional dependence of the parameters of the right-hand sides of constraint conditions [23].

Therefore, the authors of this work formulate the principles of determining the proportionality of simultaneous changes in the parameters of the right-hand sides of the constraint conditions in the case of ambiguous sensitivity analysis reports. At the same time, they identify the linear dependence of the row vectors of the left-hand side parameters of binding conditions as the reason for the ambiguity of dual prices. Under these conditions, they use various reports of ordinary sensitivity analysis available through commonly accessible software. They indicate the difference in the interpretation of the dual price corresponding to the first constraint condition for each of the sensitivity analysis reports (tables 2, 3, 6). This difference consists in the fact that the unit change of the right-hand side of the first constraint condition for each report forces the simultaneous, respectively proportional change of the right-hand side of other constraint conditions. In the example considered, the authors indicate that savings in water consumption by one unit may require simultaneous execution of one of two alternative scenarios of technical activities. At the same time, each of these two technical scenarios saves a different amount of water and results in other financial savings. The choice between these scenarios requires the inclusion of additional information not specified in this example. In the purpose of a clear presentation of the new methodology, useful in the operational processes of many complex technical systems, consideration was limited to conducting a study of a simplified model of water management.

2. A study of a simplified water management model

The authors have already indicated above that each *LP* model for a water supply network, with a single ambiguous solution of a dual task, should not be the basis for shaping market water prices or modifying parameters of this network without constructing several alternative

technical scenarios. And a full analysis of the already small *LP* model leads to many side threads, which are not significant for the presentation of the methodology of creating the foundations of alternative scenarios of technical procedure. Thus, in the simplified example below, a detailed interpretation was attached to only two constraint conditions with a partial interpretation of the majority of other constraint conditions of the *LP* model. It is a deliberate effort of the authors to point out the wide range of suitability of the methodology below, highlighting the importance of reducing water consumption and the costs of obtaining it. Making small changes in interpretations, the following example may equally well refer to a local water supply system, which is connected with three water abstraction nodes from a larger water supply system of a large urban and industrial agglomeration [cf. 38]. Such a model of water demand can be supplemented, among others, with the dynamics of seasonal changes and weekly rhythm [cf. 35]. Then, some parameters of the model change and the interpretation of particular constraint conditions change. In each of these cases the problem of a detailed interpretation of individual constraints will have to be solved individually, but each time the same problems resolved here will be revisited: 1) clear description of an incalculable set of all ambiguous dual prices, 2) attaching a practical interpretation to this ambiguity, e.g. through formulation of alternative post-optimization scenarios for the water supply network. In order to overcome these difficulties, an analyst of urban logistics systems should familiarize himself with the mathematical formalism presented here, describing the methodology of the transition from the answer reports and ambiguous sensitivity reports typical for *LP* to postulated scenarios. To make it clear, it is necessary to sketch a model for a small and not too complicated example of water distribution, which does not expose technical problems unnecessarily, but focuses on improving the methodology of engineering inference.

Therefore, in order to obtain simplicity and fix the attention on methodologies expanding the post-optimization, let us assume that the analysed urban-industrial agglomeration is supplied with water by three water abstractions located along the same watercourse in water abstraction sites P_i for $i = 1,2,3$ (as in Fig. 1).

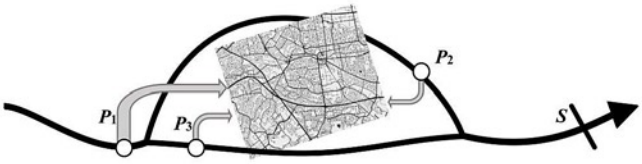


Fig. 1. Three water abstractions for an urban and industrial agglomeration located along the same watercourse, water gauge cross-section

The daily distribution of water in the agglomeration is a variable. Therefore, the amount of water sourced from the intake No. i is a decision variable $x_i \geq 0$ for $i = 1,2,3$ and periodically it can be determined in different units, e.g. [$dm^3 \cdot s^{-1}$], [$dm^3 \cdot h^{-1}$], [$m^3 \cdot d^{-1}$], [$m^3 \cdot month^{-1}$], [$m^3 \cdot year^{-1}$]. Moreover, let us assume that the current periodic agglomeration demand for water varies and is ranging between 200 and 300 units, which can be interpreted as constraints in the form of:

$$CO_1\colon\ x_1 + x_2 + x_3 \geq 200 \qquad CO_2\colon\ x_1 + x_2 + x_3 \leq 300$$

Due to the principle of the inviolability of water flow in the river through the water gauge cross-section *"S"*, or technical considerations, the dependencies between intake amounts are expressed as the following limitations:

$$CO_3\colon\ 9 \cdot x_1 + 11 \cdot x_2 + 7 \cdot x_3 \leq 1800 \quad CO_4\colon\ 5 \cdot x_1 + 6 \cdot x_2 + 4 \cdot x_3 \geq 1000$$
$$CO_5\colon\ 2 \cdot x_1 + 4 \cdot x_2 + 3 \cdot x_3 \geq 600 \quad CO_6\colon\ 0 \cdot x_1 + 2 \cdot x_2 + 1 \cdot x_3 \geq 200$$

At the same time, *instream flow* or *minimum acceptable flow* is defined as the amount of water, which should be left in the cross-section of a given stream due to biological, ecological and social considerations. The necessity of preserving this flow should not be subject to economic assessments. Therefore, the resources taken into account in water balance should be reduced by the amount of instream flow. The hydrobiological criterion determines the minimum flow needed for supporting the life of flora and fauna in the aquatic environment. The environmental criterion determines the minimum flow to maintain the level balance of surface and underground waters within national parks, nature reserves and landscape protection areas. The fishing and angling criterion defines the minimum flow allowing fish to develop. The criterion of sport and water tourism determines the minimum water levels and corresponding flows making water tourism feasible [9-10, 43]. What also plays a significant role in the discussed hydrological issues, is the technical criterion relating to the technological capabilities of the water abstraction system which requires specialised equipment. Also, some of such limitations can be related to the reliable functioning of key economic areas, such as the power generation industry and water transport.

The total cost of water abstraction from these three intakes is described by the minimised objective function $OF : \boldsymbol{c} \cdot \boldsymbol{x} \rightarrow \boldsymbol{min}$. Thereby, the vector of cost factors assumes the form of:

$$\boldsymbol{c} = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T$$

determines the vector of decision variables. In order to optimize the decision variables, the *Solver* plugin in the *Excel* application was used (see tables 1 and 2). With the constraint conditions CO_j for $j = 1,\ldots,6$ the vector of optimal decision variables:

$$\boldsymbol{x}^* = \begin{bmatrix} x_1^* & x_2^* & x_3^* \end{bmatrix}^T = \begin{bmatrix} 66.(6) & 66.(6) & 66.(6) \end{bmatrix}^T$$

indicates that three identical water abstraction amounts from three intakes determine the minimum value of the objective function at the level of 666.67 monetary units. The solution obtained will remain unchanged when the unit cost of water abstraction from the P_1 intake remains in the range of 2 to 3.5 monetary units and the cost of water consumption from other sources does not undergo change. Therefore, for example, when the unit cost of water abstraction from this intake increases by 0.3 monetary units, the optimal $\boldsymbol{x}^*$ solution will not change, but the total cost will increase by 20 monetary units. Similarly, in the upper part of the table 2, we can observe acceptable changes of the remaining singular parameters of the total function of costs that allow the optimal $\boldsymbol{x}^*$ solution to remain unchanged and for calculation of the corresponding variable of the total cost of water abstraction.

Let us assume that the critical condition of the system forces a decrease in water abstraction, i.e. an infringement of the right-hand of CO_1. *Which other parameters of the model should be used and how should they be modified in order to make the reduction in water abstraction feasible and associated with optimal cost reduction?*

Due to the 6 constraint conditions with 3 decision variables, we conclude that the row vectors for the parameters of the left-hand side of the constraints are linearly dependent. In addition, due to the 5 binding constraint conditions and 1 non-binding one, shown in table 1, with 3 decision variables, we conclude that the solution to the primary task is unambiguous, but the dual solutions form a certain

Table 1. Optimum water abstraction for an urban and industrial agglomeration from three water abstractions – PM. A report on results in the Excel application

Objective Cell (Min)

Cell	Name	Original Value	Final Value
$E$10	Total Cost	10	666.6666667

Variable Cells

Cell	Name	Original Value	Final Value	Integer
$A$2	x1	1	66.66666667	Contin
$B$2	x2	1	66.66666667	Contin
$C$2	x3	1	66.66666667	Contin

Constraints

Cell	Name	Cell Value	Formula	Status	Slack
$E$4	Left Side Constraint 1	200	$E$4>=$G$4	Binding	0
$E$5	Left Side Constraint 2	200	$E$5<=$G$5	Not Binding	100
$E$6	Left Side Constraint 3	1800	$E$6<=$G$6	Binding	0
$E$7	Left Side Constraint 4	1000	$E$7>=$G$7	Binding	0
$E$8	Left Side Constraint 5	600	$E$8>=$G$8	Binding	0
$E$9	Left Side Constraint 6	200	$E$9>=$G$9	Binding	0

Table 2. A report on optimization sensitivity in the primary model for example in the Excel application

Variable cells

Cell	Name	Final Value	Reduced Cost	Coefficient Objective	Allowable Increase	Allowable Decrease
$A$2	x1	66.66666667	0	3	0.5	1
$B$2	x2	66.66666667	0	4	0.285714286	1
$C$2	x3	66.66666667	0	3	1E+30	0.25

Constraints

Cell	Name	Final Value	Price dual	Constraint R. H. Side	Allowable Increase	Allowable Decrease
$E$4	Left Side Constraint 1	200	0	200	0	1E+30
$E$5	Left Side Constraint 2	200	0	300	1E+30	100
$E$6	Left Side Constraint 3	1800	-0.6666667	1800	0	1.42109E-14
$E$7	Left Side Constraint 4	1000	1.6666667	1000	7.10543E-15	0
$E$8	Left Side Constraint 5	600	0.3333333	600	100	2.84217E-14
$E$9	Left Side Constraint 6	200	0	200	2.84217E-14	1E+30

subset of 2-dimensional space. Therefore, in order to carry out correct post-optimization, the dual model should be thoroughly examined in comparison with the *primary model* (*PM*) presented here [24, 52]. This is due to the ambiguity of the results of the dual optimization model and the resulting problematic nature of inference. This means that the infinite number of results of the dual optimization model poses problems in the interpretation of business and technical issues [55-56]. Obtaining many nodal solutions of the dual model on the basis of *PM* constitutes a certain problem. One method to obtain more nodal solutions is to change the order in which the constraint conditions are introduced [24]. However, it requires examination of a significant number of permutations out of as many as 6! = 720. It is helpful to optimize the dual model to *PM*. The *DM* takes the form of [12, 62]:

$$OF_{DM}: 200\cdot y_1+300\cdot y_2+1800\cdot y_3+1000\cdot y_4+600\cdot y_5+200\cdot y_6 \rightarrow max$$

$$CO_{DM1}:\ 1\cdot y_1+1\cdot y_2+9\cdot y_3+5\cdot y_4+2\cdot y_5+0\cdot y_6 \le 3$$

$$CO_{DM2}:\ 1\cdot y_1+1\cdot y_2+11\cdot y_3+6\cdot y_4+4\cdot y_5+2\cdot y_6 \le 4$$

$$CO_{DM3}:\ 1\cdot y_1+1\cdot y_2+7\cdot y_3+4\cdot y_4+3\cdot y_5+1\cdot y_6 \le 3$$

$$BR_{DM}:\ \ y_j \le 0 \ \text{ for } \ j=2,3\,,\ y_j \ge 0 \ \text{ for } \ j=1,4,5,6\,.$$

CO_2 and CO_3 are non-standard inequalities in *PM*, therefore, the corresponding dual decision variables in *DM* are non-positive [49, p. 104]. Since CO_2 is a non-binding constraint condition, the corresponding value of the dual decision variable y_2 is zero. The *sensitivity report* for *PM* obtained in the *Excel* application (table 2) the *dual price* column contains one of the optimal solutions in *DM* and it takes the form of:

$$\boldsymbol{y}_A^* = \left[y_{A_1}^*, y_{A_2}^*, y_{A_3}^*, y_{A_4}^*, y_{A_5}^*, y_{A_6}^*\right]^T = \left[0 \quad 0 \quad -0.(6) \quad 1.(6) \quad 0.(3) \quad 0\right]^T = 1/3\cdot\tilde{\boldsymbol{y}}_A^*.$$

And in the *sensitivity report* for *DM* obtained in the *Excel* application (table 3) the *final value* column contains another optimal solution in DM and takes the form of:

$$\boldsymbol{y}_B^* = \left[y_{B_1}^*, y_{B_2}^*, y_{B_3}^*, y_{B_4}^*, y_{B_5}^*, y_{B_6}^*\right]^T = \left[0.(6) \quad 0 \quad 0 \quad 0.(3) \quad 0.(3) \quad 0\right]^T = 1/3\cdot\tilde{\boldsymbol{y}}_B^*.$$

The limits of acceptable increase and decrease in the value $\boldsymbol{b}$ of the right-hand side of *PM* constraint conditions can be read respectively from the lower part of the table 2 or from the upper part of the table 3.

Table 3. Sensitivity report in the dual model obtained in the Excel application

Variable cells

Cell	Name	Final Value	Reduced Cost	Coefficient Objective	Allowable Increase	Allowable Decrease
$A$1	y1	0.66666667	0	200	0	0
$B$1	y2	0	100	300	1E+30	100
$C$1	y3	0	0	1800	1E+30	0
$D$1	y4	0.33333333	0	1000	0	200
$E$1	y5	0.33333333	0	600	100	0
$F$1	y6	0	0	200	0	1E+30

Constraints

Cell	Name	Final Value	Price dual	Constraint R. H. Side	Allowable Increase	Allowable Decrease
$H$3	Left Side Constraint 1	3	66.66666667	3	0.5	1
$H$4	Left Side Constraint 2	4	66.66666667	4	0.285714286	1
$H$5	Left Side Constraint 3	3	66.66666667	3	0.5	0.25

Table 4. The vector $\boldsymbol{y}_c^*$ of dual prices and the corresponding limits of the acceptable increase and decrease of the value of $\boldsymbol{b}$ the right-hand side of the constraint conditions in PM; based on the last matrix of the simplex method obtained according to [65]

Name	Price dual	Constraint R. H. Side	Allowable Increase	Allowable Decrease
Left Side Constraint 1	0	200	0	1E+30
Left Side Constraint 2	0	300	1E+30	100
Left Side Constraint 3	-1.333333	1800	0	0
Left Side Constraint 4	3	1000	0	0
Left Side Constraint 5	0	600	0	1E+30
Left Side Constraint 6	0.333333	200	0	0

A slightly different vector of dual prices (see table 4):

$$\boldsymbol{y}_C^* = \left[y_{C_1}^*, y_{C_2}^*, y_{C_3}^*, y_{C_4}^*, y_{C_5}^*, y_{C_6}^*\right]^T = \left[0 \quad 0 \quad -1.(3) \quad 3 \quad 0 \quad 0.(3)\right]^T = 1/3\cdot\tilde{\boldsymbol{y}}_C^*$$

was obtained using the application available on the website [65] for the *PM* model. In addition, from the last matrix of the simplex method obtained there, the inverse matrix $\boldsymbol{B}^{-1}$ to the so-called base was read (supplemented with additional variables, in accordance with the simplex method rule) and the limits of acceptable increase and decrease of the value of $\boldsymbol{b}$ the right-hand side of *PM* constraints contained in the table 4 were determined by solving an appropriate inequality:

$$\left[x_3^* \quad x_5^* \quad x_1^* \quad x_8^* \quad x_4^* \quad x_2^*\right]^T = \boldsymbol{B}^{-1}\left(\boldsymbol{b}+\Delta\boldsymbol{b}\right) \ge 0$$

due to the vector $\Delta\boldsymbol{b}$ of changes on the right-hand side of the constraints [49, p. 79].

For all $\alpha,\beta,\gamma \in <0,1>$ such as $\alpha+\beta+\gamma=1$ any convex linear combination $\alpha\cdot\boldsymbol{y}_A^*+\beta\cdot\boldsymbol{y}_B^*+\gamma\cdot\boldsymbol{y}_C^*$ of three vectors $\boldsymbol{y}_A^*,\boldsymbol{y}_B^*,\boldsymbol{y}_C^*$ constitutes also a vector of dual solutions. Moreover, for the column vector $\boldsymbol{b}$ of the right-hand side constraints *PM* in accordance with the Gale-Kuhn-Tucker theorem [20], we have:

$$\begin{aligned}\boldsymbol{c}\cdot\boldsymbol{x}^* &= \boldsymbol{b}^T\cdot\boldsymbol{y}_A^* = \boldsymbol{b}^T\cdot\boldsymbol{y}_B^* = \boldsymbol{b}^T\cdot\boldsymbol{y}_C^* = \\ &= \boldsymbol{b}^T\cdot\left[\alpha\cdot\boldsymbol{y}_A^*+\beta\cdot\boldsymbol{y}_B^*+\gamma\cdot\boldsymbol{y}_C^*\right] = \\ &= 1/3\cdot\boldsymbol{b}^T\cdot\left[\alpha\cdot\tilde{\boldsymbol{y}}_A^*+\beta\cdot\tilde{\boldsymbol{y}}_B^*+\gamma\cdot\tilde{\boldsymbol{y}}_C^*\right] = \\ &= 1/3\cdot\boldsymbol{b}^T\cdot\left[\tilde{\boldsymbol{y}}_B^*+\gamma\cdot\left(\tilde{\boldsymbol{y}}_C^*-\tilde{\boldsymbol{y}}_A^*\right)-(\alpha+\gamma)\cdot\left(\tilde{\boldsymbol{y}}_B^*-\tilde{\boldsymbol{y}}_A^*\right)\right] = \\ &= 1/3\cdot\boldsymbol{b}^T\cdot\left[\tilde{\boldsymbol{y}}_B^*+\gamma\cdot\boldsymbol{w}+\lambda\cdot\mathbf{v}\right] = \boldsymbol{b}^T\cdot\left[\boldsymbol{y}^*(\gamma,\lambda)\right] = 666.(6).\end{aligned}$$

where $\lambda=\beta-1=-(\alpha+\gamma)$, $\boldsymbol{w}=\tilde{\boldsymbol{y}}_C^*-\tilde{\boldsymbol{y}}_A^*$, $\boldsymbol{v}=\tilde{\boldsymbol{y}}_B^*-\tilde{\boldsymbol{y}}_A^*$, additionally $\boldsymbol{b}^T\cdot\boldsymbol{w}=0$ and $\boldsymbol{b}^T\cdot\boldsymbol{v}=0$ for $\boldsymbol{w}^T=[0\ \ 0\ \ -2\ \ 4\ \ -1\ \ 1]$ and $\boldsymbol{v}^T=2\cdot[1\ \ 0\ \ 1\ \ -2\ \ 0\ \ 0]$ is the orthogonality of vectors $\boldsymbol{w}$ and $\boldsymbol{v}$ to the vector $\boldsymbol{b}$ of the right-hand side of the constraints in *PM*, which is simultaneously a vector of coefficients of the dual model objective function. Note that both vectors $\boldsymbol{w}$ and $\boldsymbol{v}$ orthogonal to the vector $\boldsymbol{b}$ can be indicated as those rows of the matrix $\mathbf{B}^{-1}$ inverse to the base matrix **B**, which are orthogonal to the vector **b** and thus in the product:

$$\boldsymbol{B}^{-1}\cdot\boldsymbol{b}=\left[x_3^*\ \ x_5^*\ \ x_1^*\ \ x_8^*\ \ x_4^*\ \ x_2^*\right]^T=\left[66.(6)\ \ 100\ \ 66.(6)\ \ 0\ \ 0\ \ 66.(6)\right]^T$$

they create zero values of dual variables remaining in the base in the last matrix of the simplex method. This means that the previous effort to obtain one of the dual vectors $\boldsymbol{y}_B^*$ and vectors $1/3\cdot\boldsymbol{w}$ and $1/3\cdot\boldsymbol{v}$ can be simplified using only the last matrix of the simplex method [58].

Therefore, the optimal solution $\left[66.(6)\ \ 100\ \ 66.(6)\ \ 0\ \ 0\ \ 66.(6)\right]^{\mathrm{T}}$ in *PM* with the simplex method is connected by infinitely many optimal solutions $\boldsymbol{y}^*(\gamma,\lambda)$ in *DM*, which in the space $\boldsymbol{R}^6$ form a two-dimensional convex fragment of this space with a parametric form:

$$\boldsymbol{y}^*(\gamma,\lambda)=1/3\cdot\left(\tilde{\boldsymbol{y}}_B^*+\gamma\cdot\boldsymbol{w}+\lambda\cdot\mathbf{v}\right),$$

where for each pair of parameters (γ,λ) the vector $\boldsymbol{y}^*(\gamma,\lambda)$ meets BR_{DM}. Therefore, by solving the system of 6 inequalities:

$$\frac{1}{3}\cdot\left(\begin{bmatrix}2\\0\\0\\1\\1\\0\end{bmatrix}+\gamma\cdot\begin{bmatrix}0\\0\\-2\\4\\-1\\1\end{bmatrix}+\lambda\cdot 2\cdot\begin{bmatrix}1\\0\\1\\-2\\0\\0\end{bmatrix}\right)\begin{matrix}\geq 0\\ \leq 0\\ \leq 0\\ \geq 0\\ \geq 0\\ \geq 0\end{matrix},$$

we determine the range for parameters (γ,λ) in the form of a convex figure *T*. And so $T=\left\{(\gamma,\lambda)\in\boldsymbol{R}^2: 0\leq\gamma\leq 1,-1\leq\lambda\leq\gamma\right\}$. The range T has a trapezoidal shape (Fig. 2) with ABCD vertices, where point B coincides with the beginning of the $O\gamma\lambda$ coordinate system. The trapezium *T* contains the previously determined range of parameters $T_0=\left\{(\gamma,\lambda)\in\boldsymbol{R}^2: 0\leq\gamma\leq 1,-1\leq\lambda\leq-\gamma\right\}$ in the form of a triangle with vertices ABC, corresponding to a convex linear combination of only three vectors:

$$\boldsymbol{y}^*(0;-1)=\boldsymbol{y}_A^*,\ \boldsymbol{y}^*(0;0)=\boldsymbol{y}_B^*,\ \boldsymbol{y}^*(1;-1)=\boldsymbol{y}_C^*.$$

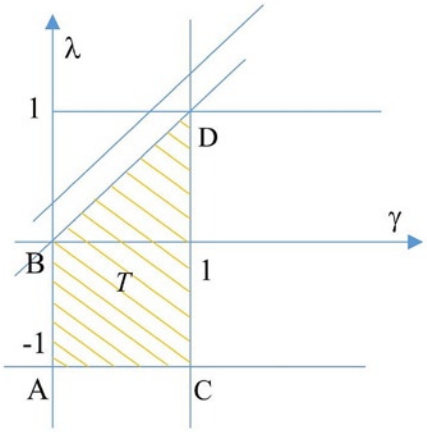


Fig. 2. The trapezium area T with ABCD vertices defines a range of parameters (γ,λ) for which the dual vector $\boldsymbol{y}^*(\gamma,\lambda)$ is allowed in DM

From here we obtain an additional nodal vector of dual prices:

$$\boldsymbol{y}^*(1;1)=\boldsymbol{y}_D^*=\left[y_{D_1}^*,y_{D_2}^*,y_{D_3}^*,y_{D_4}^*,y_{D_5}^*,y_{D_6}^*\right]^T=\left[1.(3)\ \ 0\ \ 0\ \ 0.(3)\ \ 0\ \ 0.(3)\right]^T=1/3\cdot\tilde{\boldsymbol{y}}_D^*$$

with unknown ranges of changes in the right-hand parameters of the *CO* corresponding to these prices.

For each $(\gamma,\lambda)\in T$ in *DM*, the function value $FC_{MD}(\gamma,\lambda)$ is constant and amounts to 666.67 and the vector $\boldsymbol{y}^*(\gamma,\lambda)$ is an acceptable vector, i.e. all BR_{DM} conditions are simultaneously satisfied and the constraints of the dual model are then binding conditions in *DM*. Moreover for $j=1,\ldots,6$ we obtain:

$$y_j^*(\gamma,\lambda)\in\min\left\langle\left(y_{A_j}^*,y_{B_j}^*,y_{C_j}^*,y_{D_j}^*\right),\max\left(y_{A_j}^*,y_{B_j}^*,y_{C_j}^*,y_{D_j}^*\right)\right\rangle.$$

And from here and with tables 2 – 4 because $y_2^*(\gamma,\lambda)\equiv 0$, therefore, changing the upper limit of water demand by reducing the right-hand side in CO_2 by 100 units or any increase therein will not change the optimal water abstraction plan $\boldsymbol{x}^*$ or the total cost.

Analysing tables 2 – 4, we infer that there are two alternative ways to reduce the total costs: a) by using value of the dual variable $y_{B_1}^*=0.(6)$ with only an apparently unacceptable decline of b_1 or b) using the doubled value of $y_{D_1}^*=1.(3)$ with the unknown allowed drop range of b_1.

Although $y_1^*(\gamma,\lambda)\in(0;1.(3))$, perhaps the largest decrease in the total costs in the amount of $0.(6)\cdot 25\approx 16.67$ monetary units by reducing the lower limit of water demand can be obtained by allowing declining of the right-hand side value 200 of the first constraint

condition CO_1 to be reduced by as many as supposedly 25 units when using the dual value $y^*_{B_1} = y^*_1(0,0) = 0.(6)$ read from the upper table 3 for *DM* when $[\alpha \quad \beta \quad \gamma] = [0 \quad 1 \quad 0]$, i.e. when $[\gamma \quad \lambda] = [0 \quad 0]$. But such a single variation of CO_1 may, however, contribute nothing because of the remaining 4 binding constraint conditions. Two tables inform about it: 2 and 4. This means that a single change on the right-hand side of CO_1, as in a typical sensitivity analysis, additionally requires changing the right-hand sides of other constraint conditions. As it turns out, such a proper suggestion results from the fact of unnatural blocking in table 3 both the increases and decreases in the parameters of the right side of CO_1 corresponding to a non-zero dual price. Similar objections concern the unnatural total blocking of changes in the right-hand parameters of constraints for several non-zero dual prices in tables 2 and 4.

Determining the change in the total cost of water abstraction $\Delta OF_{DM}(\boldsymbol{\Delta b})$, depending on the known vector $\boldsymbol{\Delta b} = [\Delta b_1 \quad \Delta b_2 \quad \Delta b_3 \quad \Delta b_4 \quad \Delta b_5 \quad \Delta b_6]^T$ of changes of the right-hand side of the constraint conditions in the *PM*, most simply requires re-launching appropriate software. But here an analyst faces the opposite problem, because he is looking for the whole vector of changes $\Delta \boldsymbol{b}$, the most favourable (most productive) for $y^*_{B_1}$. In addition, CO1 is not the only binding constraint condition but there are up to 5 binding constraints for the three decision variables. Therefore, having tables 2 - 4 a post-optimization question arises, not only a mathematical but also a managerial one: *Which changes Δb_j on the right-hand sides of the constraint conditions, which take into account various technical and economic problems of water abstraction and the principle of water flow in the river, must be accompanied by:*

a) $y^*_{B_1} = 0.(6)$ and a decrease in the demand for $\Delta b_1 = 25$ water units, i.e. savings $0.(6) \cdot 25 \approx 16.67$ of monetary units, and which

b) accompany twice bigger $y^*_{D_1} = 1.(3)$ with a decrease in demand unknown here Δb_1?

To answer the above questions, note that $OF = OF_{DM} = \boldsymbol{c} \cdot \boldsymbol{x}^* = \boldsymbol{b}^T \cdot \left[\boldsymbol{y}^*(\gamma,\lambda)\right]$, i.e. costs are fixed for any allowed parameter values γ and λ. But only for the properly selected fixed change vector $\Delta \boldsymbol{b}$ also new costs, i.e. the value $(\boldsymbol{b} + \Delta \boldsymbol{b})^T \cdot \boldsymbol{y}^*(\gamma,\lambda)$ is constant for any of the above-accepted values of parameters γ and λ. Hence, a vector $\Delta \boldsymbol{b}$ should be selected so that the cost change, i.e. the value $\Delta OF_{DM}(\Delta \boldsymbol{b};\gamma,\lambda) = (\Delta \boldsymbol{b})^T \cdot \boldsymbol{y}^*(\gamma,\lambda)$ would not depend on the choice of parameters γ and λ. The last equality indicates how the whole vector $\boldsymbol{y}^*(\gamma,\lambda)$ should be selected for the whole vector $\Delta \boldsymbol{b}$ used here. In particular, because also for any chosen parameters $(\gamma,\lambda) \in T$ we have:

$$\Delta OF_{DM}(\Delta \boldsymbol{b};\gamma,\lambda) = 1/3(\Delta \boldsymbol{b})^T \cdot \left[\tilde{\boldsymbol{y}}^*_B + \gamma \cdot \boldsymbol{w} + \lambda \cdot \boldsymbol{v}\right] = 1/3\left[(\Delta \boldsymbol{b})^T \cdot \tilde{\boldsymbol{y}}^*_B\right] = \Delta OF_{DM}(\Delta \boldsymbol{b};0,0),$$

when two conditions are met: $(\Delta \boldsymbol{b})^T \cdot \boldsymbol{w} = 0$ and $(\Delta \boldsymbol{b})^T \cdot \boldsymbol{v} = 0$. Therefore, the last orthogonality conditions describe this allowed proper way of selecting the constituents of the change vector $\Delta \boldsymbol{b}$, corresponding to operating the whole vector here $\boldsymbol{y}^*_B$.

Ad a) *What if, as in the classical sensitivity analysis, one should use only the one single component (read from table 3 of $y^*_{B_1} = y^*_1(0;0) = 0.(6)$ the vector $\boldsymbol{y}^*_B$, i.e. when:*

$$\Delta OF_{DM}(\Delta \boldsymbol{b};\gamma,\lambda) = \Delta b_1 \cdot y^*_{B_1} \ ?$$

Because the last equality is equivalent to the condition:

$$\sum_{j=1}^{6}\left(\Delta b_j \cdot y^*_{B_j}\right) = \Delta b_1 \cdot y^*_{B_1},$$

i.e. a requirement is created whereby in addition to $y^*_{B_1}$ the remaining non-zero components $y^*_{B_j}$ of the vector $\boldsymbol{y}^*_B$ do not affect the determined value of the change of the dual objective function. For this purpose, firstly, we accept $\Delta b_4 = \Delta b_5 = 0$. Secondly, for any chosen parameters γ and λ the equation must follow:

$$\Delta OF_{DM}(\Delta \boldsymbol{b};\gamma,\lambda) = \frac{1}{3} \cdot (\Delta \boldsymbol{b})^T \cdot \left[\tilde{\boldsymbol{y}}^*_B + \gamma \cdot \boldsymbol{w} + \lambda \cdot \boldsymbol{v}\right] =$$

$$= \frac{1}{3} \cdot \begin{bmatrix} \Delta b_1 \\ \Delta b_2 \\ \Delta b_3 \\ 0 \\ 0 \\ \Delta b_6 \end{bmatrix}^T \cdot \left(\begin{bmatrix} 2 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \gamma \cdot \begin{bmatrix} 0 \\ 0 \\ -2 \\ 4 \\ -1 \\ 1 \end{bmatrix} + \lambda \cdot \begin{bmatrix} 2 \\ 0 \\ 2 \\ -4 \\ 0 \\ 0 \end{bmatrix} \right) = \Delta OF_{DM}(\Delta \boldsymbol{b};0,0) = \frac{2}{3} \cdot \Delta b_1 \cdot$$

Hence conditions: $(\Delta \boldsymbol{b})^T \cdot \boldsymbol{w} = 0$ and $(\Delta \boldsymbol{b})^T \cdot \boldsymbol{v} = 0$ take the form of $-2\Delta b_3 + \Delta b_6 = 0$ and $2\Delta b_1 + 2\Delta b_3 = 0$. Then if $-25 \le \Delta b_1 \le 28$ and $-100 \le \Delta b_2$, the cost of water intake will change by $2/3 \cdot \Delta b_1$ when:

$$\Delta \boldsymbol{b} = \Delta b_1 \cdot [1 \quad 0 \quad -1 \quad 0 \quad 0 \quad -2]^T + \Delta b_2 \cdot [0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]^T.$$

This means that a decrease in total costs by $0.(6) \cdot 25 \approx 16.67$ monetary units through acceptable reduction of the minimum water requirement from 200 to 175 units can be obtained by concurrent (acceptable) increase of the right-hand side of the third (technical) constraint condition of CO_3 by 25 units and increase of the right-hand side of the sixth (also technical) condition restricting CO_6 by 50 units, with the appropriate limitation of the right-hand side of the second constraint condition CO_2 by 100 units. The correctness of the quoted analytical reasoning is confirmed by tables 5 and 6 for *PM* after changing 4 out of 6 parameters of the right-hand sides of the constraint conditions. This rightness exists despite the doubts that may have been raised in table 3 by zero values of acceptable growth for CO_3 and for CO_6. But the change ranges Δb_j of right-hand side changes for CO_3 and for CO_6 indicated in table 3 refer to the change of this single parameter in combination with a properly identified set of several parameters described by the vector $\Delta \boldsymbol{b}$. The correctness of the simultaneous changes carried out in the set of parameters $\Delta \boldsymbol{b}$ identified above was confirmed by conducting an analysis leading to table 5 and theoretical considerations in section 3.1.

Note that the reduction of the water supply minimums from the three sources, expressed as a reduction of the free expression in CO_1, will not result in decreasing the total water abstraction costs. It is only the collective action consisting in the simultaneous implementation of the three activities mentioned above that brings the intended effect of reducing the total cost of water consumption in the whole agglomeration from this system. The most effective measure is to maintain proportions in technical constraint conditions $\Delta b_3 / \Delta b_1 = -1$, $\Delta b_6 / \Delta b_1 = -2$ allowing for maximum reduction of water consumption $\Delta b_1 = 25$. The modification of the water abstraction system leads to the shut-off of the P_3 intake point during the minimum water demand and at the same time requires the modernization of the P_2 intake point to increase its ability to supply the agglomeration with water.

Table 5. Optimum water consumption for urban and industrial agglomeration from three water abstractions. PM after changing the 4 parameters of the right-hand sides of the constraint conditions. An answer report in the Excel application

Objective Cell (Min)

Cell	Name	Original Value	Final Value
$E$10	Total Cost	666.6666667	650

Variable Cells

Cell	Name	Original Value	Final Value	Integer
$A$2	x1	66.66666666	50	Contin
$B$2	x2	66.66666666	125	Contin
$C$2	x3	66.66666666	3.55271E-14	Contin

Constraints

Cell	Name	Cell Value	Formula	Status	Slack
$E$4	Left Side Constraint 1	175	$E$4>=$G$4	Binding	0
$E$5	Left Side Constraint 2	175	$E$5<=$G$5	Not Binding	125
$E$6	Left Side Constraint 3	1825	$E$6<=$G$6	Binding	0
$E$7	Left Side Constraint 4	1000	$E$7>=$G$7	Binding	0
$E$8	Left Side Constraint 5	600	$E$8>=$G$8	Binding	0
$E$9	Left Side Constraint 6	250	$E$9>=$G$9	Binding	0

Table 6. Sensitivity report in the Excel application for PM environment after changing 4 parameters of the right-hand sides of the constraint conditions

Variable cells

Cell	Name	Final Value	Reduced Cost	Coefficient Objective	Allowable Increase	Allowable Decrease
$A$2	x1	50	0	3	0.5	1
$B$2	x2	125	0	4	0.8	1
$C$2	x3	3.55271E-14	0	3	0.5	0.4

Constraints

Cell	Name	Final Value	Price dual	Constraint R. H. Side	Allowable Increase	Allowable Decrease
$E$4	Left Side Constraint 1	175	1.333333333	200	75	7.10543E-15
$E$5	Left Side Constraint 2	175	0	300	1E+30	125
$E$6	Left Side Constraint 3	1825	0	1800	1E+30	2.66454E-14
$E$7	Left Side Constraint 4	1000	0.333333333	1000	1.33227E-14	150
$E$8	Left Side Constraint 5	600	0	600	1.42109E-14	1E+30
$E$9	Left Side Constraint 6	250	0.333333333	200	75	1.42109E-14

Ad b) *What does the analogous adjusted sensitivity analysis procedure look like if only the highest first component* $y^*_{D_1} = y^*_1(1;1) = 1.(3)$ *of the vector* $\boldsymbol{y}^*_D$, *expressed as an amount of money, should be used, i.e. when:* $\Delta \mathrm{O}F_{DM}(\Delta \boldsymbol{b};\gamma,\lambda) = \Delta b_1 \cdot y^*_{D_1}$ *?*

Then the two-dimensional set of dual vectors should be parameterized to the form:

$$\boldsymbol{y}^*(\gamma,\lambda) = \boldsymbol{y}^*(1+\gamma_1, 1+\lambda_1) = 1/3 \cdot \left(\tilde{\boldsymbol{y}}^*_D + \gamma_1 \cdot \boldsymbol{w} + \lambda_1 \cdot \boldsymbol{v}\right)$$

constructed on the basis of vectors $\boldsymbol{y}^*_D$, $\boldsymbol{w}$, $\boldsymbol{v}$, where now D point of the trapezoid ABCD (Fig. 2) coincides with the beginning of the $O\gamma_1\lambda_1$ coordinate system. Then, as before, we accept $\Delta b_4 = \Delta b_6 = 0$ and solve the last of the following equations:

$$\Delta \mathrm{O}F_{DM}(\Delta \boldsymbol{b};\gamma_1,\lambda_1) = \frac{1}{3} \cdot (\Delta \boldsymbol{b})^T \cdot \left[\tilde{\boldsymbol{y}}^*_D + \gamma_1 \cdot \boldsymbol{w} + \lambda_1 \cdot \boldsymbol{v}\right] =$$

$$= \frac{1}{3} \cdot \begin{bmatrix} \Delta b_1 \\ \Delta b_2 \\ \Delta b_3 \\ 0 \\ \Delta b_5 \\ 0 \end{bmatrix}^T \cdot \left(\begin{bmatrix} 4 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} + \gamma_1 \cdot \begin{bmatrix} 0 \\ 0 \\ -2 \\ 4 \\ -1 \\ 1 \end{bmatrix} + \lambda_1 \cdot \begin{bmatrix} 2 \\ 0 \\ 2 \\ -4 \\ 0 \\ 0 \end{bmatrix} \right) = \frac{4}{3} \cdot \Delta b_1 \quad ,$$

where conditions $(\Delta \boldsymbol{b})^T \cdot \boldsymbol{w} = 0$ and $(\Delta \boldsymbol{b})^T \cdot \boldsymbol{v} = 0$ take the form of $-2\Delta b_3 - \Delta b_5 = 0$ and $\Delta b_1 + \Delta b_3 = 0$. Hence:

$$\Delta \boldsymbol{b} = \Delta b_1 \cdot [1 \quad 0 \quad -1 \quad 0 \quad 2 \quad 0]^T + \Delta b_2 \cdot [0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]^T .$$

This means that the decrease in the total costs by the amount of $1.(3) \cdot 20 \approx 26.67$ monetary units through reduction of the minimum water demand by only $\Delta b_1 = 20$ units from 200 to 180 units allowed by simultaneous (acceptable) increase of the right-hand side of the third (technical) condition limiting CO_3 by 20 units and decrease on the right of the fifth constraint condition CO_5 by 40 units, with the permissible drop of right-hand side of the second constraint condition CO_2 by 100 units. This time one should keep the proportions of changes $\Delta b_3 / \Delta b_1 = -1$ and $\Delta b_5 / \Delta b_1 = 2$.

The choice between the two solutions a) and b) requires additional information about the post-optimized water supply system.

3. Discussion of alternative scenarios

3.1. Two scenarios

We notice that like in a) zero values for dual prices $y^*_{B_3}$ and $y^*_{B_6}$ contained in tab. 3 indicate that only the technical binding constraint conditions of CO_3 and CO_6 together with the binding CO_1 with the exclusive exploitation of a non-zero value $y^*_{B_1} = 0.(6)$. And in b) the use of only $y^*_{D_1} = 1.(3)$ means a greater financial benefit with a smaller decrease in the minimum water demand, but at the same time requires a slightly different scope of intervention related to the technical binding constraint condition CO_3 and limitation of the parameter related to the right-hand side of another constraint condition CO_5, since $y^*_{D_3} = y^*_{D_5} = 0$. Therefore, the choice between these two solutions requires full consideration of information about the post-optimized water supply system, both included in the model and additional information from outside the model.

The fact of the above-mentioned necessity of simultaneous proportional change of several parameters of the water supply system results from the broader note below.

3.2. Interpretation of zero dual prices

Removal of the constraint condition in *PM* results in the deletion of the corresponding dual variable from *DM*. And leaving this dual variable in DM_1 equivalent to *DM* requires the recognition that in the optimal solution for DM_1 the value of this dual variable is zero. Also the non-binding constraint conditions in *PM* correspond to zero dual variable values in *DM*. Thus, the zeroing of a certain dual variable results from the already existing relaxation in the non-binding condition or indicates that to use the remaining non-zero values in the entire dual vector there is a possible necessity to make non-zero changes to the right-hand side of the constraint condition where the dual vector is zeroed. In other words, the zero optimal value of the dual variable corresponds to 1) the non-binding condition in *PM* or 2) the lack of this dual variable in *DM*. In turn, the lack of a dual variable in *DM* results from the removal of the effects of the corresponding constraint condition in *PM* which corresponds to the full relaxation of the constraints resulting from this constraint condition or due to the change in the value of the right-hand side of this constraint condition. This means that the zero value of the dual variable with a non-binding condition in *PM* corresponds to the possibility of independent change of the right-hand side of this constraint condition. And the zero value of the dual variable under the binding condition in *PM* corresponds to the possible necessity of the total change of the right-hand side of this constraint condition in the case of an attempt to change the right-hand side of another binding condition in *PM* with a corresponding non-zero value of the dual price.

3.3. Additional technical scenarios

Based on the last statement, we note that the decrease in water consumption (i.e. decrease in the right-hand side CO_1) may also be a consequence of six further scenarios changing: c) right-hand sides of binding conditions CO_3, CO_4, CO_5 in *PM* using the corresponding non-zero values $y^*_{A_3}, y^*_{A_4}, y^*_{A_5}$ of the dual price vector $\boldsymbol{y}^*_A$; d) the right-hand sides of the binding conditions CO_3, CO_4, CO_6 in *PM* using the corresponding non-zero values $y^*_{C_3}, y^*_{C_4}, y^*_{C_6}$ of the $\boldsymbol{y}^*_C$ dual price vector. The construction of these scenarios requires, among others analogous reparametrizing that shift the origin of the coordinate system to the vertices A and C of the trapezoid respectively. Each of these eight scenarios of reducing water consumption brings separate changes in the total cost of water consumed. These scenarios are the consequences of using, individually, eight of the 12 non-zero dual price values occurring for the nodes of the ABCD set. The four remaining non-zero values do not bring scenarios with the effect of reducing water consumption. The selection of the best scenario $\Delta\boldsymbol{b}$ out of 8 or a linear combination of the two best ones will be determined by a detailed analysis of, among others, financial, technical, biological, ecological and social consequences related to these scenarios. Among the considered constraint conditions, those that result from the principle of the integrity of water flow in the river either consistently do not change or should also be modified by legal, economic or additional technical steps.

For the vectors $\Delta\boldsymbol{b}$ located on a unitary sphere in $\mathbf{R}^6$ i.e. when:

$$|\Delta\boldsymbol{b}| = \sqrt[2]{\sum_{j=1}^{6}\left(\Delta b_j\right)^2} = 1,$$

changing the value of the objective function as a scalar product:

$$\Delta \mathrm{O}F_{DM}\left(\Delta\boldsymbol{b};\gamma,\lambda\right) = \left(\Delta\boldsymbol{b}\right)^T \cdot \boldsymbol{y}^*\left(\gamma,\lambda\right)$$

reaches the value of: i) maximum when vectors $\Delta\boldsymbol{b}$ and $\boldsymbol{y}^*\left(\gamma,\lambda\right)$ have identical direction and sense; (ii) minimum when vectors $\Delta\boldsymbol{b}$ and $\boldsymbol{y}^*\left(\gamma,\lambda\right)$ have the same direction and opposite sense; iii) zero when vectors $\Delta\boldsymbol{b}$ and $\boldsymbol{y}^*\left(\gamma,\lambda\right)$ are orthogonal. These facts determine the choice of scenarios with extreme or zero changes in the total costs of water consumed and require separate analysis, as well as the other 7 types of post-optimization analysis mentioned in chapter 1.5.

The significance of the result obtained here can be attested by the fact that so far some researchers, due to the analytical and interpretative difficulties, have generally discouraged the development of a similar analysis [8, p. 44]. Apart from *PM*, the authors support *DM* and software diversity to avoid the need to build a permutation algorithm supporting linear programming to generate all node solutions in the ABCD trapeze. The complexity of analytical calculations can be demonstrated by the lack of reproducibility on the basis of tables 5 and 6 return steps to table 3 allowing the return to the initial *PM*. And the post-optimization analysis made available in table 6, indicates the possibility of further recursive modification of the considered water abstraction system. But this requires repeating analytical steps similar to those described above. Of course, any changes to the right-hand constraint conditions in the model should reasonably reflect the operational needs of the agglomeration and the environmental potential of the hydrological system.

4. Conclusions

1. In its classic form, linear programming models for the exploitation of complex technical systems most often in the sensitivity reports contain ambiguous responses. Therefore, in section 1.4, some results referring to the dual price of water presented in the literature were critically evaluated. Because this study is devoted to eliminating the unreliability of standard idealistic sensitivity analysis procedures, which combine the value of each dual price only with the change of a single parameter of the right-hand constraint conditions, the proposed post-optimization methodology can be widely used by engineers, not only in the management of operational processes.
2. Inference of the market price of water based on an ambiguous dual price (and such situations almost always occur while examining technical issues [29]) leads to:
   a) lack of recognition of many scenarios for modification of the water supply system, and in turn, to a selection of a random scenario, which may require carrying out a difficult or even unachievable modernization of the water supply system.
   b) achieving lower savings in the functioning of the water supply network, than it is feasible in another way.

   Presenting the foundations of technical scenarios for a simplified example, the authors present the methodology of:

   ad. a) identification, through constraint conditions, of parts of the water supply system which are subject to change when implementing various technical scenarios,

   ad. b) finding various ways to obtain savings associated with the implementation of alternative scenarios, which allows for selecting the best system modification.
3. The methodology developed in the study allows for generating eight alternative technical water saving scenarios, of which two were outlined.
4. The ratios of $\Delta\boldsymbol{b}$ components determined in the work are used to simplify and improve the procedures of single and multiparametric linear programming [54]. Included among those

procedures are: post-optimization analysis with a functional dependence of changes in the parameters of the right-hand sides of constraint conditions, or changes in the coefficients of the objective function, which are analogous to the case of type 6 post-optimization in section 1.5. These problems are not only without descriptions available in the literature, but also become more complicated with a larger number of decision variables.

5. The study examines the use of dual vectors $\boldsymbol{y}^*$ and $\boldsymbol{w}$, $\boldsymbol{v}$ vectors orthogonal to **b** in determining effective proportional changes $\Delta\boldsymbol{b}$. Thus, it generalises the importance of vectors balancing in the transport model and orthogonal to **b** and allows for generalization of the dual price matrix method [24, 55-58] for models with inequality constraint conditions.
6. The methodology presented above can be directly adapted to the post-optimization of various operational processes of complex technical systems, e.g.
   a) in the water abstraction system for an urban agglomeration from several sources under various operating conditions [9-10, 57];
   b) to operating costs of other multi-source systems, e.g. supplying gas through the municipal gas network [34, 50];
   c) to operating costs of medical equipment and technical infrastructure [51];
   d) in the problem of improving the regionalization of emergency medical stations [39, 52];
   e) to the average system residence times in subsets of reliability states [31-32].
7. Currently, the authors are working on a full algorithmizing of the methodology presented here.

References

1. Abdy Sayyed M A H, Gupta R, Tanyimboh T T. Flow-based penalty in GA for opitmal design of water distribution network. Hydro-2017 International, L.D.College of Engineering, Ahmedabad, India, 2017; 1439 – 1446.
2. Agyei E, Munch D, Burger P. Application of Optimization Modeling to Water Resource Planning in East-Central Florida. Technical Publication SJ2005-2. St. Johns River Water Management District, Palatka, Florida, 2005 ftp://secure.sjrwmd.com/technicalreports/TP/SJ2005-2.pdf.
3. Amani N, Ali N M, Mohammed A H, Samat R A. Maintenance and management of wastewater system components using the condition index system, prediction process and costs estimation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2013; 15(2):161–168.
4. Arsham H. Construction of the largest sensitivity region for general linear programs. Appl. Math. Comput. 2007; 189(2):1435–1447, https://doi.org/10.1016/j.amc.2006.12.020.
5. Arsham H, Oblak M. Perturbation analysis of general LP models: a unified approach to sensitivity, parametric, tolerance, and more-for-less analysis. Math Comput. Model. 1990; 13(8): 79–102, https://doi.org/10.1016/0895-7177(90)90073-V.
6. Bradley S P, Hax A C, Magnanti T L. Applied mathematical programming. Addison-Wesley Publishing Company, Boston, 1977.
7. Brown L R. 2030, Who will feed China? World watch, 1994.
8. Busłowski A. Stabilność rozwiązania optymalnego zadania programowania liniowego. Wyd. Uniwersytetu w Białymstoku, 2000.
9. Chmielowski W. Zastosowania optymalizacji w gospodarce wodnej. Wyd. Politechniki Krakowskiej, 2005.
10. Chmielowski W. Management of Complex Multi-reservoir Water Distribution Systems using Advanced Control Theoretic Tools and Techniques. Springer-Verlag, 2013, https://doi.org/10.1007/978-3-319-00239-2.
11. Conradie B I, Hoag D L. A review of mathematical programming models of irrigation water values. Water SA 2004; 30(3): 287-292, https://doi.org/10.4314/wsa.v30i3.5076.
12. Dantzig G B, Thapa M N. Linear programming. 1: Introduction. New York, Berlin, Heidelberg: Springer 1997.
13. Datta B, Harikrishna V. Optimization Applications in Water Resources Systems Engineering. Research Journal of IIT Kanpur 2005:57-64.
14. Drobny N L. Linear programming applications in water resources 1. Journal of the American Water Resources Association 1971; 7(6): 1180-1193, https://doi.org/10.1111/j.1752-1688.1971.tb05055.x.
15. Easter K W, Rosegrant M W, Dinar A. Markets for water: Potential and performance. Springer Science & Business Media, 1998. https://doi.org/10.1007/b102479
16. Ehrgott M, Wiecek M M. Multiobjective programming. International Series in Operations Research and Science 2005: 667-722
17. European Commission. Report on the Review of the European Water Scarcity and Droughts Policy; European Commission: Brussels, Belgium, 2012.
18. Freire-González J, DeckerCA, HallJW. A Linear Programming Approach to Water Allocation during a Drought. Water 2018; 10(4): 363, https://doi.org/10.3390/w10040363.
19. Frizzone J A, Coelho R D, Dourado-Neto D, Soliani R. Linear programming model to optimize the water resource use in irrigation projects: An application to the Senator Nilo Coelho Project. Scientia Agricola 1997; 54(SPE): 136-148.
20. Gale D, Kuhn H W, Tucker A W. Linear programming and the theory of games, Activity Analysis of Production and Allocation, Ed. TC Koopmans, 1951.
21. Gass S I, Saaty T L. Parametric objective function (part 2)-generalization. Oper. Res. 1955; 3(4):395–401.
22. Gibbons D. The Economic Value of Water. Resources for the Future. Publisher: Resources for the Future, Washington, 1986.
23. Hanafizadeh P, Ghaemi A, Tavana M. Local perturbation sensitivity analysis of linear programming with functional relation among parameters. Int. J. Oper. Res. Inf. Syst. 2011; 2(1): 42–65, https://doi.org/10.4018/joris.2011010102.
24. Ignaciuk S. Post-optimization procedures for the selected transport issues in agricultural engineering. PhD thesis. Department of Applied Mathematics and Computer Science, Faculty of Production Engineering, University of Life Sciences in Lublin (in Polish), 2018.
25. Jacovkis P M, Gradowczyk H, Freisztav A M, Tabak E G. A linear programming approach to water-resources optimization. Zeitschrift für Operations Research 1989; 33(5): 341-362, https://doi.org/10.1007/BF01416081.
26. Jansen B, de Jong J J, Roos C, Terlaky T. Sensitivity analysis in linear programming: just be careful!. European Journal of Operational Research 1997; 101(1): 15-28, https://doi.org/10.1016/S0377-2217(96)00172-5.

27. Keshari A K, Datta B. Multiobjective Management of a Contaminated Aquifer for Agricultural Use. Water Resources Management 1996; 10(5): 373-395, https://doi.org/10.1007/BF00452952.
28. Klabjan D, Adelman D. Existence of optimal policies for semi-Markov decision processes using duality for infinite linear programming. Siam Journal on Control and Optimization 2006; 44(6): 2104-2122, https://doi.org/10.1137/S0363012903437290.
29. Koltai T, Terlaky T. The difference between the managerial and mathematical interpretation of sensitivity analysis results in linear programming. Int. J. Production Economics 2000; 65: 257-274, https://doi.org/10.1016/S0925-5273(99)00036-5.
30. Koltai T, Tatay V. A practical approach to sensitivity analysis in linear programming under degeneracy for management decision making. Int. J. Production Economics 2011; 131: 392-398, https://doi.org/10.1016/j.ijpe.2010.04.037.
31. Kołowrocki K, Soszyńska J. Optymalizacja procesów eksploatacji złożonych systemów technicznych. Problemy Eksploatacji 2010; 31-40.
32. Kolowrocki K, Soszynska-Budny J. Reliability and Safety of Complex Technical Systems and Processes: Modeling – Identification – Prediction- Optimization. Springer Science & Business Media, 2011, https://doi.org/10.1007/978-0-85729-694-8.
33. Kołowrocki K, Soszyńska J. Integrated safety and reliability decision support system. Journal of KONBiN 2011; 20(1): 17-28, https://doi.org/10.2478/v10040-012-0033-5.
34. Li J, Yan M, Yu J. Evaluation on gas supply reliability of urban gas pipeline network. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20(3): 471–477, https://doi.org/10.17531/ein.2018.3.17.
35. Lipinski P, Filipiak P, Rychlikowski P, Stanczyk J, Kajewska-Szkudlarek J, Lomotowski J, Konieczny T. Discovering weekly seasonality for water demand prediction using evolutionary algorithms. In Proceedings of the Genetic and Evolutionary Computation Conference Companion 2017; 07: 33-34, https://doi.org/10.1145/3067695.3082049.
36. Liu X, Chen X, Wang S. Evaluating and Predicting Shadow Prices of Water Resources in China and its Nine Major River Basins. Water Resour. Manag. 2009; 23(8):1467–1478, https://doi.org/10.1007/s11269-008-9336-7.
37. Madadpour S, Hanafizadeh P, Habibi R. Sensitivity Analysis of Asset Allocation: In The Presence of Correlation. Journal of Accounting, Finance and Auditing Studies 2018; 4(1): 151-166.
38. Malinowski J. A newly developed method for computing reliability measures in a water supply network. Operations Research and Decisions 2016; 26(4): 49–64.
39. Manikowski A. Optymalizacja rejonizacji Stacji Ratownictwa Medycznego z wykorzystaniem zmodyfikowanego zagadnienia transportowego na przykładzie Warszawy. Problemy Zarządzania 2011; 9(3): 227 – 241.
40. Mavrotas G. Effective implementation of the ε-constraint method in Multi-Objective Mathematical Programming problems. Applied Mathematics and Computation 2009; 213(2): 455-465, https://doi.org/10.1016/j.amc.2009.03.037.
41. McKee P W, Clark B R, Czarnecki J B. Conjunctive-use optimization model and sustainable-yield estimation for the Sparta aquifer of southeastern Arkansas and north-central Louisiana. Water-Resources Investigations Report 2003; 2003-4231, https://doi.org/10.3133/wri034231.
42. Nieuwoudt W L, Backeberg G R. A review of the modelling of water values in different use sectors in South Africa. Water SA 2011; 37(5): 703-710, https://doi.org/10.4314/wsa.v37i5.7.
43. Połomska J, Chmielowski W. Optimization of water management system operation while quality and quantity of water determined. Annals of Warsaw University of Life Sciences-SGGW. Land Reclamation 2012; 44(1): 15-24.
44. Romaniuk M. Optimization of maintenance costs of a pipeline for a V-shaped hazard rate of malfunction intensities. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20(1): 46–56, https://doi.org/10.17531/ein.2018.1.7.
45. Romaniuk M. Application of Markov chain and interest rate process for forecasting of costs of maintenance of pipelines. In: Simulation in Umwelt-und Geowissenschaften. Workshop Müncheberg. 2015: 147-156.
46. Romaniuk M. On simulation of maintenance costs for water distribution system with fuzzy parameters. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18(4): 514–527, https://doi.org/10.17531/ein.2016.4.6.
47. Saaty T L, Gass S I. Parametric objective function (part 1). Oper. Res. 1954; 2(3): 316–319.
48. Shahin A., Hanafizadeh P, Hladík M. Sensitivity analysis of linear programming in the presence of correlation among right-hand side parameters or objective function coefficients. Central European Journal of Operations Research 2016; 24(3): 563-593, https://doi.org/10.1007/s10100-014-0353-8.
49. Sioshansi R, Conejo A J. Optimization in Engineering: Models and Algorithms (Vol. 120). Springer, 2017, https://doi.org/10.1007/978-3-319-56769-3.
50. Słowiński B. Inżynieria eksploatacji maszyn. Wyd. Politechniki Koszalińskiej, 2014.
51. Sukharev M G, Karasevich A M. Reliability models for gas supply systems. Automation and Remote Control 2010; 71(7): 1415–1424, https://doi.org/10.1134/S0005117910070155.
52. Taha H A. Operations research: An introduction. Prentice Hall, 2007.
53. Wang Z. The linear programming model of water resource utilization in the irrigated area of Huaibei, Institute of Water Resource of Huai River Committee and Anhui Province, Bengbu, 2018
54. Ward J E, Wendell R E. Approaches to sensitivity analysis in linear programming. Ann. Oper. Res. 1990; 27(1): 3–38, https://doi.org/10.1007/BF02055188.
55. Wawrzosek J, Ignaciuk S. Use of extended sensitivity reports of linear programming in emergency medicinal services issues. Logistyka (Poznan) 2015; No 4 attachment CD ROM no 2 part 5: 8473-8481 (in Polish).
56. Wawrzosek J, Ignaciuk S. Selected postoptimization problems of the logistics of supplying urban-industrial agglomerations with water. Logistyka (Poznan) 2016; No 1 attachment CD ROM part 2: 425-431 (in Polish).
57. Wawrzosek J, Ignaciuk S. Dual model for classic transportation problem as a tool for dynamizing management in a logistics company. Econtechmod 2016;5(3): 95-100.
58. Wawrzosek J, Ignaciuk S. Postoptimization of the model of water supply for urban and industrial agglomeration. XLVIII Seminar of Applied Mathematics, ITM Web of Conferences 23, 00035, 1-5, https://doi.org/10.1051/itmconf/20182300035.
59. Wendell R E. A preview of a tolerance approach to sensitivity analysis in linear programming. Discrete Math. 1982; 38(1): 121–124, https://doi.org/10.1016/0012-365X(82)90178-9.

60. Wendell R E. Using bounds on the data in linear programming: the tolerance approach to sensitivity analysis. Math. Program. 1984; 29(3): 304–322, https://doi.org/10.1007/BF02591999.
61. Wendell R E. The tolerance approach to sensitivity analysis in linear programming. Manag Sci. 1985; 31(5): 564–578, https://doi.org/10.1287/mnsc.31.5.564.
62. Wendell R E. Sensitivity analysis revisited and extended. Decision Sciences 1992; 23(5): 1127-1142, https://doi.org/10.1111/j.1540-5915.1992.tb00439.x.
63. Wondolowski F R. A generalization of Wendell's tolerance approach to sensitivity analysis in linear programming. Decis Sci 1991; 22(4): 792–811, https://doi.org/10.1111/j.1540-5915.1991.tb00365.x.
64. World Bank. High and Dry: Climate Change, Water, and the Economy; World Bank: Washington, DC, USA, 2016.
65. http://www.maslowski.pl/index.php?id=simplex
66. Zio E. An introduction to the basics of reliability and risk analysis. World Scientific Publishing Co. Pte. Ltd, 2006.

Jacek WAWRZOSEK
Szymon IGNACIUK
Andrzej BOCHNIAK
Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin
ul. Głęboka 28, 20-612 Lublin, Poland

E-mails: jacek.wawrzosek@up.lublin.pl, szymon.ignaciuk@up.lublin.pl, andrzej.bochniak@up.lublin.pl

Article citation info:
SIKORA M, SZCZYRBA K, WRÓBEL Ł, MICHALAK M. Monitoring and maintenance of a gantry based on a wireless system for measurement and analysis of the vibration level. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 341–350, http://dx.doi.org/10.17531/ein.2019.2.19.

Marek SIKORA
Krzysztof SZCZYRBA
Łukasz WRÓBEL
Marcin MICHALAK

MONITORING AND MAINTENANCE OF A GANTRY BASED ON A WIRELESS SYSTEM FOR MEASUREMENT AND ANALYSIS OF THE VIBRATION LEVEL

MONITOROWANIE I UTRZYMANIE SUWNICY BRAMOWEJ NA PODSTAWIE BEZPRZEWODOWEGO SYSTEMU POMIARU I ANALIZY POZIOMU DRGAŃ*

The paper describes a system for monitoring and diagnosing a gantry. The main goal of the system is to acquire, visualize and monitor vibration levels of the gantry crucial elements. The system is also equipped with a computing and analytical part which enables predictive maintenance related to the vibration level assessment. The system architecture can be used in other applications too, i.e. those which require a wireless network of vibration sensors to carry out diagnostic tasks.

***Keywords**: vibration sensor, monitoring system, maintenance, gantry, predictive maintenance, data mining, trend analysis.*

W artykule przedstawiono system monitorowania i diagnostyki suwnicy bramowej. Głównym zadaniem systemu jest akwizycja, wizualizacja i monitorowanie poziomu drgań newralgicznych elementów suwnicy. System wyposażony jest również w część obliczeniowo-analityczną, umożliwiającą realizację zadań predykcyjnego utrzymania ruchu (ang. predictive maintenance) związanych z oceną poziomu drgań. Architektura systemu umożliwia wykorzystanie go również do innych zastosowań, w których dla realizacji zadania diagnostyki wymagana jest bezprzewodowa sieć czujników drgań.

***Słowa kluczowe**: czujnik drgań, system monitorowania, utrzymanie, suwnica, utrzymanie predykcyjne, eksploracja danych, analiza trendu.*

1. Introduction

Today's IT systems for monitoring and diagnosing industrial facilities offer extensive data storing and processing possibilities. For several years, in accordance with the ISA95 standard [11], a distinct specialization has been forming which allows for distinguishing the following systems: measurement data acquisition systems, online monitoring and process visualization systems (HMI – Human Machine Interface; SCADA – Supervisory Control and Data Acquisition), process execution systems (MES – Manufacturing Execution System), and ERP (Enterprise Resource Planning) systems which ensure comprehensive support of enterprise management.

Data analysis is a dynamically developing area of IT systems application. New analysis-based functionalities support suppliers and manufacturers in gaining competitive advantage on the market. There are two groups of applications related to the analysis of data collected by SCADA and MES systems: CMMS (Computerized Maintenance Management Systems) and DAP (Data Analytics Platforms). In the ISA95 architecture, these applications should be placed between MES and ERP. DAP applications provide tools and methods to build diagnostic models of devices or to analyze the collected data with a view to improving manufacturing processes.

Modern organizations have been increasingly more aware of the necessity to employ data scientists who develop and maintain diagnostic modes used, for example, in Predictive Maintenance (PdM) [18].

The paper presents a gantry monitoring and diagnosing system. The diagnosed gantry is an element of a raw material transportation line for a converter steel plant in a foundry. The gantry is one of the key elements of the production line, therefore it is necessary to monitor its operations and to identify, at least with some approximation, a moment in time when the gantry operation starts to be dangerous or impossible.

The source of data involves a wireless network of vibration sensors. The data are processed by a dedicated SCADA system, while the analytical part is based on R analytical environment [21].

Section 2 of the paper features examples of related works about systems dedicated to machine diagnosing based on historical data analysis, particularly that of the vibration level. Section 3 presents the monitored object and the issue of its diagnostics. Section 4 includes a description of the hardware and software parts of the monitoring system. Section 5 features the data analysis process and the resulting diagnostic procedure whose practical application is presented in section 6. Section 7 includes conclusions and a system development concept.

2. Related works

Computer systems dedicated to defining, monitoring and predicting diagnostic states of machines can be categorized from different points of view. One of them is the division into frameworks [13] and

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

systems dedicated to a specific solution or a group of solutions (e.g. [17]). Methodologies for performing diagnostic tasks, maintaining machines and devices [29], predictive maintenance [14,26], or diagnosing damages [20,25,10] allow for an increasingly more efficient use of equipment owned by the organization.

In the literature one can find many publications about diagnostics. Depending on the degree of machine complexity, its structure, intended use, and operation mode, the diagnosing process is carried out using diverse research methods. For example, the diagnostics of combustion engines in terms of their wear is very often performed based on the spectrometric analysis of the composition of oil taken from these engines [12]. Some works apply an analytical approach to the modelling of the oil combustion process in pistons in order to describe the relation between the content of pollutants and the passed time [16].This example illustrates two main approaches to the construction of diagnostic models: one is based on the analysis of the collected measurement data or other data related to the machine work monitoring (e.g. thermal images [9]), the other – on modelling/simulation of the diagnosed object operations.

Due to the topic of the paper, it is particularly important to refer to works in which the vibrations generated by a working machine are the input signal being the basis for the diagnosis. In the work [6] the authors monitored the behavior of a plunger pump in a truck crane through the acquisition of its vibrations in three dimensions. The spectra of the machine operations in a correct state were known, along with its operations spectra in five incorrect diagnostic states. The classification of the pump's current state was conducted by means of Bayesian reasoning whose parameters were tuned with the use of the particle swarm optimization algorithm. In the work [8] the authors described the use of advanced features extraction from the vibration signals describing the work of a commutator motor in the range of frequency.

In order to analyze data from vibration sensors, the wavelet transform is used. The works [19,27] feature a review of this methodology application for machine diagnosing. The method was used particularly to detect failures of an induction engine rotor [28], failures in transmission networks [23] or damages of bearings [15].

The vibration amplitude waves are usually not stationary. For this reason, their analysis is often performed with the use of the so called spectral kurtosis [1]. The method has been used to detect failures in wind turbines drives [7] or to assess the condition of bearings [5].

Samuel's work [22] presents a comprehensive approach to the issue of building diagnostic models on the basis of the vibrations analysis. The work features many variants of using vibrations measurement (from simple statistical indicators of time and frequency, through wavelet transform, use of neural networks, to mathematical modeling) to construct diagnostic models of machines.

In works [2, 3] Bartelmus and others point out that, while analyzing the vibrations, it is necessary to take into account the load under which the machine (e.g. conveyor or wheel excavator) works at a given moment. The component of vibrations induced by the changing load of the machines is isolated and deducted from the originally observed series of vibrations. This allows, in special cases, to obtain proper input data for the diagnostic process.

3. Object and goal of monitoring

The monitored object involves two gantries, each with a lifting capacity of 500 tons, working in a steel plant. The gantries are elements of a raw material transportation line in a converter plant. Each gantry takes full ladles of crude steel from transportation trolleys, transports them towards the converter mouth, pours the crude steel inside by tilting the ladle and, eventually, puts the empty ladles back onto the trolleys.

Figure 1 features a diagram of a drive system responsible for lifting and lowering the ladle. The system is the place where vibration sensors are located. Each gantry has two drives – left (L) and right (R). The drives are connected by rollers with toothed gears which reduce the rotary speed of the roller in the main lifting mechanism. In addition, the figure presents the location of vibration sensors which are placed on the rollers (1, 6), pinion bearings (2, 5) and on the supporting beam. Sensors 3 and 4 are, in fact, auxiliary ones and their task is to monitor the vibration level of the background. They may serve as reference points for other sensors. For the diagnostic system (particularly for the data preparation phase) it is important to note that for different gantries (even gantries of the same type) it is very difficult to locate vibration sensors exactly in the same place.

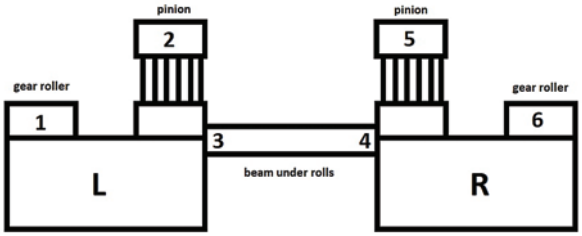


Fig. 1. Drive system of a gantry with vibration sensors locations

Accurate operation of the gantries is one of key elements to maintain steel production continuity. Their damage results in the failure to supply raw steel from the blast furnace to the converters. This, in turn, prevents the supply of raw material to the successive production stages (continuous casting and rolling) and, in extreme cases, leads to the necessity to extinguish the furnace.

The gantry works in difficult environmental conditions, in a production hall (Figure 2). The gear mechanism operates in the presence of penetrating metal oxide dust. The impact of dust can be observed by the rising vibration level of the whole gantry, as the gear vibrations are transmitted into other elements, including the structural elements of the hall. In an extreme situation, the work of the gantry becomes impossible or the operators are forced to reduce the engine gears, which has an unfavorable impact on the production efficiency. In such situations the gear needs a major renovation.

The key diagnostic issue is to analyze the vibration level and to determine, on this basis, whether further normal operation will be possible within the next several days. The diagnostic procedure should make it easier for the maintenance personnel to make decisions about

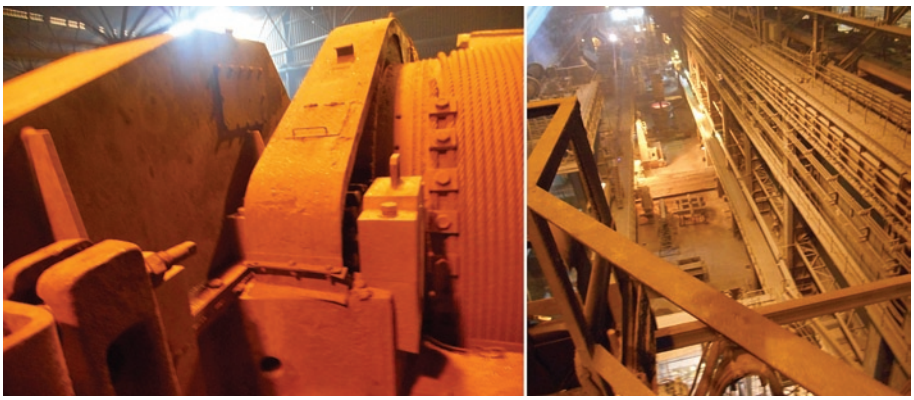


Fig. 2. Part of the housing of the gantry drive and the view of the production hall from the operator's stand

renovations or interim maintenance works that were not included in the renovation plan, such as oil change in the gear or control of capstan teeth.

From the practical point of view, the diagnostic procedure should be parameterizable as the impossibility to install vibration sensors in the same places each time leads to a situation in which vibration levels that are non-acceptable in certain operating conditions could prove acceptable in others. In the analysis in section 5, it is important to note the crucial role of information from the so called event log where all maintenance and renovation works performed on a given object are recorded. This information was used to develop the presented method and to assess its efficiency. Thanks to this, it was possible to refer the vibration level to the performed renovations and, in particular, to not treat sudden drops of the vibration level as data errors.

4. Monitoring system

The main goal of the monitoring system is to acquire, monitor and analyze the level of vibrations registered by sensors. Apart from ongoing visualization, tracking the level and the changing dynamics of the vibrations allows one to determine the moment of warning and alarm levels of vibrations.

The main source of data for the presented system involves vibration sensors installed in crucial points of the monitored objects. The practice shows that these points are often located in places with difficult access. Movable parts of machines are often monitored as well.

These conditions determined the following main features of the described system:

- the sensors are installed permanently in measurement points,
- the sensors are wireless (supplied by batteries),
- the sensors are resistant to so-called temperature surges as they may be periodically exposed to flames,
- the sensors are dustproof which makes them protected against metal dust,
- communication with the sensor network is provided via radio, with the use of a network of transceivers,
- the IT part of the system is created in the client-server architecture:
  - the server records, stores and gives access to measurements collected in the database,
  - the SCADA client application enables defining managers' panels and cockpits (allowing for visualizing device operations, including the registered vibration level) and generating summary reports,
- the analytical server, which makes use of the R environment [21], is an integral part of the system; the server is used for launching diagnostic procedures related to data analysis and provides information about the current state of the device.

Figure 3 depicts the architecture of the system dedicated to the gantry monitoring. The Server computer features a communication software (for data acquisition), a database server (PostgreSQL) and the R analytical environment. The Maintenance station computer represents the client's application (there may be many of such applications). In addition, the server can provide access to data to other applications, e.g. MES or ERP systems.

Further part of this section contains a brief description of the vibration sensor selected to monitor the gantry, as well as of the selected elements of the client application.

4.1. Vibration sensor

Vibrations are understood as oscillations of physical bodies of a certain weight in an assumed reference system. Vibrations can be described by one of the following three parameters: dislocation, velocity or acceleration. To diagnose vibrations it is enough to measure one of the three parameters – the remaining ones can be obtained based on the known physical relations. The easiest way to measure the level of higher-frequency vibrations is to measure acceleration, because the amplitude of this size is measureable within a wide range of the vibration spectrum.

From the point of view of diagnostic requirements, the sensor should have the best possible measurement accuracy and resolution, a considerable bandwidth and measurement linearity in this band.

Table 1 features a list of sample sensors which are analyzed for use as sources of measurement data (the first sensor is manufactured by Banner Engineering, the second – by KCF Technologies, the third – by Swift Sensors, the fourth – by EMAG-Serwis, and the fifth – by Somar).

The SD-VSN-2 sensor was one of two considered for use. However, due to communication problems encountered during the system implementation in the production hall, eventually the other sensor was selected: WS-VT1 created by Somar S.A. As seen in Table 1, the selected sensor has all the features required for the considered application. In particular, its communication band is sufficient for accurate data transmission in the production hall. Additionally, the sensor has been integrated with a radio transmitter.

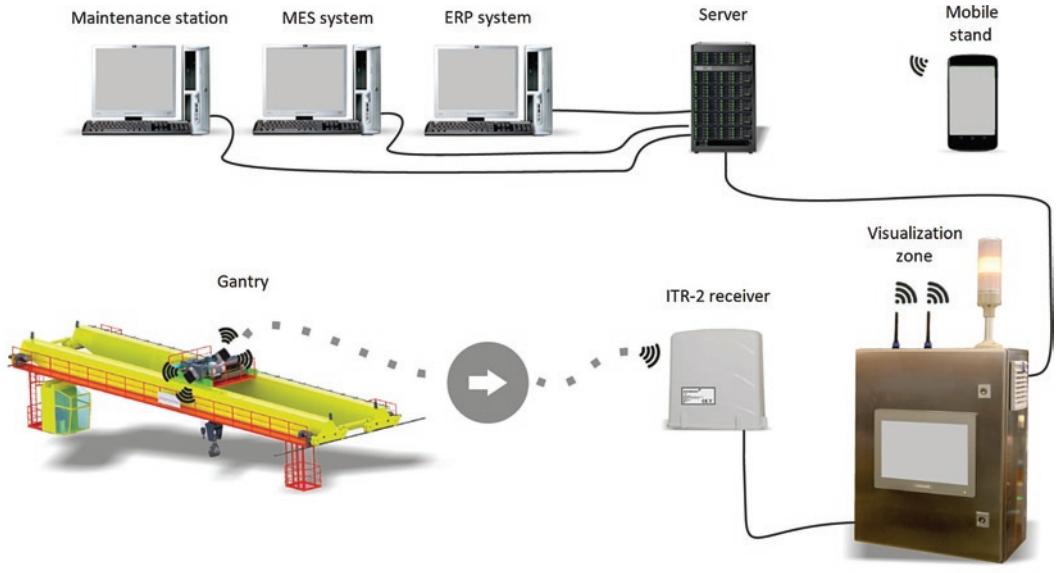


Figure 3. System for monitoring vibrations and diagnosing machines with the use of wireless vibration sensors

WS-VT1 allows for measuring acceleration in three axes, in the ±2g/±4g/±8g/±16g ranges, and in the band up to 2.5 kHz. In addition, the sensor allows one to measure ambient temperature, which is performed by a digital temperature converter. The converter makes measurements with a 12-bit resolution and accuracy of 1°C within the range of –55÷+85°C. Both converters can be put in sleep mode in order to save energy. The measurements are managed by an energy-saving, 16-bit micro-controller equipped with a radio transceiver working in the 860 MHz band. The whole

Table 1. List of sensors enabling acceleration measurement

Sensor type	QM42VT1	SD-VSN-2	201R	Pantera	WS-VT1
Resistance to difficult environmental conditions (humidity, dustiness, temperature surges)	-	+	-	-	+
Ability to work in atmospheres with gas and dust explosion hazards	-	-	-	-	+
Communication band	2.4GHz	2.4GHz	860MHz	None	860MHz
Time of work with battery backup	+	+	+	-	+

element is supplied by a lithium-ion ½ AA battery with a 1.2 Ah capacity.

In its basic working mode the sensor is periodically woken up from the sleeping mode so that the acceleration, temperature and battery charge level could be measured. The sensor sends the following information via radio:

- its identification number,
- RMS and maximal value of vibrations,
- temperature value,
- battery charge level.

Putting the sensor to sleep and then waking it up, performed periodically, is not the only way to carry out the measurements. The developed embedded software allows for changing the frequency of the measurements' acquisition, depending on the observed vibration level. This solution allows one to save energy during machine downtimes (lack of or low level of vibrations) and to raise the frequency of data acquisition when the registered vibrations are above the user-predefined boundary value. Theoretically, the sensor can be specially programmed to raise the frequency of measurements while crossing the successive, increasingly higher, vibration levels.

4.2. Maintenance station

The sensors and transmitters are data sources for the application which monitors the state of the device. Data from the sensors are collected with the use of radio transceivers. The transceivers work in Ethernet, in a 10Base-T or 100Base-Tx standard. Their basic task is to receive measurement data from the sensors which are within their range. These data are supplemented with the indication of the receiver signal strength and stored in suitable database tables.

The database server collects the measurement data and grants access to them to the maintenance station application, being a SCADA program, and to the R environment server where the data analysis is performed.

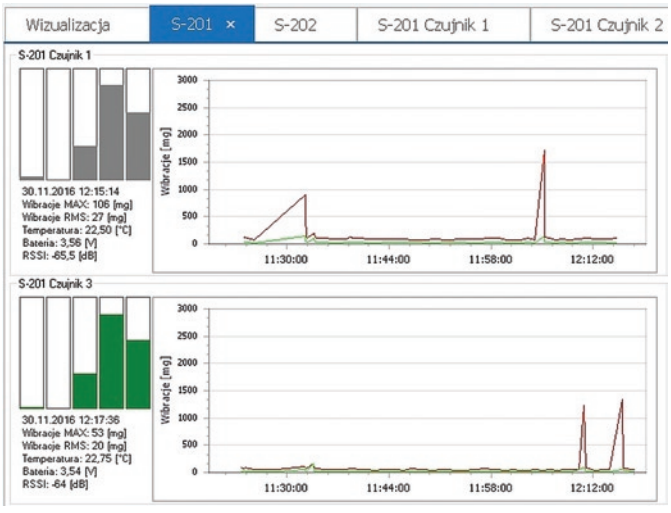


Fig. 4. Visualization of measurements for a group of sensors – current measurements are visible (bar diagram), along with historical measurements (in the figure these are measurements from the last hour)

The maintenance station application enables the following:

- to present current values of vibrations, temperature and battery charge level,
- to configure and set accurate, warning and alarm vibration levels (in general, the parameters of a diagnostic procedure) as well as the levels of temperature and battery charge,
- to present current and legacy data on diagrams (Figure 4. Visualization of measurements for a group of sensors – current measurements are visible (bar diagram), along with historical measurements (in the figure these are measurements from the last hour)),
- to present selected results of data analyses performed by the analytical part of the system,
- to generate periodic reports (shift-, day- and long-term reports).
- In addition, the application makes it possible to prepare a log of failures, renovations and maintenance works. As it was mentioned before, this is a very important feature of the system which enables to use information from the log during data analysis. In particular, the information about replacing an element of the gantry results in resetting the information about non-failure working time of this element.

5. Diagnostic procedure

5.1. Data preprocessing

Two gantries were monitored, each via a set of six sensors. The analysis used information about the registered vibration values – maximum (VibrMAX) and effective (VibrRMS) value.

The first stage of data processing was based on the preparation of two data sets (one for each gantry) in which each successive line included information about the maximum and effective level of vibrations registered by each of the six sensors in successive seconds (Table 2).

In a situation in which no measurements were recorded from a given sensor during a given second and the gap in data transfer was shorter than 30 seconds, the lacking data were supplemented with the last registered value. The period of 30 seconds resulted from the transmission configuration which forced the transfer of data from the sensors at least every 30 seconds. Gaps longer than 120 seconds were treated as missing data.

The diagnostic procedure is based on the analysis of the variables vector which describes a single work cycle of the gantry. The work cycle should reflect the operations of the gantry, i.e. lifting an element, transferring it to another place and putting it down. As there was no information available about the work of elements which are responsible for lifting and moving, the cycle was defined as a period of time during which the sensors registered increased levels of vibrations. In addition, the following additional assumptions were adopted to identify work cycles:

Table 2. Set of basic data describing the operations of each gantry

Date/time	VibrRMS1	...	VibrRMS12	VibrMAX1	...	VibrMAX12
2017-05-05 15:12:00	28	...	29	71	...	74
2017-05-05 15:12:01	28	...	30	70	...	72

- the basis for cycle identification included measurements for which the registered value of effective vibrations level exceeded 150 mg, lower values were treated as a temporary stoppage of the gantry,
- identified cycles shorter than 70 seconds were not taken into account,
- if the stoppage was shorter than 60 seconds, two successive cycles were combined.

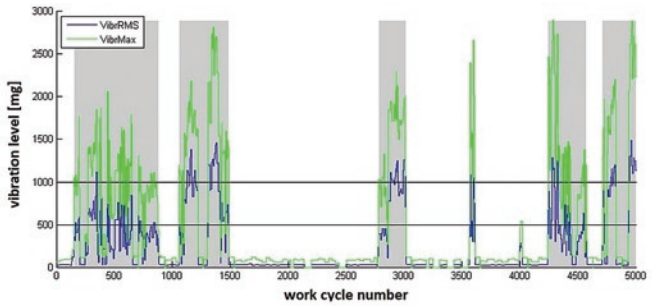


Fig. 5. Sample identification of the gantry work cycles – gray areas depict identified work cycles

For each of the identified work cycles a number of parameters were calculated. The parameters describe the course characteristics of VibrMAX and VibrRMS variables. The calculations were performed along the entire range of the registered values and for filtered values of these features in the ranges 2,000-2,500 mg, 2,500-3,000 mg, and 3,000-3,500 mg. For VibrMAX and VibrRMS, the following were determined:

- minimal value,
- average value and its 5th(P5) and 95th (P95) percentile,
- the number of measurements above 3,500mg,
- harmonic analysis results of the signal of the effective vibrations value: the frequency of the first 10 harmonics (including the constant) and their amplitude,
- harmonic analysis results of the signal of the vibrations maximal value: the frequency of the first 10 harmonics (including the constant) and their amplitude,
- 1st quartile (Q1),
- median (Q2),
- 3rd quartile (Q3),
- maximal value,
- interquartile range (Q3-Q1),
- percentile range (P95-P5),
- mode of registered values.

For each sensor the final effect of the conducted analysis was a vector of 115 features characterizing each work cycle of the gantry.

The first analysis dealt with checking whether, on the basis of the above parameters, it is possible to clearly recognize the work cycles in which the gantry is loaded by a full ladle. The tests were conducted with a very limited testing set containing 48 runs of the gantry of which the personnel qualified 24 as "full runs" and 24 as "empty runs." On the basis of the available sample runs and with the use of the RPART algorithm [24] a decision tree was generated (see: Figure 6).

The features based on Fourier's analysis allowed for obtaining a tree with the following classification ability: 0.792 accuracy of the empty class and 0.958 accuracy of the full class. Unfortunately, after launching the classifier on the data that had been collected for a long time, it was observed that most samples (5,467 out of 6,439) were classified as empty. In spite of the fact that the samples did not have class labels, such a classification result was recognized as bad, because according to the specification of the gantry operations the distribution of samples should be close to the regular one. Eventually, during a further analysis of the gantry work cycles no differentiation was introduced between empty and full runs.

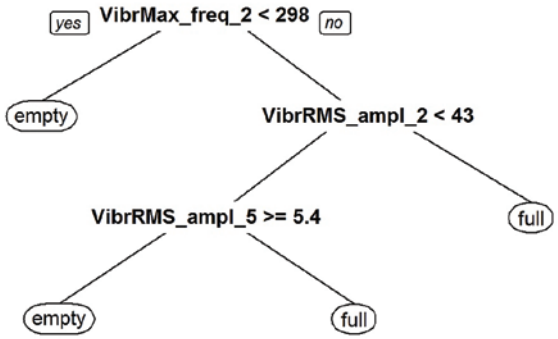


Fig. 6. Decision tree to identify full and empty gantry runs

In the next phase of testing, the diagnostic procedure was determined using data from one gantry. The data from the other gantry were used to verify the conducted tests.

In order to identify the features with the highest diagnostic significance, a feature selection method was used. The method is based on ranking the features of a classifier made with the use of the random forest methodology [4]. The vector of successive cycles from the period of 4 months of the gantry operations (a total of 11,818 cycles) was labelled so that the cycles registered in the period up to two weeks directly before the failure received a 0 label (operational gantry), while the cycles registered within the period of two weeks before the failure up to the failure point were labelled 1 (pre-failure gantry state). This way, a data set was obtained in which 9,275 samples pointed to class 0 and 2,543 indicated class 1. On a data set prepared for each vibration sensor separately, a classifier created of 1,000 decision trees was trained. The main criterion of the node assessment involved the Gini index [4]]. The following quality of classification was obtained for the successive sensors:

- sensor 1 – specificity 0.984, sensitivity 0.898,
- sensor 2– specificity 0.991, sensitivity 0.722,
- sensor 3 – specificity 0.986, sensitivity 0.828.

Based on the analysis of trees creating the random forest, the relevance of each feature was assessed from the point of view of its importance in the developed classifier. Based on the random forest, the importance of feature X was determined according to the following formula:

$$Imp(X)=\frac{1}{M}\sum_{T}\ \sum_{t\in T:v(t)=X}\frac{N_t}{N}\cdot\left\{G(t)-\left[\frac{N_{t_L}}{N_t}G(t_L)+\frac{N_{t_R}}{N_t}G(t_R)\right]\right\} \quad (1)$$

where:

M – the number of all T trees in the random forest (in this case 1,000),

t – the node of tree T ,

t_L, t_R – left and right child of node t,
$v(t)$ – a feature occurring in the condition of node t,
N – the number of all observations in the training set (in this case 11,818),
N_t – the number of observations in node t,
$G(t)$ – the Gini index for node t; $G(t)=1-\left(\frac{N_t^+}{N_t}\right)^2-\left(\frac{N_t^-}{N_t}\right)^{-2}$, where N_t^+ and N_t^- are, respectively, the number of positive samples (class 1) and the number of negative samples (class 0) in node t.

The higher the value of the measure (1), the better the recognition is of diagnostic states by the feature describing the gantry work cycle. This way, it is more eligible as a diagnostic feature. If the feature did not occur in any trees of the random forest, its importance was 0.

The features were ranked according to measure (1) for each sensor separately. This way, three feature ranks were obtained. The final rank of a given feature was an arithmetic average of the positions the feature held in each rank.

The first five features in the rank were used for further work on the diagnostic procedure. These included the following:

- VibrMax2000_Q3 – the value of the 3rd quartile of the VibrMAX feature; only VibrMAX>2000 values were taken into account in the calculations,
- VibrMax3000_Q3 – as above,
- VibrMax2500_Q3 – as above,
- VibrMax_P95 – the value of the 95th centile of the VibrMAX feature;
- VibrMax2000_P95 – the value of the 95th centile of the VibrMAX feature; only VibrMAX>2000 values were taken into account in the calculations.

5.2. Trend analysis

The basis of the diagnostic procedure is the trend analysis of changes of the features values (let us call them diagnostic variables) identified as the key ones in the previous subsection. The trend analysis is conducted on the basis of historical data, simultaneously for the *h1, h2, h3* work cycles of the gantry. Here *h1* describes a short period of time while *h2=2h1, h3=3h1*. This allows for identifying dynamic (though maybe short-term) and stable change trends in the analyzed time series.

The procedure considers the following parameters as well:

- smoothing degree (*sm*) – due to a very big variance of diagnostic variable values, before the trend analysis the values are smoothed by a moving average which takes into account 100 previous values (100 work cycles of the gantry),
- boundary level of a diagnostic variable value (*cvl*) – a reference level which allows for estimating the number of work cycles of the gantry remaining until the level is exceeded; in the case of testing the increase of the diagnostic variable value, this level does not have to be identified with the gantry failure.

The trend extrapolation allows for estimating the number of work cycles remaining until the *cvl* level is exceeded. Instead of the number of cycles, it is possible to use the number of days after which the *cvl* level is exceeded. For this purpose, it is assumed that the average daily number of cycles (*awc*) is arbitrarily set or calculated (based on historical data) and it is an equivalent of one day of the gantry operation. This value can be modified as the information about the operation of a given gantry is being collected.

Let *S* denote a time series of consecutive values of the diagnostic variable in the range of work cycles. Let S_{sm} denote a time series of value *S* smoothed by means of the moving average from the last sm of value *S*. The analysis of a linear trend in the S_{sm} series, in which the estimation of parameters $hi\left(i\in\{1,2,3\}\right)$ of the last values of the S_{sm} series were used, leads to the following model:

$$\hat{S}_{sm,h}(i)=Ai+B \tag{2}$$

where: *i* is the number of the gantry work cycle, *A* and *B* are parameters of the model estimated on the basis of h of the last values of the S_{sm} series.

If $\hat{S}_{sm,h}(i)$ is the current value of the model (2), the number of work cycles remaining until the boundary value *cvl* is exceeded can be calculated according to the following formula:

$$\Delta i=\begin{cases}\dfrac{cvl-\hat{S}_{sm,h}(i)-B}{A} & A\neq 0\\ \infty & A=0\end{cases} \tag{3}$$

Assuming that the daily average number of gantry work cycles is *awc*, it is possible to determine the number of gantry workdays after which the value of the considered diagnostic variable will exceed the *cvl* level. The values of indicator (3) can be grouped so that their interpretation could be better understood by the end user. This way, we will obtain a meta-indicator marked as Δi_+ whose values are interpreted as follows:

- $\Delta i_+=0$ if and only if $\Delta i<0$ or $\Delta i_+=\infty$ – this refers to a situation in which we have a downward or lateral trend,
- $\Delta i_+=1$ if and only if $\Delta i_+\in[0,1]$ – this refers to a situation in whih the *cvl* level is exceeded or will be exceeded in less than *awc* work cycles of the gantry,
- $\Delta i_+=\left(\frac{1}{awc}\right)\Delta_i$ in all other cases.

Due to the way it is defined, the Δ_{i+} indicator does not reflect the drop of the diagnostic variable value. To put it more precisely, in the case of a constant or downward trend $\Delta i_+=0$, particularly when the values of the considered variable get stabilized on a certain level (including a the level of >*cvl*).

The Δi_- meta-indicator allows for indicating time ranges in which the trend of the monitored value is downward. The threshold value for this indicator was set as 0. An=other solution is to set this value as an average vibration level during the period for which we are sure that the gantry was operational. The interpretation of Δi_- – after modifying the formula (3) by replacing the *cvl* value with the *nvl* value – is as follows:

- $\Delta i_-=0$ if and only if $\Delta i<0$ or $\Delta i_-=\infty$ –this refers to a situation in which we have an upward or a lateral trend,
- $\Delta i_-=1$ if and only if $\Delta i_-\in[0,1]$ – this refers to a situation in which the 0 level is exceeded or will be exceeded in less than *awc* work cycles of the gantry,
- $\Delta i_-=\left(\frac{1}{awc}\right)\Delta i$ in all other cases.

The values of $\Delta i+$ and Δi_- can be visualized on a diagram. Coming back to the analyzed data set: in the first half of February we can observe two moments of a significantly worse diagnostic condition of the device (the analysis was conducted for *h1*=7 days, *awc*=40, *cvl*=4000). From the Δi_+ diagram we can see that the first moment is in its maximum at a time of less than 7 days remaining to the time when the admissible level of the diagnostic variable value is exceeded, while for the second moment it is 10 days.This can be clearly seen on diagrams of trends built on the basis of the analysis of the last 3 (black) and 7 (red) days of the gantry operations. The diagram of Δi_- diagram, in turn, does not demonstrate any significant improvement

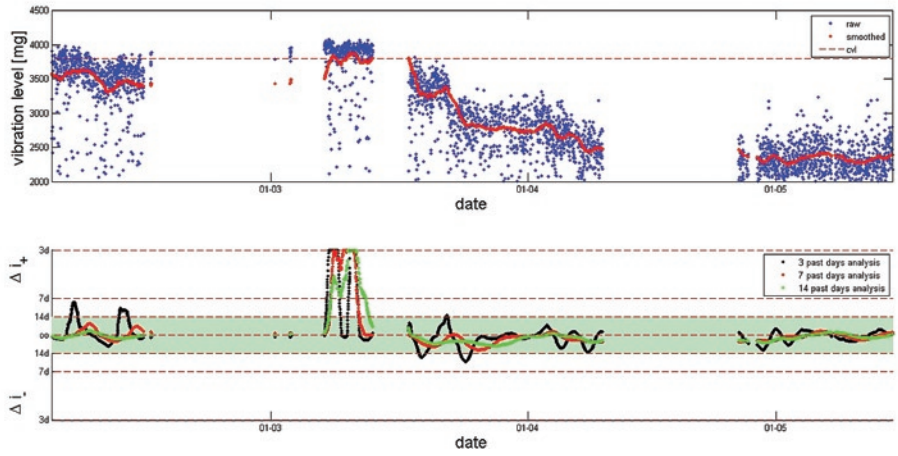


Fig. 7. Diagram of monitoring the value of VibrMax2000_Q3 – gantry 1. Upper diagram: original values (blue), smoothed values (red) and the cvl level. Lower diagram: values of the Δi indicator generated on the basis of a trend analysis conducted based on 3, 7 and 14 past days of the gantry operations

in the diagnostic state between these two periods, though the analysis of the smoothed course of the diagnostic variable (Figure 7, upper diagram) shows a small temporary drop of the diagnostic variable value.

For legibility reasons, the Y axis of in the lower diagram of Fig. 7 was cut to a value corresponding to 3 days after exceeding the *cvl* level. In addition, the value of *nvl* was set as 0, which lowers the warning and alarm thresholds and, at the same time, ramps up the values placed in the report.

In the considered period, the report log did not contain any information about repairs; however, it was noted that the operator recorded an increased vibration level. A break in the measurement data resulted from a break in the gantry operations. Then, around the 10th of March, the gantry was operating again but, as it could be seen from the beginning of its operations in fact, the Δi_+ indicator reached the boundary value very quickly. This was confirmed by an increasing vibration level which followed the values of Δi_+. Next, the gantry was under renovation for a few days, after which it was put back in operation. The renovation period was reflected by a decreasing value of Δi_+ and a rising value of Δi_-, followed by stabilized trends of 7- and 14-day values of the Δi_+ indicator.

The interpretation of changes in the Δi_+ and Δi_- values, as well as the value of the diagnostic variable based on which these indicators are calculated, can be placed in the report. Table 1 features a sample report.

The values in the report are determined according to the following rules:

- OK! if the Δi_+ indicator is equal to 0 and the values of the diagnostic variable are below *cvl* 0.7; additionally, if the trend is upward, then the number of "days" is given after which the *cvl* level will be exceeded, provided that this number is below 35;
- W if Δi_+ is equal to 0 and the values of the diagnostic variable are within the range (*cvl* 0.7, *cvl* 0.8]; additionally, if the trend is upward, then the number of "days" is given after which the *cvl* level will be exceeded, provided that this number is below 28; if the number is above 28, only the information about the trend direction is given,
- A if the values of the diagnostic variable are above cvl 0.8; additionally, if the trend is upward, then the number of "days" is given after which the *cvl* level will be exceeded, provided that this number is below 21; if the number is above 21, only the information about the trend direction is given.

Please note that 1 "day" is the average number of gantry activations during the day. This value can be set arbitrarily or calculated based on the history of the gantry work cycles.

The far right column contains the number of the same decisions for trend analyses conducted on the basis of the last 3, 7 and 14 "days". This column allows for tracking the stability of changes in the diagnostic variable values.

Depending on the user's needs, the report may refer to one or more sensors as well as to a bigger number of diagnostic variables.

The report does not give any suggestions about the time in which a failure will occur or the gantry will have to be renovated. Still, it allows one to observe, in an organized manner, the trends in the changes of a variable (or variables) identified as diagnostic variables.

Table 3. Diagnostic report with trends of the values of the diagnostic variable describing vibration levels in particular work cycles of the gantry (gantry 1)

Gantry 1 trends in changes of $\Delta i+$ values				
Diagnostic variable Vibr_Max2000_Q3	Trend analysis based on the last:			
Date	3 days	7 days	14 days	Compatible decisions
2017-02-04	A (8)	A(↑)	A (↑)	3
2017-02-05	A (10)	A(↑)	A(↑)	3
...	...	...	...	
2017-02-08	A(↓)	A(↑)	A(↑)	2
...	...	...	...	...
2017-03-24	W(↓)	W(↓)	W(↓)	3
...	...	...	...	
2017-05-09	OK!	OK!	OK!	3

5.3. Application phase

A practical application of the presented method requires the following:

- distributing sensors on the monitored gantry;
- identifying the gantry work cycles in accordance with the method presented at the beginning of section 5.2;
- calculating the value of VibrMax2000_Q3 for each sensor and, possibly, successive features from the rank included in section 5.2;
- determining the values of *cvl*, *awc*;
- calculating the values of the Δi_+, Δi_- indicators regularly, once a day; during the calculation of these values, the values of the sm and h1 parameters are set at 100 and 7 *awc*, respectively;
- generating the diagrams of Δi_+, Δi_- values and a report presented in Table 3.

From the practical point of view, it is also necessary to specify the principles of operation during renovation works (whether the history of the diagnostic variable value must be reset) and when the gantry works in the so-called start-up mode. There are frequent situations when the gantry is operated in the start-up mode despite the fact that it should not be operated at all – in such instances the vibration level is considerably lower.

6. Verification tests

Verification tests were conducted on a second identical gantry used in the same foundry. The renovations log claimed that the gantry was renovated in the period of 16-18.08.2017. Still, it was not done properly and another renovation was necessary in a short time (in October). The latter renovation included the replacement of a pinion. From the beginning of September to the time of the second renovation, the gantry was working much less frequently than when it was fully operable.

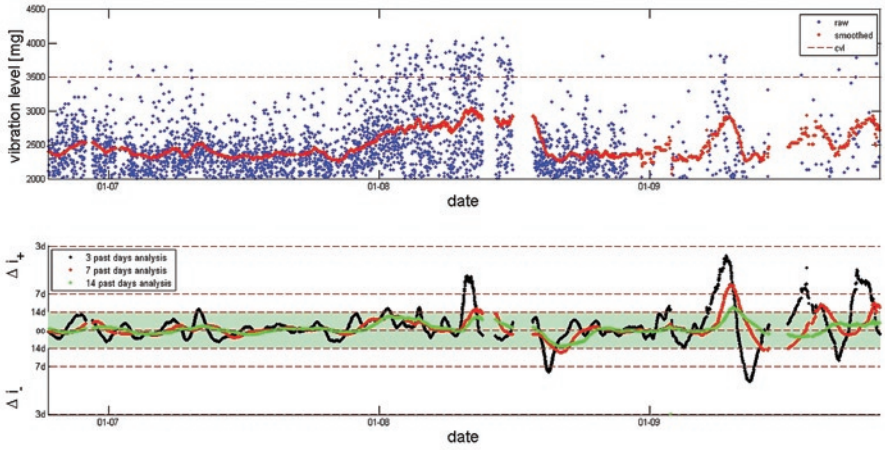


Fig. 8. Diagram of monitoring value of VibrMax2000_Q3 – gantry 2. Upper diagram: original values (blue), smoothed values (red) and the cvl level. Lower diagram: values of the Δi indicator generated on the basis of a trend analysis conducted based on 3, 7 and 14 past days of the gantry operations

Table 4. Diagnostic report with trends of the values of the diagnostic variable describing vibration levels in particular work cycles of the gantry (gantry 2)

Gantry 1 trends in changes of $\Delta i+$ values				
Diagnostic variable Vibr_Max2000_Q3	Trend analysis based on the last:			
Date	3 days	Date	3 days	Compatible decisions
2017-07-25	OK!	OK!	OK!	3
...	...	...	...	...
2017-08-07	W (↓)	W(↑)	W(↑)	2
...	...	...	...	
2017-08-11	A(5)	A(20)	A(↑)	3

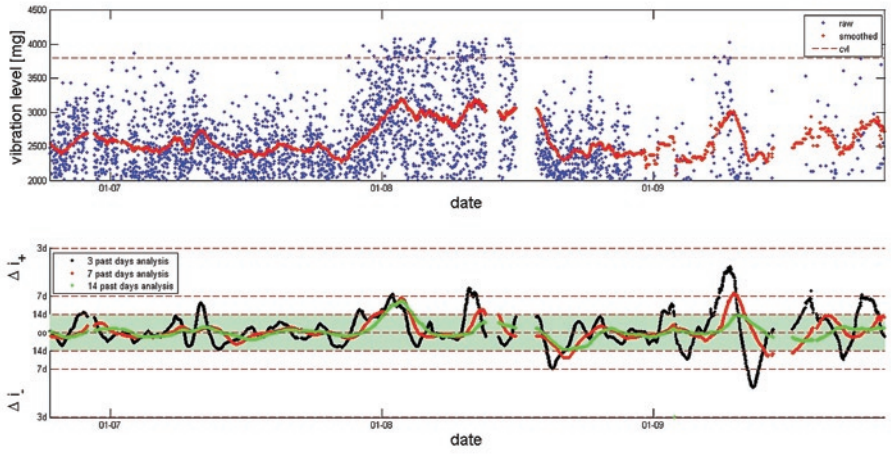


Fig. 9. Diagram of monitoring value of VibrMax2000_P95 – gantry 2. Upper diagram: original values (blue), smoothed values (red) and the cvl level. Lower diagram: values of the Δi– indicator generated on the basis of a trend analysis conducted based on 3, 7 and 14 past days of the gantry operations

Figure 8 presents original and smoothed values of the VibrMAx_Q3 diagnostic variable, along with the values of the $\Delta i+$ and $\Delta i-$ indicators. Table 4 features a report for the first few days of August. It is important to note that before the renovation the gantry worked in the start-up mode only, which slightly weakened the dynamics of the vibration level increase. This can be seen on the diagram of the VibrMax2000_Q3 diagnostic variable value.

Figure 9 also features a diagram of the VibrMax2000_P95 diagnostic variable. This is the last of the relevant variables identified in subsection 5.2.

This variable shows the process of the gantry deterioration very well as clear upward trends are observed at the beginning of August when the gantry was still working in normal conditions (in the working mode, not start-up mode). Then, in spite of lowering the course of operations, the moment for performing necessary renovation is predicted properly. After the improper renovation, irregularities in the machine work are signalled as well.

This example shows how important it is to generate diagrams and diagnostic reports for all five diagnostic variables identified in section 5.2, for all monitored sensors. From the practical point of view, the analysis and interpretation of such a big number of diagrams and reports is impossible for the operator to carry out. Therefore, a data analyst should prepare recommendations as to which diagnostic variable and which vibration levels *cvl* and *nvl* should be applied as the basis for generating the report. Such work is increasingly more often conducted as a maintenance service for a decision-support system by data science departments of companies which supply these types of systems.

7. Summary and conclusions

Monitoring and diagnostics of technical devices is not an easy task as each time it requires an individual approach to the analyzed object.

The paper presents a system for monitoring vibrations which makes use of a wireless network of sensors. In addition, it features the use of measurements collected by sensors which monitor the diagnostic state of gantries.

In the second part of the paper the authors presented a method which allows for analyzing trends in the changes of the vibration level and for associating this level with the deteriorating condition of the gantry. Additionally, they listed the requirements regarding the practical use of the method. As it was not possible to simulate

different kinds of damages and failures, the diagnostic methodology is relatively simple – it is based on the assumption that the deteriorating condition of the gantry is demonstrated by an increasing vibration level (which is in accordance with the gantry operators' intuition). The key issues of the diagnostic procedures include: the process of identifying work cycles, identification of statistics which most reflect the impact of the vibration level on the machine state, as well as short- and medium-term trends of changes in the statistics values. The paper features solutions to these issues. Moreover, the authors proposed a structure of a diagnostic report generated for the maintenance operator.

The presented measurement system was implemented in a foundry in Upper Silesia, Poland.

Further works will focus on the system application for monitoring and diagnosing other devices – it is planned to monitor the vibrations of rollers on which a conveyor belt moves. In order to implement the system, it will be necessary to develop a new diagnostic procedure dedicated to this type of task. The remaining parts of the system (hardware and software – monitoring part) are ready to use on any object.

The developed system is, to a large extent, in compliance with the architecture of data driven frameworks for decision support. In such systems the diagnostic models are made to meet the users' specific needs on the basis of historical data analyses, while the modules for data acquisition, storage and visualization are ready to use.

Acknowledgements

The work was carried out partially within the statutory research project of the Institute of Informatics, BK-213/RAU2/2018. The work was carried out partially (participation of the second author) from the funds of the Ministry of Science and Higher Education (agreement number 10/DW/201701/1).

References

1. Antoni J. The spectral kurtosis: a useful tool for characterising non-stationary signals. Mechanical Systems and Signal Processing 2006; 20(2): 282-307, https://doi.org/10.1016/j.ymssp.2004.09.001.
2. Bartelmus W, Zimroz R. Vibration condition monitoring of planetary gearbox under varying external load. Mechanical Systems and Signal Processing 2008; 23: 246-257, https://doi.org/10.1016/j.ymssp.2008.03.016.
3. Bartelmus W, Zimroz R. A new feature for monitoring the condition of gearboxes in non-stationary operating conditions. Mechanical Systems and Signal Processing 2009; 23(5):1528-1534, https://doi.org/10.1016/j.ymssp.2009.01.014.
4. Breiman L. Random forests. Machine Learning 2001; 45(1): 5-32, https://doi.org/10.1023/A:1010933404324.
5. Chen B, Yin P, Gao Y, Peng F. Use of the correlated EEMD and time-spectral kurtosis for bearing defect detection under large speed variation. Mechanism and Machine Theory 2018; 129: 162-174, https://doi.org/10.1016/j.mechmachtheory.2018.07.017.
6. Du W, Li A, Ye P, Liu C. Fault diagnosis of plunger pump in truck crane based on relevance vector machine with particle swarm optimization algorithm. Shock and Vibration 2013; 20(4): 781-792, https://doi.org/10.1155/2013/610235.
7. Elforjani M, Bechhoefer E. Analysis of extremely modulated faulty wind turbine data using spectral kurtosis and signal intensity estimator. Renewable Energy 2018; 127: 258-268.
8. Głowacz A, Głowacz W. Vibration-Based Fault Diagnosis of Commutator Motor. Shock and Vibration 2018; art. id 7460419, https://doi.org/10.1155/2018/7460419.
9. Głowacz A, Głowacz Z. Diagnosis of the three-phase induction motor using thermal imaging. Infrared Physics & Technology 2017; 81: 7-16, https://doi.org/10.1016/j.infrared.2016.12.003.
10. Henao H, Capolino G, Manes F. Trends in fault diagnosis for electrical machines: A review of diagnostic techniques. IEEE industrial electronics magazine 2014; 8(2): 31-42, https://doi.org/10.1109/MIE.2013.2287651.
11. ISA95 – Enterpise-Control System Integration Standard (https://www.isa.org/isa95/)
12. Jingwei G, Niaoqin H, Lehua J, Jianyi F. A New Condition Monitoring and Fault Diagnosis Method of Engine Based on Spectrometric Oil Analysis. Advances in Intelligent and Soft Computing 2011, 110:117-124, https://doi.org/10.1007/978-3-642-25185-6_16.
13. Korbicz J, Kościelny M (eds.).Modeling, Diagnostics and Process Control. Implementation in the DiaSter System. Springer-Verlag Berlin, Heidelberg 2011, https://doi.org/10.1007/978-3-642-16653-2.
14. Korbicz J, Kościelny M, Kowalczuk Z, Cholewa W (eds.). Fault Diagnosis. Models, Artificial Intelligence, Appications. Springer-Verlag Berlin Heidelberg 2004, https://doi.org/10.1007/978-3-642-18615-8.
15. Li Y, Liang X, Xu M, Huang W. Early fault feature extraction of rolling bearing based on ICD and tunable Q-factor wavelet transform. Mechanical Systems and Signal Processing 2017, 86(Part A): 204-223.
16. Macián V, Tormos B, Olmeda P, Montoro L. Analytical approach to wear rate determination for internal combustion engine condition monitoring based on oil analysis. Tribology International 2003; 36: 771–776, https://doi.org/10.1016/S0301-679X(03)00060-4.
17. Mazurkiewicz, D. Computer-aided maintenance and reliability management systems for conveyor belts. Eksploatacja i Niezawodność-Maintenance and Reliability 2014; 16(3):377-382.
18. Mobley R. An Introduction to Predictive Maintenance. Second Edition. Butterworth-Heinemann 2013.
19. Peng Z, Chu F. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. Mechanical Systems and Signal Processing 2004; 18(2): 199-221, https://doi.org/10.1016/S0888-3270(03)00075-X.
20. Przystałka P, Moczulski W. Methodology of neural modelling in fault detection with the use of chaos engineering. Engineering Applications of Artificial Intelligence 2015; 41: 25-40, https://doi.org/10.1016/j.engappai.2015.01.016.
21. R Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014, http://www.R-project.org
22. Samuel P, Pines D. A review of vibration-based techniques for helicopter transmission diagnostics. Journal of Sound and Vibration 2005; 282(1-2): 475-508, https://doi.org/10.1016/j.jsv.2004.02.058.
23. Silva S, Costa P, Gouvea M, Lacerda A, Alves F, Leite D. High impedance fault detection in power distribution systems using wavelet transform and evolving neural network. Electric Power Systems Research 2018, 154: 474-483, https://doi.org/10.1016/j.epsr.2017.08.039.
24. Therneau T, Atkinson B. Package: rpart (http://cran.r-project.org/web/packages/rpart/rpart.pdf)
25. Wachla D,Moczulski W. Identification of dynamic diagnostic models with the use of methodology of knowledge discovery in databases. Engineering Applications of Artificial Intelligence 2007; 20(5): 699-707, https://doi.org/10.1016/j.engappai.2006.11.002.

26. Wu S, Zuo M. Linear and Nonlinear Preventive Maintenance Models. IEEE Transactions on Reliability 2010; 59(1): 242-249, https://doi.org/10.1109/TR.2010.2041972.
27. Yan R, Gaob R, Chen X. Wavelets for fault diagnosis of rotary machines: A review with applications. Signal Processing 2014; 96(A): 1-15.
28. Ye Z, Wu B, Zargari N.: Online mechanical fault diagnostics of induction motor by wavelet artificial neural network using stator current. IECON Proceedings 2000; 2: 1183–1188.
29. Zio E. Some challenges and opportunities in reliability engineering. IEEE Transactions on Reliability 2016; 65(4): 1769-1782, https://doi.org/10.1109/TR.2016.2591504.

Marek SIKORA
Krzysztof SZCZYRBA
Łukasz WRÓBEL
Marcin MICHALAK
Institute of Informatics, Silesian University of Technology
ul. Akademicka 16, 44-100 Gliwice, Poland

E-mails: marek.sikora@polsl.pl, k.szczyrba@somar.com.pl, lukasz.wrobel@polsl.pl, marcin.michalak@polsl.pl

Article citation info:
KARWAT B, MACHNIK R, NIEDŹWIEDZKI J, NOGAJ M, RUBACHA P, STAŃCZYK E. Calibration of bulk material model in Discrete Element Method on example of perlite D18-DN. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 351–357, http://dx.doi.org/10.17531/ein.2019.2.20.

Bolesław KARWAT
Ryszard MACHNIK
Jerzy NIEDŹWIEDZKI
Magdalena NOGAJ
Piotr RUBACHA
Emil STAŃCZYK

CALIBRATION OF BULK MATERIAL MODEL IN DISCRETE ELEMENT METHOD ON EXAMPLE OF PERLITE D18-DN

KALIBRACJA MODELU MATERIAŁU SYPKIEGO W METODZIE ELEMENTÓW DYSKRETNYCH NA PRZYKŁADZIE PERLITU D18-DN

Analytical methods for calculations of the transport machinery are often insufficient especially when untypical granular materials are considered. Discrete Element Method (DEM) is a very useful numerical tool supporting designing and optimization of the transport equipment. However, to obtain reliable DEM simulation results an accurate set of input parameters values is needed. The most common calibration approach is to make use of a procedure where laboratory tests are performed and then the same experiments are numerically replicated in DEM. The article presents calibration of the DEM input parameters on the example of perlite D18-DN Based on the performed calibration, the model of perlite transport in a screw conveyor has been shown.

***Keywords**: discrete element method, modelling, computer simulations, calibration, bulk materials, perlite.*

Analityczne metody obliczeniowe parametrów konstrukcyjnych maszyn i urządzeń transportowych są często niewystarczające, zawłaszcza w przypadku transportu nietypowych materiałów sypkich. Pomocnym narzędziem numerycznym wspierającym proces projektowania i optymalizacji urządzeń do transportu materiałów sypkich jest Metoda Elementów Dyskretnych (DEM). Uzyskanie wiarygodnych wyników symulacji wymaga kalibracji parametrów wejściowych modelowanego materiału wykorzystując wyniki badań laboratoryjnych właściwości fizykochemicznych rzeczywistych materiałów. W artykule przedstawiono metodologię kalibracji modelu DEM na przykładzie perlitu D18-DN. W oparciu o przeprowadzoną kalibrację zaprezentowano możliwości zastosowania metody DEM do symulowania transportu materiału przenośnikiem ślimakowym.

***Słowa kluczowe**: metoda elementów dyskretnych, modelowanie, symulacje komputerowe, kalibracja, materiały sypkie, perlit.*

1. Introduction

Optimization of the design of machines and devices for the transport of bulk materials is aimed at reducing the costs of their production while maintaining quality and functionality. This process involves the analysis of the functional features of the device and the search for possibilities of its cheaper generation or improvement of functionality through structural improvements, saving or substituting of construction materials. There are some of the main optimization criteria adopted by design offices and companies producing industrial transport equipment like reducing the weight of the device, energy consumption and increasing productivity. The analytical approach to the design of machines and transport devices allows satisfactory accuracy to choose the characteristic dimensions of components, to predict efficiency and power demand, and to estimate the strength and durability of the device. However, this approach does not provide any information about behavior of the bulk material during transport and interactions with equipment. This knowledge is necessary to be able to start the process of design optimization. Computer-Aided-Engineering tools are helpful in this extent. Increasing computing power of computers allows the use of numerical methods to solve complex engineering problems. In the field of bulk material simulation, the Discrete Element Method DEM is widely used [2-3, 6, 9, 12]. It allows to simulate the behavior of granular materials with any physical properties subjected to various extortions results from the operating conditions of the considered device. An example of a DEM application in the conveyor belt simulation is shown in Figure 1.

Considering the simulation model the physicochemical and mechanical properties of the transported granular material requires calibrating the input parameters of the model (among others determining: coefficient of friction, bulk density, shape and size of particles) based on the results of laboratory tests [4-5,10-11,13]. This paper presents the methodology for testing bulk materials in terms of their use for the calibration of the model created in the DEM environment and the calibration process on the example of granular material – D18-DN perlite.

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl***

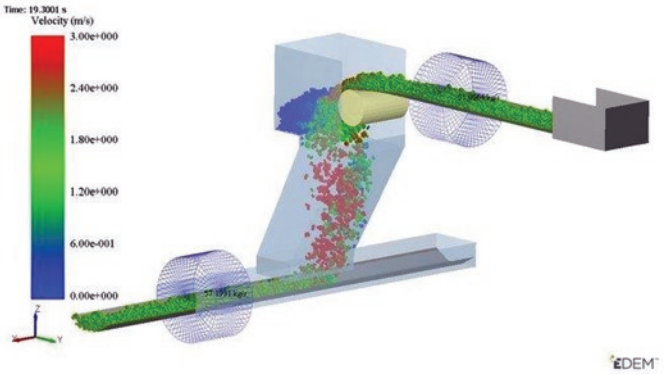


Fig. 1. An example of the application of the DEM method in belt conveyor simulations

2. Physicochemical properties of the material: D18-DN perlite

The simulation model of the material transport process requires properly defined parameters describing the physicochemical properties of the transported material. These parameters were determined on the basis of the laboratory tests results of D18-DN perlite. The scope of research included designation:

a) relative density of the material – ρ_{mat} [g/cm^3],
b) bulk weight – ρ_b [g/cm^3],
c) angle of natural repose – α_u [°],
d) outer friction coefficient of the steel-perlite friction pair – φ_z [–].

2.1. Determination of relative material density

The relative density of the material was determined by the pycnometric method. Measurements were made with a 50 cm^3 Gay-Lussac pycnometer. In the pycnometric method, distilled water is most often used as the standard liquid. Due to the chemical properties of perlite (the possibility of leaching in the aquatic environment), anhydrous ethyl alcohol was used as the standard liquid. The density value of the tested material was determined from the relation (1):

$$\rho_{mat} = \rho^{t}_{alk} \cdot \frac{m_3 - m_1}{(m_2 - m_1) - (m_4 - m_3)} \tag{1}$$

where:

ρ_{mat} – material density [g/cm^3],
ρ^{t}_{alk} – density of ethyl alcohol at temperature t [g/cm^3],
m_1 – weight of an empty pycnometer [g],
m_2 – weight of a pycnometer with alcohol [g],
m_3 – weight of the pycnometer with the material [g],
m_4 – weight of a pycnometer with alcohol and material [g].

The calculations include the change in the relative density of ethyl alcohol as a function of temperature. The mean value obtained from three measurements was taken as the relative density value of the examined materials.

2.2. Determination of material bulk density

The bulk density is the ratio of the weight of the material to its total volume, taking into account pores, spaces between the grains and any contaminants. The bulk weight of the material in the loose state depends on the specific weight of the material, its moisture content and the degree of filling of the given volume. The measurement was carried out in a measuring container with a volume of 1 dm^3, determining the bulk density of the material in the loose state ρ_b from the relation (2):

$$\rho_b = \frac{m_{2b} - m_1}{V} \quad \text{[g/cm3]} \tag{2}$$

where:

ρ_b – bulk density in the loose condition [g/cm^3],
m_{2b} – weight of the measuring cylinder and samples in loose condition [g],
m_1 – weight of the measuring cylinder [g],
V – capacity of the measuring cylinder [cm^3].

The average value obtained from three measurements was taken as the bulk density value of the tested materials.

2.3. The angle of natural repose

The angle of repose α_u is the inclination angle between descent of the material cone formed by free pouring and horizontal plane. The tests were carried out in a measuring stand with a loose charge of the tested material. The value of the angle of repose was directly measured at the test stand, and the photos which were taken for further analysis. Using the image analysis, the angles of material placement with respect to the horizontal plane were determined (Fig. 2.).

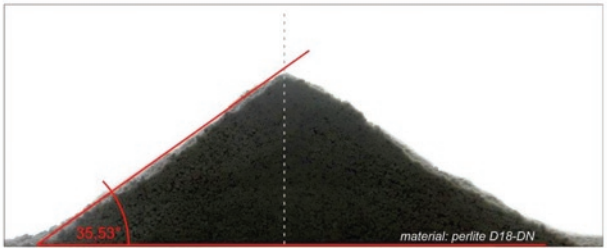


Fig. 2. Measurement of the angle of natural repose of perlite by using image analysis

2.4. Outer friction coefficient

Determination of outer friction was carried out in the AB-2A direct shearing apparatus used for soil cohesion studies. The tests were carried out in a 60 x 60 mm box at a normal stress of 500 kPa. Sample of material was subjected to normal stress and displaced into horizontal direction due to the horizontal force. On the basis of registered time characteristics or changes of the shear force value in the displacement function, the outer friction coefficient was determined from the relation (3):

$$\phi_z = \frac{F_S}{F_N} \tag{3}$$

where:

ϕ_Z – outer friction coefficient [–],
F_S – shear force [kN],
F_N – normal force [kN].

An example of registered changes of the shear force value in the displacement function is presented in figure 3. The calculations were carried out assuming the maximum tangential stress value or the tangential stress value corresponding to the box displacement by 6 mm (10% of the box dimension).

The results of the physicochemical properties of the model material – D18-DN perlite – are presented in Table 1.

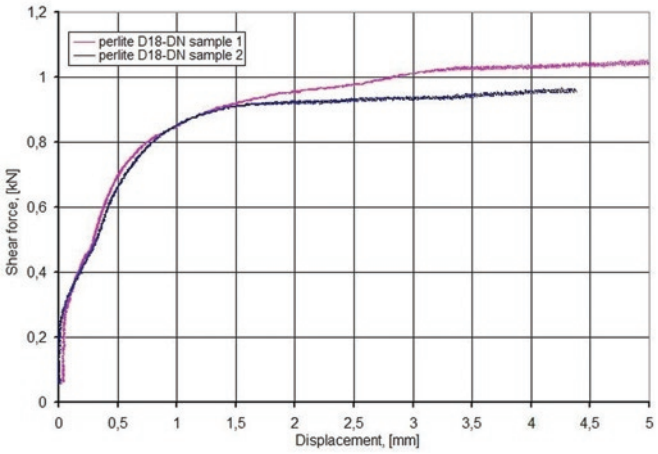


Fig. 3. Change of shear force as a function of displacement for a steel-perlite friction pair

Table 1. Physicochemical properties of D18-DN perlite

Physicochemical property	Value
Relative density ρ_{mat} [g/cm^3]	1.24
Bulk density ρ_b [g/cm^3]	0.10
The angle of natural repose [°]	35.50
Outer friction coefficient (friction pair steel-perlite D18-DN)	0.54

3. Description of the DEM method

Computer-Aided-Engineering programs have become important tools used by engineers in the process of designing new machines and devices in recent years. A number of engineering and scientific issues, which are too complicated to be described by analytical models or difficult to test in the laboratory, have been solved using numerical methods [1,7-8,14-15,17-18]. In addition, relatively cheap and easy access to high computing power and rapid development of computer programs implementing specific numerical methods causes that physical phenomena are modeled with incredible accuracy in relation to reality. The fundamental example is the Finite Element Method, which is commonly used to solve the problems of continuum mechanics (continuum). The base of the FEM method is the assumption that the modeled material is considered as a continuous in the macroscopic sense, and therefore it is characterized by a continuous distribution of matter (mass) in space, while its properties resulting from the atomic (or microscopic) structure are neglected. The idea of Finite Element Method is to divide the considered continuous area into a finite number of sub-areas (elements) connected to each other at nodes. In addition, it is assumed that the continuity of the center is preserved under load, hence the concept of FEM excludes its use in simulating the behavior of bulk materials. Therefore, for the needs of mechanics of discontinuity, the Discrete Element Method was developed. In DEM the considered material is modeled as a set of individual particles interacting with one another through appropriate contact models. Hence, the macroscopic features of the material are determined by the accepted properties of the particles and their interactions [19].

The algorithm of the DEM method is based on the contact mechanic, where forces acting on particular particles are calculated by appropriate models. Based on Newton's motion equations, particle accelerations are calculated. The velocities and positions of particles are determined by integration over time of these equations. By definition, the particles are rigid solids. However, the DEM method takes into account the deformation of particles in an artificial way, namely during the collision the depth of mutual penetration is calculated, as shown in Figure 4.

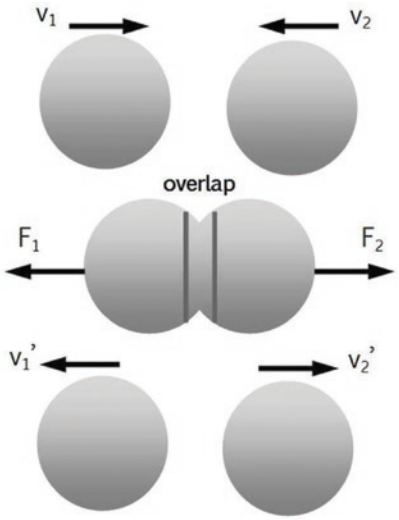


Fig. 4. Collision of rigid particles

The purpose of the contact model is to link the depth of penetration (in the normal and tangential direction in regard to the colliding particles) with the magnitude of contact force. Figure 5 shows the contact model of two particles during the collision. Usually, DEM simulation software offers a range of different contact models that allow modeling of elastic and plastic collisions with or without cohesion.

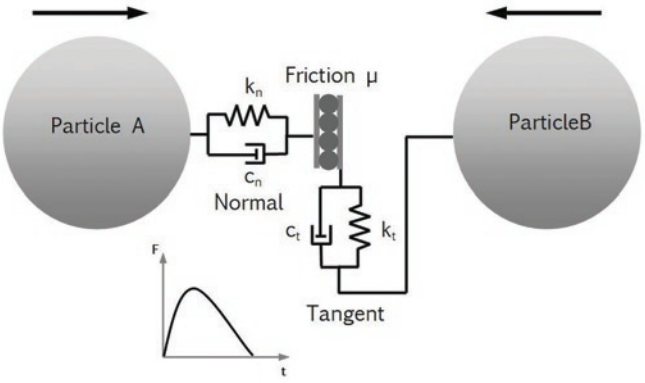


Fig. 5. Contact model of two particles during a collision

The parameters of individual particles (size, shape, density) and contact parameters (coefficients of friction, rolling resistance, coefficient of restitution) have a large influence on the behavior of the modeled bulk material [4]. The basic shape of the DEM particle is the sphere, however, due to the fact that the actual shapes of individual grains rarely take the ideal, spherical form, it is possible to create more complex shapes as a result of permanent connection of spherical particles. However, it has to be noted that the use of very complex shapes of clumps results in the extension of the calculation time.

Therefore, it is important to find a compromise between the actual shape and size of the solids, and those used in DEM simulations. To provide reliable behavior of the material in DEM simulations, the

input parameters of the model should be calibrated (i.e. friction coefficients, density, etc.).

4. Calibration of DEM parameters

In DEM analyzes, both physical and contact parameters are defined at the microscopic level of the particle [16]. Calibration consists in selecting of input parameters on a micro scale to obtain a convergent result of the modeled soil with the behavior of the real soil (macro scale). Therefore, it is more important to accurately reproduce the behavior of the entire soil than individual particle interactions. Parameters are calibrated by performing a suitable laboratory tests, e.g. measuring of the angle of repose or shear test. Then the same experiment is modeled in the DEM environment. Assuming a specific shape and size of particles, an iterative selection of the input parameters of the model is made until the convergent result with the experiment has been obtained [19]. As part of this work, the process of calibration of DEM input parameters has been presented on the example of laboratory measurements of physicochemical properties of perlite D18-DN.

Calibration of the DEM model begins with the selection of the shape and size of the particles. There is no unambiguous rule how to simplify the modeled shape of a solid and how to accept its size. A number of simulation studies performed by many researchers indicate that this selection is dictated by the specific purpose of the simulation. Both complex clumps and spherical particles are able to reflect the real angle of repose. In the case of spherical particles, an additional coefficient limiting the rolling of spheres should be introduced in the DEM model. An important factor in DEM simulation is the calculation time, which is influenced, inter alia, by the number of particles used and their size (in accordance with Rayleigh's theory). In the case of fine materials or powders, the dimensions of DEM particles are scaled so as to provide a reasonable compromise between the accuracy of the calculation and the runtime of simulation. Typically, the particle size of the modeled material is selected based on the proportions of the characteristic dimension of the device (e.g. screw pitch) to the particle radius.

Taking into account the regular shape of the perlite, spherical particles with 2, 5 and 8 mm radii were assumed. Sizes have been scaled relative to the actual grain dimensions to reduce their number in the soil, thereby limiting the calculation time. Due to the fact that perlite does not show cohesive properties, the simulation uses a non-linear contact model according to the Hertz-Mindlin theory. Then, bulk density calibration was performed for each particle size. It consisted of free filling up a measuring cylinder of a specified volume and evaluating the mass of the material in the filled cylinder space (Fig. 6). The input parameter of the model was the actual density of a single particle. This density was selected in an iterative manner to receive a value of bulk density convergent with obtained one as a result of laboratory measurements. The results of the above analysis also provided information on how the size of the selected particle influences the value of the bulk density. Table 2 presents the results of the bulk density depending on the particle's radius and the actual density.

The study of the angle of natural repose was aimed at calibrating the contact parameters of the model, i.e. the coefficient of internal friction and rolling resistance. In the DEM simulation, the laboratory test stand for examining the angle of repose was replicated [10]. The calibration consisted of an iterative selection of the values of both coefficients (inner and rolling friction), which made it possible to obtain the natural angle of the perlite close to the real one. For this purpose, a research plan was adopted consisting in changing the value of only one input parameter, i.e. the coefficient of internal friction at

Table 2. Calibration results of perlite bulk density

Particle radius [mm]	True particle density [kg/m^3]	Bulk density [kg/m^3]
2	200	119.4
5	200	119.9
8	200	119.7

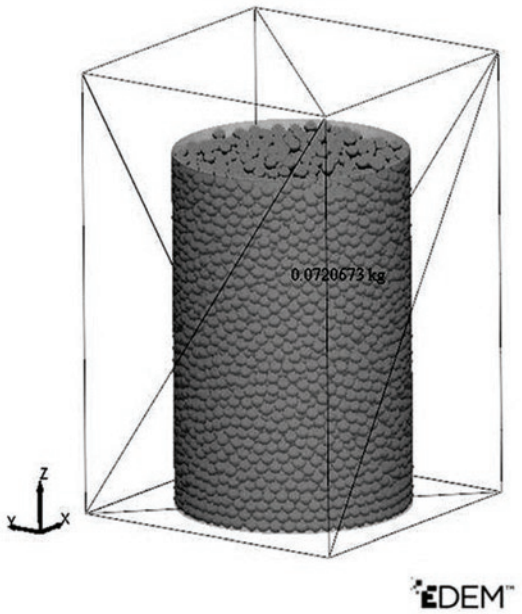


Fig. 6. Calibration of bulk density

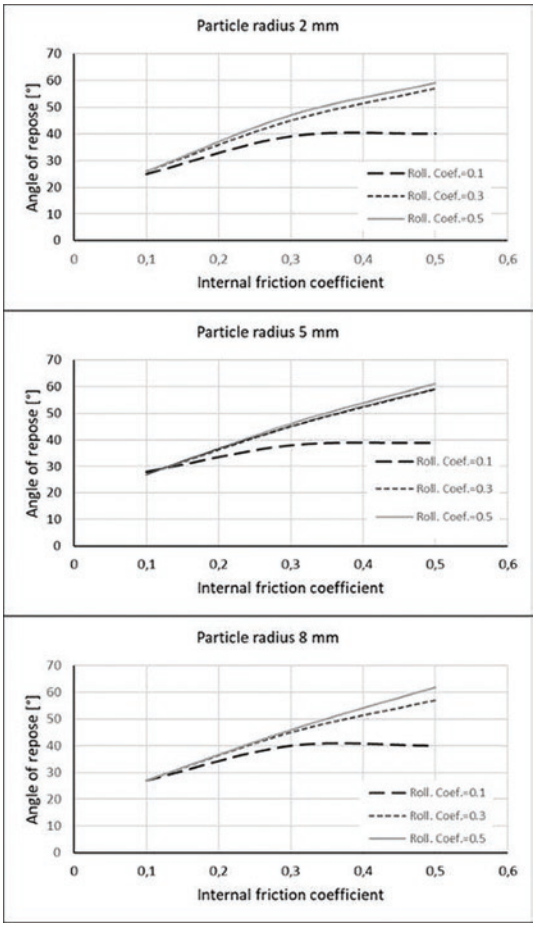


Fig. 7. The influence of the input parameters to the value of the angle of natural repose

a constant value of the rolling resistance coefficient. The series of simulation tests was repeated for three values of the rolling resistance coefficient. In results, characteristics combining the coefficients of internal friction and rolling resistance with the value of the angle of repose were obtained. Figure 7 presents graphs of changes in the angle of repose depending on the assumed input parameters.

Figure 8 shows the calibrated repose cone. This approach, assuming three different particle radii, three levels of internal friction coefficient and three values of rolling resistance coefficient, requires 27 simulations. These studies are undoubtedly time-consuming, but they provide important information about which input parameters determine the value of the angle of repose. Calibration of the outer friction coefficient (material-steel) was made on the model of a box apparatus for direct shear testing of the soil.

The model of the DEM test stand is shown in Figure 9. A sample of calibrated material was placed inside the box bounded from the bottom by a stationary plate and from the top by a stiff punch generating pressure on the material. Through the iterative change of the friction coefficient of the material-plate contact pair and the registration of the change of the normal and tangent force, it was possible to estimate (due to the relation 3) the outer friction coefficient.

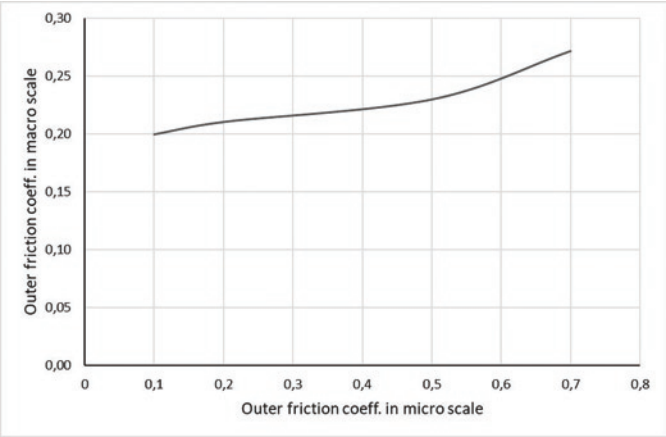


Fig. 8. Calibrated angle of natural repose

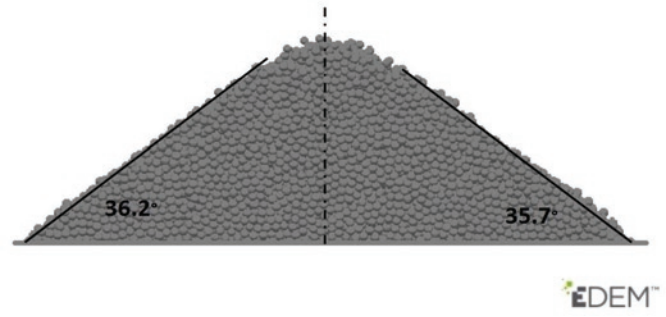


Fig. 9. Calibration of the outer friction coefficient

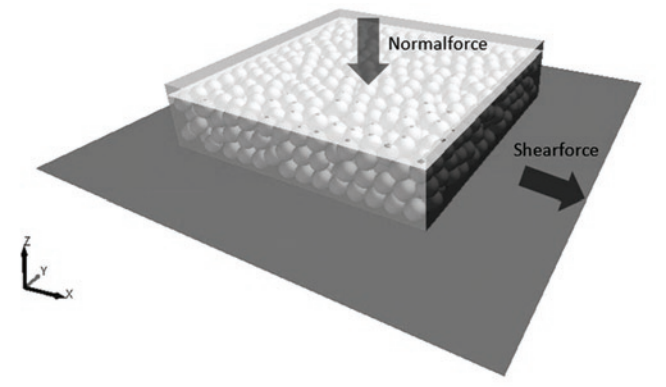


Fig. 10. Dependence of the outer friction coefficient on the deposit on the friction coefficient of a single particle

The influence of the coefficient of friction in the micro scale on the value of the coefficient of friction in the macro scale is presented in Figure 10.

5. Simulation of transport of D18-DN perlite in a screw conveyor

By having properly calibrated input parameters of the DEM model, it can be started performing the right simulations. Simulations can be used, among others, to determine the expected transport performance of the designed device or to determine the required drive power. The following is an example of a simulation of the transport of D18-DN perlite in screw conveyors. Two screws with the same external and internal diameters, differing only in the dimension of the pitch, were analyzed. Variant 1 that was assumed is a screw with a pitch of 250 mm and variant 2 - a screw with a pitch of 150 mm. The basic geometrical dimensions are shown in Table 3.

The results of the simulation provided information how the screw pitch influences the efficiency and power demand at different rotational speeds of the shaft. For the analysis, 5 mm particles were selected. The simulations for both variants of the screw were made by changing its rotational speed in the range of 10-70 rpm. The model of the screw conveyor with the material is shown in Figure 11.

Table 3. Parameters of the analyzed screw conveyors

Screw variant	External diameter [mm]	Internal diameter [mm]	Screw pitch [mm]
Variant I	250	139	250
Variant II	250	139	150

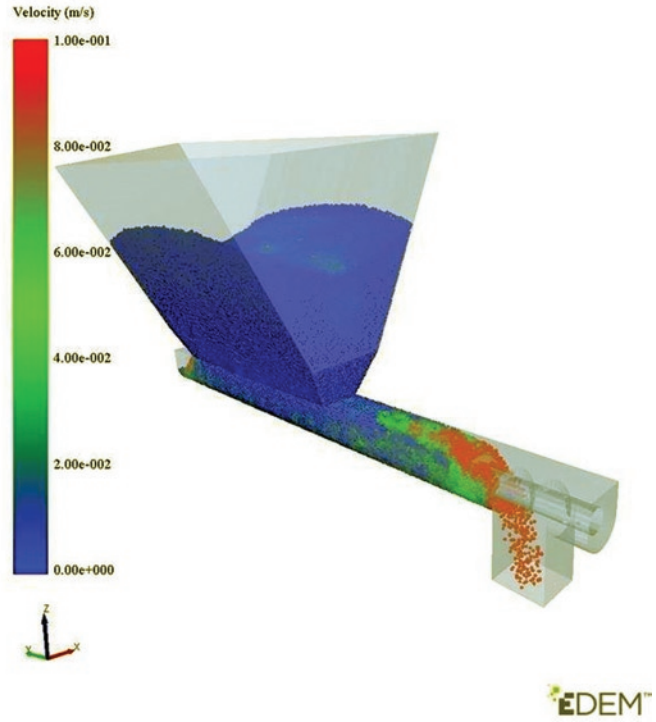


Fig. 11. Model of the screw conveyor

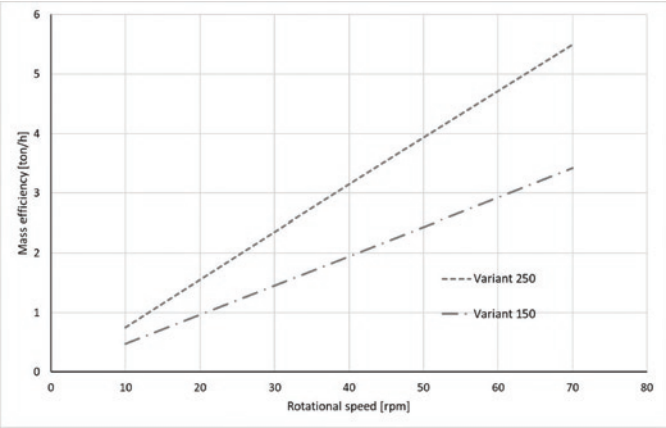


Fig. 12. Mass efficiency of the conveyor depending on rotational speed and screw pitch

The simulation results for both conveyor variants showing the mass efficiency depending on the rotational speed of the screw pitch are shown in Figure 12.

The change in the helix of the conveyor directly affects its transport efficiency. With increasing pitch at the same rotational speed of the shaft, mass efficiency increases. Having an adequately correlated material model, one can easily examine the functionality of a given transport device or select its appropriate geometric dimensions. In addition to determining geometrical characteristics, it also can be determined the required drive power. Figure 13 presents the demand for the power of the screw conveyor depending on the rotational speed and pitch of the screw in both accepted variants.

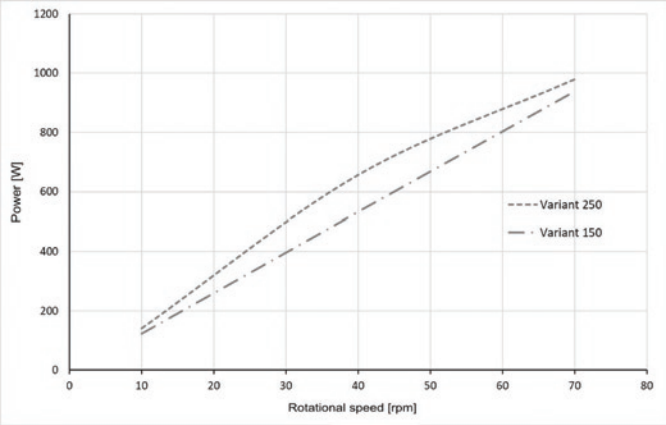


Fig. 13. Demand for the power of the conveyor depending on the rotational speed and screw pitch

From the graph above, it can be seen that the need of the conveyor for power increases with the increase of the rotational speed. For the same rotational speed of the shaft, reducing the screw pitch results in a reduction of the power demand. The increase in power demand is not directly proportional to the increase of mass efficiency, which is particularly visible at shaft speeds above 50 rpm. For example, at a speed of 70 rpm and a change in pitch from 150 mm to 250 mm, mass efficiency increased by 60%, while the power demand increased by only 6.5%. With the characteristics of the device's operation, it is possible to optimize the dimensions and mass of the screw or determine the preferred speed range. This speeds up the design process and allows for virtual testing of various solutions.

6. Conclusion

The process of calibration of input parameters of the DEM model is a factor that strongly determines the reliability of the final results of the simulation of the flow of bulk materials in transport machines. Laboratory tests determining physical and chemical properties of materials are relatively easy and cheap, especially in the case of fine-grained materials. This is undoubtedly a huge advantage, because on their basis numerical model can be calibrated. As demonstrated by the DEM calibration process itself, due to the iterative selection of individual parameters, it is a time-consuming process. Simulation tests have proved that the radius of a spherical particle does not affect the bulk density of the soil. Therefore, one size of spherical particles can be using. Likewise, the size of the particles does not substantially affect the value of the angle of repose. The above tests show that the same value of the cone angle can be obtained by using different sets of parameters of the internal friction coefficient and the coefficient of rolling resistance. The values of rolling resistance coefficients such as $F = 0.3$ and $F = 0.5$ have a slight influence on the value of the angle of repose. Different results were obtained by assuming the rolling resistance coefficient $F = 0.1$. In this case, for all three particle diameters (2, 5 and 8 mm) assumed in the simulations with an increase in the internal friction coefficient μ_w value of the angle of natural repose α_u initially grows and then assumes a constant value $\alpha_u \approx 40°$ for $\mu_w \geq 0.3$. This phenomenon can be explained by too low moment of rolling resistance of particles causing their rolling along the slope of the cone. This means that even with a very high value of the internal friction coefficient, the angle of repose will be small. Hence, in the case of spherical particles, it is necessary to use a rolling resistance model with an appropriate value (calibrated). The outer friction coefficient of soil depends non-linearly on the value of the single particle coefficient.

The DEM material model calibrated in this way should be verified by simulating a specific application, e.g. transport in a screw conveyor. As comparative features, the device operating parameters should be determined, i.e. screw shaft torque, mass efficiency or power consumption, which can be confronted with the results obtained on the laboratory line. In addition, further simulation studies should focus on non-spherical particles due to the fact that real solids have more complex shapes.

Undoubtedly, an important advantage of DEM simulation is the ability to quickly compare several constructions in terms of accepted criteria, e.g. transport efficiency or power. Quantitative and qualitative assessment of these differences allows for a proper selection of geometric features of the screw conveyor transporting a specific type of bulk material. This improves the design process and allows to explore different configurations without having to make expensive prototypes of devices.

References

1. Balac M, Grbovic A, Petrovic A, Popovic V. FEM analysis of pressure vessel with an investigation of crack growth on cylindrical surface. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (3): 378–386, https://doi.org/10.17531/ein.2018.3.5.
2. Campello M B E. A computational model for the simulation of dry granular materials. International Journal of Non-Linear Mechanics 2018; (106): 89-107, https://doi.org/10.1016/j.ijnonlinmec.2018.08.010.
3. Chen W, Roberts A, Katterfeld A, Wheeler C. Modelling the stability of iron ore bulk cargoes during marine transport. Powder Technology 2018; (326): 255-264, https://doi.org/10.1016/j.powtec.2017.12.006.

4. Coetzee C J. Review: Calibration of the discrete element method. Powder Technology 2017; 310: 104-142, https://doi.org/10.1016/j.powtec.2017.01.015.
5. Coetzee C J. Calibration of the discrete element method and the effect of particle shape. Powder Technology 2016; 297: 50-70, https://doi.org/10.1016/j.powtec.2016.04.003.
6. Elskamp F, Kruggel-Emden H. DEM simulations of screening processes under the influence of moisture. Chemical Engineering Research and Design 2018; (136): 593-609, https://doi.org/10.1016/j.cherd.2018.06.022.
7. Falkowicz K, Ferdynus M, Dębski H. Numerical analysis of compressed plates with a cut-out operating in the geometrically nonlinear range. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2015; 17 (2): 222–227, http://dx.doi.org/10.17531/ein.2015.2.8.
8. Ferdynus M, Kotełko M, Kral J. Energy absorption capability numerical analysis of thin-walled prismatic tubes with corner dents under axial impact. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (1): 252–259, https://doi.org/10.17531/ein.2018.2.10.
9. Fernandes M A P, Chaves Pereira M J, Pereira C F J. Discrete particle simulation in horizontally rotating drum: Uncertainty quantification of granular material physical parameters. Powder Technology 2018; (339): 199-210, https://doi.org/10.1016/j.powtec.2018.07.042.
10. Grima A, Wypych P. Discrete element simulation of a conveyor impact-plate transfer: calibration, validation and scale-up. Australian Bulk Handling Review 2010.
11. Gröger T, Katterfeld T. On the numerical calibration of discrete element models for the simulation of bulk solids. Conveying and Handling of Particulate Solids. Italy 2006, https://doi.org/10.1016/S1570-7946(06)80100-8.
12. Gröger T, Katterfeld A. Application of the Discrete Element Method in Materials Handling: Basics and Calibration. Bulk Solid Handling 2007; 27.
13. Gröger T, Katterfeld A. Verified Discrete Element Simulations of Bulk Solids Handling Equipment. AIChE Spring Meeting and Global Congress on Process Safety 2006.
14. Labra C, Oñate E, Kargl H, Rojek J. Discrete Element Simulation of rock cutting. International Journal of Rock Mechanics and Mining Sciences 2011; 48: 996-1010, https://doi.org/10.1016/j.ijrmms.2011.06.003.
15. Mitura A, Gawryluk J, Teter A. Numerical and experimental studies on the rotating rotor with three active composite blades. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (3): 571-579, https://doi.org/10.17531/ein.2017.4.11.
16. Powder Calibration: Challenges, Advances & Guidelines. edem.com 2018.
17. Rubacha P, Kulinowski P, Furmanik K, Badania laboratoryjne i symulacyjne ruchu ziaren w rynnie przesypowej przenośnika taśmowego. Transport Przemysłowy i Maszyny Robocze 2014; (1): 25-29.
18. Śledziewski K. Experimental and numerical studies of continuous composite beams taking into consideration slab cracking. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (3): 578-589, https://doi.org/10.17531/ein.2016.4.13.
19. WHAT IS DEM Theoretical background behind the Discrete Element Method (DEM). edem.com 2018.

Bolesław KARWAT
Ryszard MACHNIK
Jerzy NIEDŹWIEDZKI
Magdalena NOGAJ
Piotr RUBACHA
AGH-University of Science And Technology
al. Mickiewicza 30, 30-059 Cracow, Poland

Emil STAŃCZYK
ESOS Projekt
Cieśle 109, 29-105 Krasocin, Poland

E-mails: karwat@agh.edu.pl, machnik@agh.edu.pl, jniedz@agh.edu.pl, nogaj@agh.edu.pl, rubacha.piotr@gmail.com, stanczykemil@gmail.com

The Journal is indexed and abstracted in the Journal Citation Reports (JCR Science Edition), Scopus, Science Citation Index Expanded (SciSearch®) and Index Copernicus International.

The Quarterly appears on the list of journals credited with a high impact factor by the Polish Ministry of Science and Higher Education and is indexed in the Polish Technical Journal Contents database – BAZTECH and the database of the Digital Library Federation.

Ministry of Science and Higher Education
Republic of Poland

Task *„Implementation of procedures ensuring the originality of scientific papers published in the quarterly „Eksploatacja i Niezawodność – Maintenance and Reliability"* financed under contract 532/P-DUN/2018 from the funds of the Minister of Science and Higher Education for science dissemination activities.

All the scientific articles have received two positive reviews from independent reviewers.
Our 2017 Impact Factor is 1.383

Editorial staff:	Dariusz Mazurkiewicz, PhD, DSc (Eng), Associate Professor (Editor-in-Chief, Secretary of the Editorial Board) Tomasz Klepka, PhD, DSc (Eng), Associate Professor (Deputy Editor-in-Chief) Teresa Błachnio-Krolopp, MSc (Eng) (Editorial secretary) Andrzej Koma (Typesetting and text makeup) Krzysztof Olszewski, PhD (Eng) (Webmaster)
Publisher:	Polish Maintenance Society, Warsaw
Scientific patronage:	Polish Academy of Sciences Branch in Lublin
Address for correspondence:	"Eksploatacja i Niezawodność" – Editorial Office ul. Nadbystrzycka 36, 20-618 Lublin, Poland e-mail: office@ein.org.pl http://www.ein.org.pl/
Circulation:	550 copies

INFORMATION FOR AUTHORS

Eksploatacja i Niezawodnosc – Maintenance and Reliability – the journal of the Polish Maintenance Society, under the scientific supervision of the Polish Academy of Sciences (Branch in Lublin), published four times a year.

The scope of the Quarterly

The quarterly *Eksploatacja i Niezawodnosc – Maintenance and Reliability* publishes articles containing original results of experimental research on the durability and reliability of technical objects. We also accept papers presenting theoretical analyses supported by physical interpretation of causes or ones that have been verified empirically. *Eksploatacja i Niezawodność – Maintenance and Reliability* also publishes articles on innovative modeling approaches and research methods regarding the durability and reliability of objects.

The following research areas are particularly relevant to the journal:

1. degradation processes of mechanical and biomechanical systems,
2. diagnosis and prognosis of operational malfunctions and failures.
3. analysis of failure risk/wear,
4. reliability-and-environmental-safety engineering in the design, manufacturing and maintenance of objects,
5. management and rationalization of object maintenance,
6. risk management in the processes of operation and maintenance,
7. the human factor and human reliability in operation and maintenance systems.

Terms and Conditions of Publication

The quarterly *Eksploatacja i Niezawodnosc – Maintenance and Reliability* publishes only original papers written in English or in Polish with an English translation. Translation into English is done by the Authors after they have received information from the Editorial Office about the outcome of the review process and have introduced the necessary modifications in accordance with the suggestions of the referees! Acceptance of papers for publication is based on two independent reviews commissioned by the Editor. The publication fee for one text is 320 EUR (all graphics in color included).

The quarterly *Eksploatacja i Niezawodnosc – Maintenance and Reliability* proceeds entirely online at submission.ein.org.pl

Technical requirements

- After receiving positive reviews and after acceptance of the paper for publication, the text must be submitted in a Microsoft Word document format.
- Drawings and photos should be additionally submitted in the form of high resolution separate graphical files in the TIFF, SVG, AI or JPG formats.
- A manuscript should include: names of authors, title, abstract, and key words that should complement the title and abstract (in Polish and in English), the text in Polish and in English with a clear division into sections (please, do not divide words in the text); tables, drawings, graphs, and photos included in the text should have descriptive two-language captions, if this can be avoided, no formulae and symbols should be inserted into text paragraphs by means of a formula editor; references (written in accordance with the required reference format); author data – first names and surnames along with scientific titles, affiliation, address, phone number, fax, and e-mail address.

The Editor reserves the right to abridge and adjust the manuscripts. All submissions should be accompanied by a submission form.

Detailed instructions to Authors, including evaluation criteria can be found on the journal's website: www.ein.org.pl

Editor contact info

Editorial Office of „Eksploatacja i Niezawodnosc – Maintenance and Reliability"
Nadbystrzycka 36, 20-618 Lublin, Poland
e-mail: office@ein.org.pl

INFORMATION FOR SUBSCRIBERS

Fees

Yearly subscription fee (four issues) is 100 złoty and includes delivery costs. Subscribers receive any additional special issues published during their year of subscription free of charge.

Orders

Subscription orders along with authorization to issue a VAT invoice without receiver's signature should be sent to the Editor's address.

Note

In accordance with the requirements of citation databases, proper citation of publications appearing in our Quarterly should include the full name of the journal in Polish and English without Polish diacritical marks, i.e.,

Eksploatacja i Niezawodnosc – Maintenance and Reliability.

dr inż. Leszek Gomółka

JAWO TECH S.R.O. Sp. z o.o. Oddział w Polsce
Niepodległości 67, 44-370 Pszów
E-mail: leszek.gomolkaa@gmail.com

dr hab. inż. prof. PO Andrzej Augustynowicz

Wydział Mechaniczny
Politechnika Opolska
Mikołajczyka 5, 45-271 Opole
E-mail: a.augustynowicz@po.opole.pl

Ocena przydatności stałej dielektrycznej do monitorowania procesu starzenia oleju silnikowego

Słowa kluczowe: *eksploatacyjne zużycie oleju, diagnostyka oleju silnikowego, stała dielektryczna*

Streszczenie: Celem pracy było określenie związków pomiędzy wytypowanymi właściwościami fizykochemicznymi oleju silnikowego a zmianą jego stałej dielektrycznej, rozważanej jako parametr oceniający stan jego zestarzenia. W celu szczegółowego poznania wybranych zjawisk towarzyszących starzeniu się oleju zdecydowano się na przebadanie oleju w różnych stadiach jego użytkowania. Poza badaniem stałej dielektrycznej wykonano również badania całkowitej liczby zasadowej TBN oraz badania spektrometryczne w podczerwieni IR. Stopień współzależności badanych wielkości przedstawiono wykorzystując do tego współczynnik korelacji liniowej Pearsona. W efekcie oceniono zmienność całkowitej liczby zasadowej oraz absorbancji IR w kontekście zmiany stałej dielektrycznej. Po przeprowadzeniu badań eksperymentalnych, stwierdzono, że uzyskane wyniki wskazują na zadawalającą korelację pomiędzy wytypowanymi własnościami fizykochemicznymi oleju, co pozwoliło z dużą dozą pewności przyjąć, iż stała dielektryczna może być użyta w charakterze parametru diagnostycznego do monitorowania stanu oleju silnikowego.

1. Wprowadzenie

W dobie wciąż intensywnie rozwijającej się techniki, stosowaną jednostką napędową pojazdów, maszyn i urządzeń jest nadal spalinowy silnik tłokowy. Należy sądzić, że pomimo wad, jeszcze przez kilka lat będzie w tej roli dominował [9,12]. Jednym z najważniejszych układów silnika jest system smarowania, którego integralnym czynnikiem jest olej smarujący. W celu prawidłowej weryfikacji własności użytkowych oleju w czasie eksploatacji, rodzi się potrzeba łatwego i precyzyjnego diagnozowania jego jakości. Fakt ten, zmusza nas do wytypowania najbardziej reprezentatywnego parametru, za pomocą którego możliwe jest monitorowanie zachodzących w nim zmian podczas użytkowania. Nowoczesne oleje składają się z różnych komponentów i dodatków co pozwala osiągać i spełniać szereg surowych wymagań eksploatacyjnych. Znaczące jest więc przyjęcie i dobór takich metod badawczych oraz wielkości, które spełniają wymagania parametru diagnostycznego, za pomocą którego jest możliwa analiza jakości oleju. Ocena zużycia eksploatacyjnego oleju silnikowego należy obecnie do najbardziej dynamicznie rozwijającego się nurtu badań i jest przedmiotem zainteresowania wielu instytucji naukowych [7]. Tematyka ta wzbudza również coraz to szersze zainteresowanie firm prywatnych eksploatujących pojazdy ze względu na istotne dla

nich efekty praktyczne i ekonomiczne [5,6]. Wszystkie mechanizmy czy podzespoły urządzeń ulegają w naturalny sposób procesom starzenia i zużycia, tak i olej silnikowy zmienia swoje właściwości smarne, niekorzystnie wpływając na stan techniczny całego układu smarowania. Niejednorodne warunki eksploatacji pojazdów i maszyn bezpośrednio wpływają na intensywność zmian zachodzących w substancjach smarnych a więc narzucają indywidualne podejścia do oceny jakości oleju, uwzględniając przyjęty okres użytkowania silnika. Z tego powodu prace badawcze poprowadzono w kierunku poszukiwania właściwości, które efektywnie i dokładnie odzwierciedlają charakter zmian eksploatacyjnych i które z kolei można w prosty sposób mierzyć i rejestrować.

Na bazie uzyskanych charakterystyk zmian wytypowanych właściwości fizykochemicznych oleju oraz silnych korelacji zmian stałej dielektrycznej z porównywanymi właściwościami zaproponowano nową metodykę diagnozowania jakości oleju. Zatem uznano za celowe wnikliwe podejście do badań nad stanem oleju smarującego, ponieważ szczególnie uzasadnione jest to w takich gałęziach gospodarki jak przemysł maszynowy – głównie samochodowy, rolniczy.

Niniejsza praca została ukierunkowana na wykazanie związku pomiędzy właściwościami fizykochemicznymi eksploatowanego oleju a stałą dielektryczną, która jest łatwo mierzalna. Przeanalizowano związki pomiędzy podstawowymi parametrami opisującymi proces starzenia oleju na podstawie wyników pomiarów zmian wartości stałej dielektrycznej oleju w czasie eksploatacji, w odniesieniu do olejów świeżych. Ze względu na ekspansywną elektronizację pojazdów szczegółowo scharakteryzowano procedury diagnozowania stanu oleju oraz metody prognozowania jego wymiany opierając się na wytycznych producentów pojazdów a także logice algorytmów z tym związanych. Na tej podstawie, jako drugorzędny cel zaproponowano zarys nowej metody umożliwiającej podjęcie właściwej decyzji o wymianie oleju silnikowego w trakcie eksploatacji.

2. Założenia wstępne

Z dokonanego przeglądu wybranych instrukcji obsługi samochodów wynika, że dla producenta pojazdu wskazanie użytkownikowi harmonogramu wymiany oleju jest zaleceniem niejednoznacznym ze względu na brak warunków technicznych do przeprowadzenia w pojeździe oceny jego zdatności. Zaawansowany poziom elektrotechniki i elektroniki umożliwia stosowanie różnych wielkości mierzonych oraz stosowanych form kontroli układów mechatronicznych [2,4]. Jednak z przeprowadzonej analizy stanu wiedzy wynika, że jednym z podstawowych i trudnych do zdefiniowania procesów fizykochemicznych jest starzenie oleju smarującego w silnikach spalinowych. Proces ten można opóźnić poprzez dodawanie w czasie eksploatacji do oleju dodatków uszlachetniających które wpływają na wzrost jego własności przeciwzatarciowych i przeciwzużyciowych. Uzyskuje się również spowolnienie tendencji spadkowej wartości liczby zasadowej [10,11]. Pomiar całkowitej liczby zasadowej (TBN) należy do rutynowych badań i jest testem określającym poziom utraty rezerwy alkalicznej w czasie eksploatacji. Przeprowadza się go również w celu monitorowania stanu dodatków uszlachetniających olej.

Proponowane przez producenta wymiany oleju są często wykonywane na długo przed zanim oleje zaczyna znacząco tracić właściwości smarujące. Sprawdzenie właściwości oleju przy użyciu laboratoryjnych metod analitycznych jest kosztowny. Korzystanie z tych technik podczas eksploatacji może być niezwykle trudne. W związku z tym rzadko diagnozuje się olej przed wymianą, co powoduje jego marnowanie.

W czasie użytkowania oleju silnikowego zmienia się wiele jego parametrów, jak choćby lepkość kinematyczna, skład frakcyjny, zawartość dodatków, zawartość zanieczyszczeń.

Badanie lepkości kinematycznej w temperaturze 40 i 100 °C wykazuje wąski obszar zmian w procesie użytkowania. Parametr diagnostyczny oparty na parametrze lepkościowym obarczony jest małą nośnością informacyjną, szczególnie w zmiennych warunkach temperaturowych.

Olej smarujący silnik jest typowym materiałem dielektrycznym. W pracy [7] wykazano, że spektroskopia dielektryczna jest jedną z najpotężniejszych i najbardziej dominujących technik analizy strukturalnej dla materiałów dielektrycznych. W łatwy sposób identyfikuje olej według klasyfikacji lepkościowej. Przypuszcza się że spektroskopia dielektryczna może dostarczyć przydatnych informacji na temat składu i struktury oleju silnika. Podobnie w pracy [14] wykazano, że współczynnik temperaturowy stałej dielektrycznej olejów silnikowych zależy od tego, czy olej jest świeży, czy też był używany. Stąd uznano, że pomiar stałej dielektrycznej daje możliwości do określania stanu oleju silnikowego.

Do podobnego wniosku doszli badacze w publikacji [16] w której zaproponowano zasadę monitorowania zanieczyszczenia oleju przez Internet w oparciu o pomiar stałej dielektrycznej. Opracowano również system pomiarowy, który obejmuje czujnik pojemności, mały obwód wykrywania pojemności oraz oprogramowanie do monitorowania i analizy. Wyniki pokazują, że względna stała dielektryczna oleju może być skutecznie śledzona i odpowiednio klasyfikowana za pomocą opracowanego systemu pomiarowego, który może być użyty do określenia właściwego okresu wymiany oleju.

Pomiar stałej dielektrycznej oleju może dostarczyć ważnych informacji na temat jakości oleju. Dedykowany do tego prosty test może w pewnych okolicznościach natychmiast określić, czy po wymianie oleju zastosowano właściwy olej. Zwiększenie się stałej dielektrycznej oleju wskazuje na obecność zanieczyszczeń lub zmianę jego składu chemicznego [3,15]. Ustalono, że szybkość zmiany stałej dielektrycznej zależy od fizycznych właściwości oleju, które wpływają na szybkość osadzania się w nim zanieczyszczeń, takich jak gęstość i lepkość, a także od użytych dodatków uszlachetniających.

Silnikowe oleje syntetyczne i mineralne nie są doskonałymi izolatorami, charakteryzują się minimalnym przewodnictwem elektrycznym. Dlatego znajomość właściwości dielektrycznych, takich jak właśnie względna przenikalność elektryczna jest istotna dla rozwoju i prawidłowego działania współczesnych silników spalinowych. Monitorowanie stanu oleju i jego degradacji w systemie online oraz prognozowanie jego wymiany jest celem współczesnych badaczy [18]. Monitorowanie stanu oleju i jego degradacji w systemie online, przy użyciu stałej dielektrycznej wiąże się z poznaniem jej zmian w funkcji temperatury. Wykonane badania [8,13] wykazały istotny wzrost stałej dielektrycznej w obszarach wyższych temperatur co ułatwia jej pomiar i klasyfikację. W pracy przedstawiono również możliwość wykorzystania czujnika pojemnościowego oraz jego korelację ze stałą dielektryczną. Stwierdzono, że w chwili zmiany stałej dielektrycznej oleju, pojemność również się zmienia.

Spektrometria w podczerwieni (IR) pozwala oceniać jakość oleju oraz zachodzące w nim zmiany poprzez analizę widma. Jest to analiza instrumentalna oparta na badaniach widm elektronowych związków chemicznych. Jest również metodą stosowaną do oznaczania zawartości niektórych związków chemicznych w mieszaninach, dla których pomiar jest dokonywany przy ustalonej długości fali. W efekcie badań na podstawie uzyskanych widm i widm wzorców wnioskuje się o zawartości związku w badanej substancji. Metody spektroskopii w podczerwieni są stosowane między innymi do oznaczania zawartości składników niepożądanych oraz zawartości niektórych dodatków [1].

Mając na uwadze powyższe zdecydowano się na zbadanie takich parametrów jak: stała dielektryczna, liczba zasadowa oraz poziom absorbancji.

3. Program badań

Na potrzeby badań określono pięć głównych jego etapów:
- wytypowanie obiektów do pobrania próbek oleju,
- opracowanie metodyki pobierania oleju,
- przebadanie wpływu temperatury na zmianę wartości stałej dielektrycznej oleju,
- przeprowadzenie badań laboratoryjnych próbek,
- dokonanie analizy otrzymanych wyników.

Ważną kwestią było ustalenie odpowiedniej reprezentatywnej ilości oleju potrzebnej do analiz laboratoryjnych. Dla celów porównawczych z góry określono badania mające na celu zbadanie degradacji oleju podczas użytkowania standardowego przebiegu pojazdu – 15 000 km. Po każdorazowym pobraniu próbki (od 100 do 150 ml), sprawdzano poziom oleju w silniku i uzupełniano go do poziomu zalecanego przez producenta. Wykonano również serię próbną bez dolewania oleju po pobraniu próbki w celu sprawdzenia takiego charakteru eksploatacji, spodziewając się szybszego starzenia oleju. Ze względu na długoterminowe badanie tego typu zdecydowano na przeprowadzenie tylko jednej, takiej serii na jednym obiekcie badań.

4. Obiekt badań

Jako obiekty badań wytypowano sześć samochodów osobowych klasy średniej, użytkowane w systemie poza miejskim i miejskim, głównie jako dojazd do miejsca pracy. Minimalny pokonywany dystans dzienny wynosił ok. 40 km. Samochody zasilane były benzyną, alternatywnie benzyną i LPG oraz ON. Każdy z pojazdów eksploatowany był przez jednego kierowcę. Pojazdy użytkowane były na olejach półsyntetycznych i syntetycznych.

Dobór obiektów badań uwarunkowano również, takimi czynnikami jak:
- popularność samochodów klasy średniej na krajowym rynku motoryzacyjnym,
- zróżnicowany wiek i przebieg pojazdów,
- zbliżony charakter eksploatacji (ruch miejski, dzienne przebiegi, itp.),
- dyspozycyjność pojazdów do cyklicznego pobierania próbek.

W tabeli 1 zestawiono podstawowe dane pojazdów.

Tabela 1. Zestawienie badanych pojazdów

Lp	Marka	Rodzaj paliwa	Pojemność silnika [dm^3]	Rodzaj oleju	Przebieg [km]
1	Opel Astra 1,7D	Olej napędowy	1,7	Lotos synt. 5W/40; API SN/CF, ACEA A3/B4	brak danych
2	Daewoo Nexia 1,5	Benzyna + LPG	1,5	Lotos synt. 5W/40;API SN/CF, ACEA A3/B4	98 400
3	Fiat Punto 1,1	Benzyna	1,1	Elf Semi-Syntetic 10W/40;API SL/CF, ACEA A3/B4	81 700
4	Opel Astra 1,6 *dwie serie badań S1 i S2*	Benzyna + LPG	1,6	Genuine GM 10W/40;API SL/CF, ACEA A3/B3	125 000
5	Volkswagen Passat 1,8	Benzyna + LPG	1,8	Castrol Magnatec 10W/40;API SL /CF, ACEA A3/B3	242 038
6	Fiat Seicento 1,1 *dwie serie badań S1 i S2*	Benzyna	1,1	Mobil 1 Formula S 10W/40;API SL/SJ/CF, ACEA A3/B3	99 382

5. Badania laboratoryjne

Oprócz przyjętych założeń wytypowania obiektów badań określono warunki techniczne pozyskania i przechowywania próbek oleju do badań laboratoryjnych. Prawidłowy schemat czynności przedstawiał się następująco:
- pobieranie próbek z silnika realizowano zawsze po jego nagrzaniu; w przypadkach długotrwałego postoju pojazdu pobranie próbki poprzedzał przejazd dystansu

zapewniającego osiągnięcie przez silnik równowagi cieplnej. Gwarantowało to dobre wymieszanie oleju w magistrali oraz sprawiało, że z oleju została odprowadzona woda,
- ze względu na relatywnie wysoką temperaturę oleju w chwili wykonywania pobrania, zachowywano szczególną ostrożność w zakresie BHP,
- jako przyrządy do pobierania próbek wykorzystano strzykawkę wraz z giętkim wężykiem olejoodpornym,
- miejscem, z którego pobierano stałą, określoną ilość oleju była rurka osłonowa wskaźnika poziomu oleju,
- pobierano określoną ilość oleju – od 100 do 150 ml,

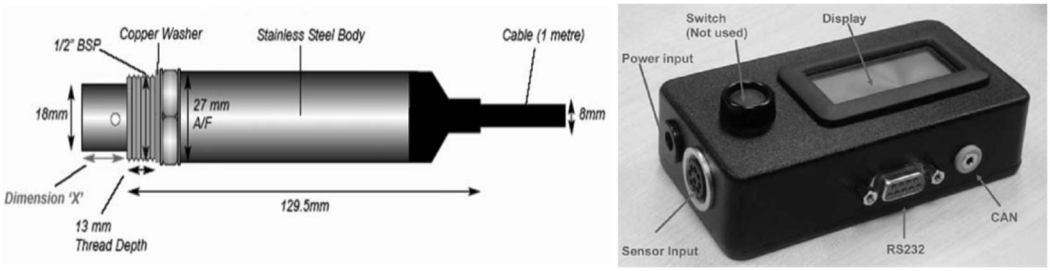
Zachowując powyższe wytyczne podjęto wszelkie starania wyeliminowania jakiekolwiek błędów już w trakcie pobierania próbek, co zapewniało takie same warunki dla wszystkich badanych pojazdów. W celu przechowywania olejów w różnych stadiach eksploatacji zakupiono specjalne dedykowane pojemniki ze szkła laboratoryjnego. Stanowi to bardzo przydatne rozwiązanie do magazynowania oleju do czasu przeprowadzenia badań. Zastosowane szkło laboratoryjne chroni olej od dodatkowych zanieczyszczeń i niekontrolowanego starzenia podczas magazynowania.

Ze względu na dobór obiektów oraz metodyki badań, analiza dotyczyła zmian właściwości oleju w silniku o zapłonie ZI i ZS. Takie ujęcie tematu potrzebne było ze względu na faktyczną utratę właściwości fizykochemicznych, zależących od charakteru użytkowania i stanu technicznego pojazdu oraz dokonywanych aktualnie wymian oleju po przebyciu danego przebiegu. Poprzez przeprowadzenie badań laboratoryjnych zweryfikowano sensowność przebiegowych wymian oleju oraz możliwość szybkiej i skutecznej oceny jego stanu na podstawie przyjętego parametru diagnostycznego.

Stanowisko badawcze wyposażone było w niezbędne urządzenia. Przed każdorazowym pomiarem czujnik był myty z resztek oleju za pomocą benzyny ekstrakcyjnej oraz wyczyszczony za pomocą ręcznika papierowego a na koniec suszony powietrzem pod ciśnieniem. Zapewniało to rzetelny pomiar oraz dołożono wszelkich starań aby na elektrodach pomiarowych czujnika Analexrs nie znajdował się olej z poprzedniego badania. Kolejnym bardzo ważnym momentem w procesie badawczym było zapewnienie jednorodności oleju w próbkach. Ze względu na czasowe magazynowanie pozyskanych próbek, olej poddawany był niepożądanej sedymentacji, co prowadziło do wytrącenia osadów oraz rozwarstwień. Niektóre nawet procesy są niedostrzegalne „okiem nieuzbrojonym", dlatego dla zapewnienia jednorodności zawiesiny, przed pomiarem z każdej badanej serii, próbki oleju delikatnie wstrząśnięto w celu wymieszania się osiadłych produktów starzeniowych.

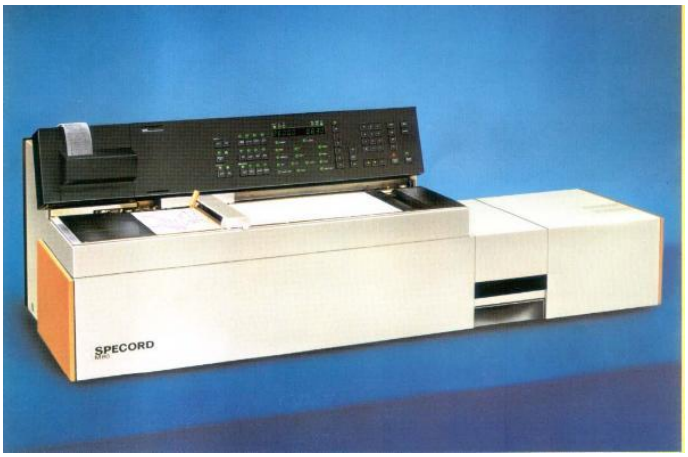
Do wykonanych podstawowych badań zaliczyć można:
- pomiar stałej dielektrycznej przyrządem Lubri Sensor,
- pomiar stałej dielektrycznej przyrządem Analexrs,
- zmienność całkowitej liczby zasadowej TBN,
- badania spektrometryczne w podczerwieni, zmian poziomu absorbancji – IR.

Wykorzystanymi przyrządami mierzącymi jakość oleju poprzez określenie cechy oleju, zwanej również stałą dielektryczną to: mechatroniczny czujnik Analexrs oraz przenośny miernik Lubri–Sensor przedstawione na rys. 1 i 2. Zmiany absorbancji promieniowania IR w funkcji przebiegu przebadano za pomocą spektrofotometru siatkowego Specord M80 (rys. 3).



Rys.1. Zestaw do pomiaru stałej dielektrycznej Analexrs [19]

Rys.2. Zestaw do pomiaru stałej dielektrycznej Lubri Sensor [20]



Rys.3. Ogólny widok spektrofotometru Specord M80 [21]

6. Analiza wieloaspektowa

Istotą koncepcji będącej przedmiotem publikacji jest analiza degradacji oleju na podstawie zmian stałej dielektrycznej w wybranych stanach pracy. Założenie takie wiązało się z koniecznością szczegółowego poznania wpływu pozostałych własności fizykochemicznych procesu starzenia oleju i określenia na podstawie przeprowadzonej identyfikacji zmian funkcji korelacji pomiędzy nimi. Mimo wielu opracowań, wiedza w tym zakresie nie jest pełna i wymaga uzupełnienia zwłaszcza z uwagi na coraz to nowe komponenty olejowe, które

w założeniu powinny zapewnić jak największe spełnienie parametrów technicznych i ekologicznych przy uzyskaniu minimalnych kosztów produkcji.

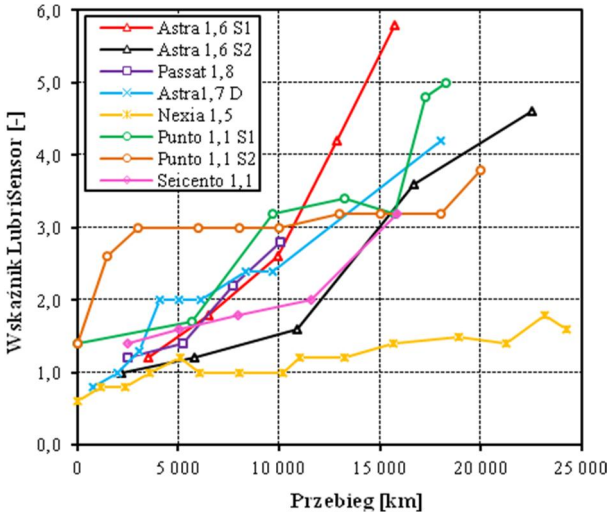
6.1. Stała dielektryczna

Pomiar stałej dielektrycznej wykonano w oparciu o dwa urządzenia: Lubri Sensor i Analexrs. Charakterystyki badanego parametru przedstawiono i opisano poniżej.

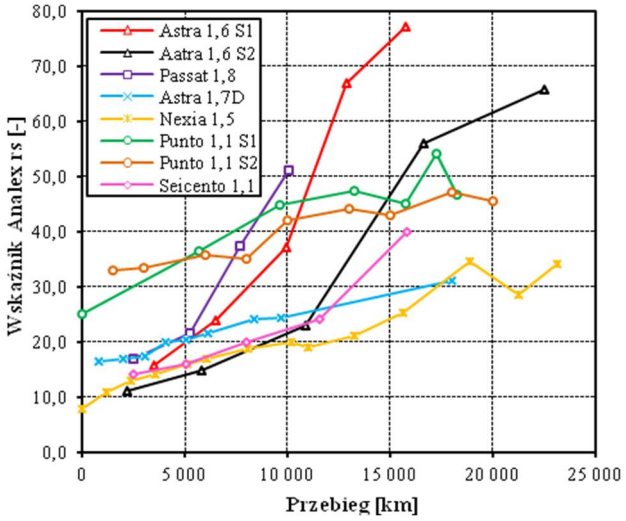
Na wykresach zbiorczych – rys.4, przedstawiono charakterystyki zmian stałej dielektrycznej dla poszczególnych obiektów badań. Dla przedmiotowych analiz odniesieniem był przebieg pojazdu, po przebyciu którego dokonywano w poszczególnych obiektach badania oleju.

Analizując przedstawiane wykresy zmian stałej dielektrycznej mierzonej dwoma niezależnymi przyrządami stwierdzono, iż widoczny jest znaczący wzrost wartości stałej dielektrycznej podczas użytkowania. Niewielkie różnice wynikają z sposobu pomiaru oraz nominalnej skali zastosowanych przyrządów, jak również z odstępu czasowego wykonanych badań. Stwierdzono zbliżony trend zmian charakterystyki w odniesieniu obu serii wykonanych badań na tej samej jednostce napędowej Fiata Punto, zmierzone Analex'em i Lubri Sensorem. Podobnie jest w przypadku przebiegu funkcji (podobieństwo jakościowe) w seriach Opla Astry 1,6. Dla serii S1 uzyskano szybszą degradację oleju w trakcie eksploatacji bez stosowania dolewek co skutkowało wcześniejszym osiągnięciem wartości granicznych wymagających wymiany oleju. Jest to równoznaczny dowód na to, iż konieczne jest sprawdzanie poziomu oleju w silniku podczas użytkowania, w celu utrzymywania wymaganej jakości oleju smarującego. Zapewnia to bezawaryjną pracę układu smarowania, a stosowane dolewki „odświeżają" olej i tym samym spowalniają proces jego stężania – po osiągnięciu przebiegu na poziomie 25 000 km wskaźnik stałej dielektrycznej osiągnął poziom 65 jednostek (Analexrs) dla serii w Oplu Astrze 1,6 z uzupełnianym olejem. Charakterystyki zmian przyjętego parametru diagnostycznego wyróżniają się wysokimi współczynnikami korelacji, mieszczącymi się w zakresie od R^2=0,87 do 0,98.

a)



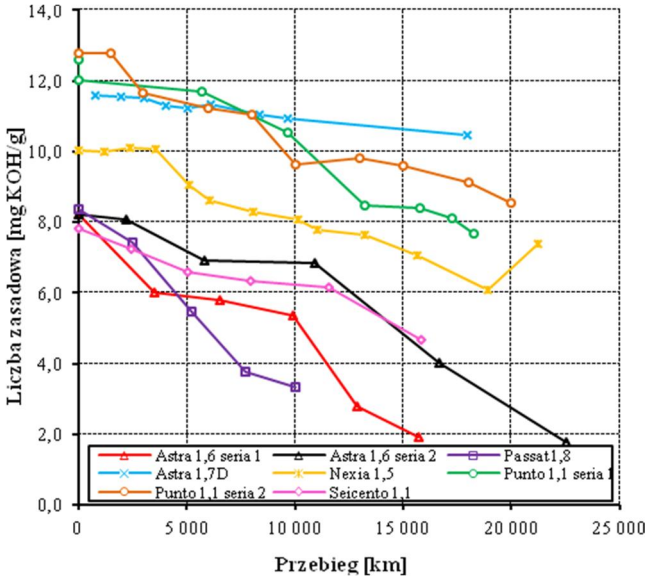
b)



Rys.4. Zmiana stałej dielektrycznej mierzonej za pomocą: a) Lubri Sensor, b) Analexrs

6.2. Całkowita liczba zasadowa TBN

Pomiar liczby zasadowej wykonano metodą miareczkowania potencjometrycznego wg PN – 88/C-04049. Jednoznacznie wynika, że stosowanie dodatków do olejów decyduje o odpowiednim poziomie liczby zasadowej w procesie użytkowania. Z tego też względu oleje różnych gatunków nie mogą być ze sobą mieszane. Takie działanie prowadzi do tego, że oleje w silniku mogą reagować ze między sobą, powodując zmniejszenie liczby zasadowej i powstawanie związków chemicznych obojętnych, a nawet szkodliwych dla silnika. Poniżej na rysunku 5 zobrazowano charakterystyki zmian tej właściwości w badanych pojazdach.



Rys.5. Zmiana liczby zasadowej

Z przeprowadzonej analizy TBN w oleju podczas użytkowania stwierdzono, że bardzo ważna jest początkowa wartość liczby zasadowej stanowiąca rezerwę alkaliczną na dalszy okres eksploatacji. W prezentowanych wynikach startowa wartość TBN mieści się w zakresie od 8 do 13 [mgKOH/g] i głównie zależy od rodzaju oleju silnikowego danego producenta – im wyższa tym dłuższy okres użytkowania. Dla Fiata Punto, po porównaniu obu serii danych,

zauważono bardzo zbliżoną utratę rezerwy alkalicznej (zmniejszającą się liczbę zasadową) co świadczy najprawdopodobniej o podobnym sposobie użytkowania pojazdu. Z kolei dla samochodu Opel Astra 1,6 (seria pierwsza) stwierdzono efekt przyspieszonego starzenia się oleju ponieważ nie dokonywano dolewek kompensujących po każdorazowym pobieraniu oleju do badania. Bazą odniesienia była seria druga samochodu Opel Astra 1,6, w trakcie której po pobraniu oleju do badań niedobór uzupełniano olejem świeżym.

Ciekawym przypadkiem jest przebieg zmian liczby zasadowej dla samochodu Opel Astra 1,7D zasilanego olejem napędowym. Dla tego pojazdu odnotowano najmniejszy spadek rezerwy alkalicznej. Z uwagi na to, iż do badań wykorzystano tylko jeden pojazd z silnikiem o ZS nie jest możliwe jednoznaczne wyjaśnienie tego faktu.

Kolejnym interesującym zestawieniem jest para pojazdów, VW Passat 1,8 zasilany benzyną i LPG oraz Fiat Seicento 1,1 zasilany benzyną. Porównując na rys.5 przebiegi dla tych dwóch pojazdów zauważono gwałtowniejszy spadek liczby zasadowej dla pojazdu VW w stosunku do Fiata, którego olej posiadał na początku badań niższą, o 0,52 mgKOH/g, liczbę zasadową . Zjawisko można tłumaczyć tym, że pojazdy różniły się pojemnością jednostki napędowej a nadto VW Passat posiadał na początku badań znacząco wyższy przebieg od Fiata Seicento (Tabela 1).

Zmiana wartości liczby zasadowej w eksploatacji, czyli zmniejszanie się rezerwy alkalicznej jest procesem znanym a progres ten zależy głównie od użytkowania, stanu jednostki napędowej, czyli podstawowych czynników wpływających na degradację oleju. Wstępną analizę obiektów można dokonać pod kątem zastosowanego oleju silnikowego, który charakteryzuje się różnymi wartościami początkowej liczby zasadowej. Mając na uwadze powyższe zaproponowano parametr za pomocą którego można porównać stopień zmian pomiędzy konkretnymi pojazdami. Dla przykładu wzięto pod uwagę tylko parametry brzegowe zmian liczby zasadowej, czyli TBN_{max} i TBN_{min}, po czym odniesiono to do przebiegu oleju. Dla szerszej analizy możemy rozpatrywać zmianę w poszczególnych okresach użytkowania.

Dla obliczenia ΔTBN użyto wzoru:

$$\Delta TBN = TBN_{max} - TBN_{min} \tag{1}$$

Zmianę procentową wyznaczono ze wzoru:

$$zTBN = 100\% - \frac{TBN_{min} \cdot 100\%}{TBN_{max}} \tag{2}$$

Porównując zmianę procentową liczby zasadowej (utratę procentową rezerwy alkalicznej) w stosunku do przebiegu pojazdu możemy porównać pomiędzy sobą wszystkie badane pojazdy (Tabela 2). Znając wytyczne producenta konkretnego oleju silnikowego co do momentu wymiany oleju w kontekście spadku rezerwy alkalicznej jesteśmy w stanie wyznaczyć ten moment poprzez korelację z przebiegiem pojazdu lub czasookresem (motogodziny). W tym momencie można również określić czy podawany przez producenta czasookres wynikający z przebiegu lub godzin pracy jest prawidłowym wyznacznikiem czynności serwisowych.

Tabela 2. Zmiana liczy zasadowej

Rodzaj pojazdu	Liczba zasadowa TBN_{max} [mgKOH/g]	Liczba zasadowa TBN_{min} [mgKOH/g]	ΔTBN [mgKOH/g]	Zamiana procentowa zTBN [%]	Przebieg [km]
Punto 1,1 seria 1	12,59	7,69	4,90	38,92	18250
Punto 1,1 seria 2	12,78	8,54	4,24	33,18	20000
OpelAstra 1,7D	11,58	10,45	1,13	9,76	17983
Daewoo Nexia 1,5	10,2	7,04	3,16	30,99	24259
Astra 1,6 seria 1	8,22	1,94	6,28	76,40	15730
Astra 1,6 seria 2	8,22	1,79	6,43	78,22	22529
VW Passat 1,8	8,36	3,34	5,02	60,05	10038
Fiat Seicento 1,1	7,84	4,70	3,14	40,05	15828

Z analiz wynika, że w niektórych przypadkach przyjęty przebieg z jednej strony jest zbyt krótki a z drugiej za długi, patrząc pod kątem samochodu VW Passat, gdzie liczba zasadowa po przebiegu 10 000 km osiągała wartość 3,34 [mgKOH/r], co stanowi 60% utraty rezerwy alkalicznej. Ewenementem jest wynik dla Opla Astry 1,6 seria druga, gdzie mimo niskiej wartości początkowej liczby zasadowej, stwierdzono znaczący spadek rezerwy alkalicznej do około 78%, dla serii w której olej był podczas pobierania próbek uzupełniany olejem świeżym.

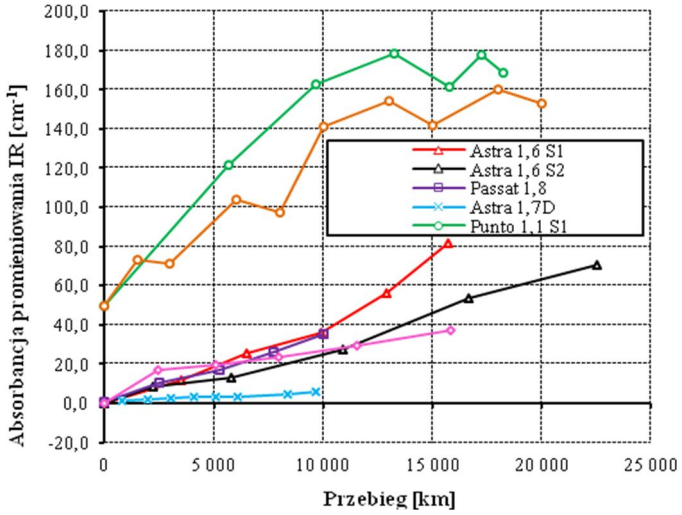
Z porównań również wynika, że ważnym parametrem oleju jest jego startowa liczba zasadowa co oznacza, że należy stosować oleje lepszej jakości o właściwie dobranej początkowej TBN. Dopuszczalne limity TBN w oleju są regulowane przez ACEA (Association des Constructeurs Européens d'Automobiles) i zależą od stosowanego paliwa. Pojazdy napędzane benzyną, gazem LPG i CNG oraz olejem napędowym o ultra niskiej zawartości siarki nie wymagają wysokiego TBN. Natomiast zasilanie silnika biopaliwem wiąże się z doborem oleju o wyższej wartości TBN [17].

Po dokonaniu szczegółowej analizy zmiany całkowitej liczby zasadowej celowe zatem jest zestawienie wyników pomiarowych z rozważanym parametrem diagnostycznym wyznaczanym za pomocą wskaźnika Lubri Sensor i Analexrs.

6.3. Rozkładu widma promieniowania podczerwonego w paśmie tlenowym IR

Rozkład widma (absorbancja) promieniowania podczerwonego dla olejów silnikowych jest najbardziej efektywny w paśmie tlenowym, tzn. w zakresie liczby falowej około 1824-1520 [cm^{-1}]. W tym obszarze następują znaczące zmiany eksploatacyjne, które możemy stosunkowo w prosty sposób badać oraz poddawać dalszej analizie. Podstawowe warunki rejestracji widma IR to: spektrofotometr typu SPECORD M80, zakres pomiarowy: 1860 - 1480 cm^{-1}, grubość warstwy: 0,105 mm, rozdzielczość: 4 cm^{-1}, jako odnośnik zastosowany olej świeży, właściwy dla każdej serii badanych pojazdów. Zarejestrowane widma IR przeliczono wg linii bazowej opartej na punktach 1764 cm^{-1} i 1516 cm^{-1}, a następnie scałkowano numerycznie w granicach wyznaczonych przez punkty bazowe. Na rys. 6 przedstawiono zmianę absorbancji dla poszczególnych obiektów badawczych. Pomiar i analiza widm absorbcyjnych promieniowania w podczerwieni (IR) pozwala przedstawić dynamikę zmian termooksydacyjnych oleju w eksploatacji, lecz nie spełnia kryteriów parametru diagnostycznego. Określenie wartości granicznych kwalifikujących olej do wymiany jest bardzo trudne lub niemożliwe (szczególnie dla Opla Astry 1,7D), ponieważ otrzymane funkcje zmian wartości absorbancji nie posiadają charakterystycznych punktów przegięcia, odzwierciedlających zmianę stanu.

Na rys. 6 zauważono również, że zmiany wartości absorbancji dla Fiata Punto (w dwóch seriach) znacząco odbiegają od zmian pozostałych przebiegów. Zaistniały fakt można tłumaczyć tym, iż obliczenia wykonane dla Fiata Punto były realizowane w szerszym paśmie tlenowym niż dla pozostałych obiektów badań.

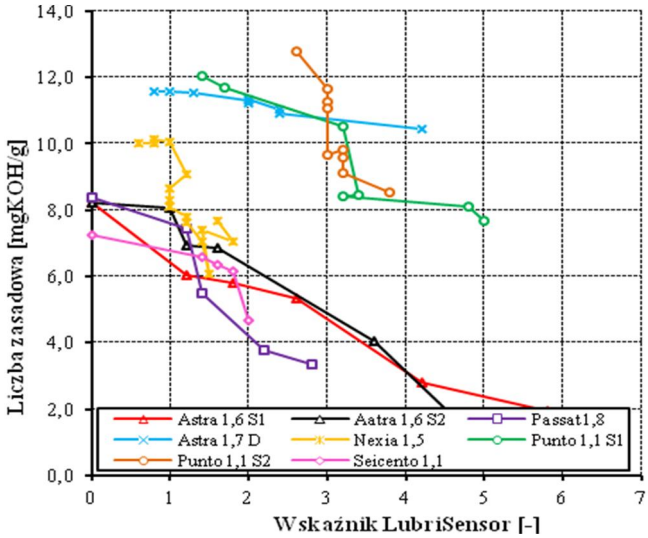


Rys.6. Rozkład widma promieniowania podczerwonego

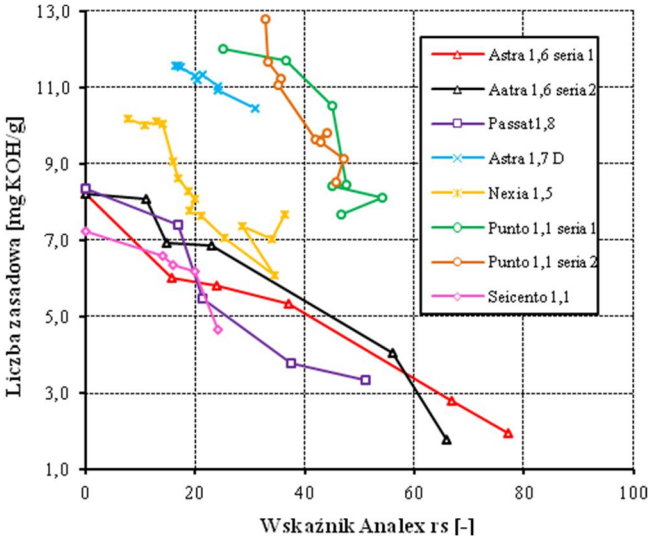
6.4. Analizy porównawcze

Wszystkie pomiary przedstawionych własności oleju silnikowego zostały wykonane z pewnym błędem wynikającym z niedoskonałości zastosowanych przyrządów pomiarowych. Z tych powodów niemożliwe jest absolutnie dokładne wyznaczenie wartości mierzonej wielkości i dlatego ważna jest ocena wiarygodności otrzymanych wyników pomiarowych. W związku z tym, dla lepszego zobrazowania zmienności oraz określenia współzależności pomiędzy badanymi własnościami eksploatacyjnymi oleju silnikowego, wykonano analizy porównawcze. Stopień współzależności przedstawiono wykorzystując współczynnik korelacji liniowej Pearsona. W tym celu porównano zmienność całkowitej liczby zasadowej (rys. 7) oraz absorbancji IR (rys. 8) w kontekście stałej dielektrycznej, z zamiarem zastosowania jej w charakterze parametru diagnostycznego.

a)

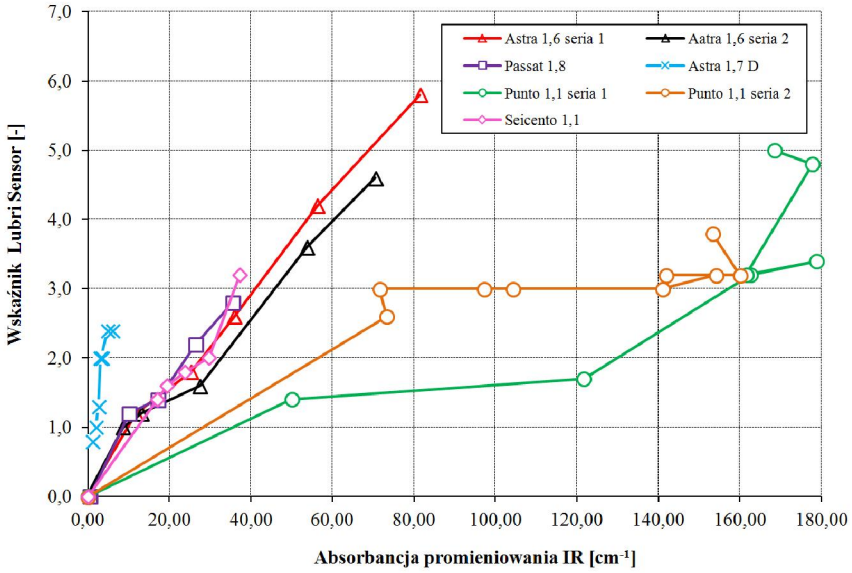


b)

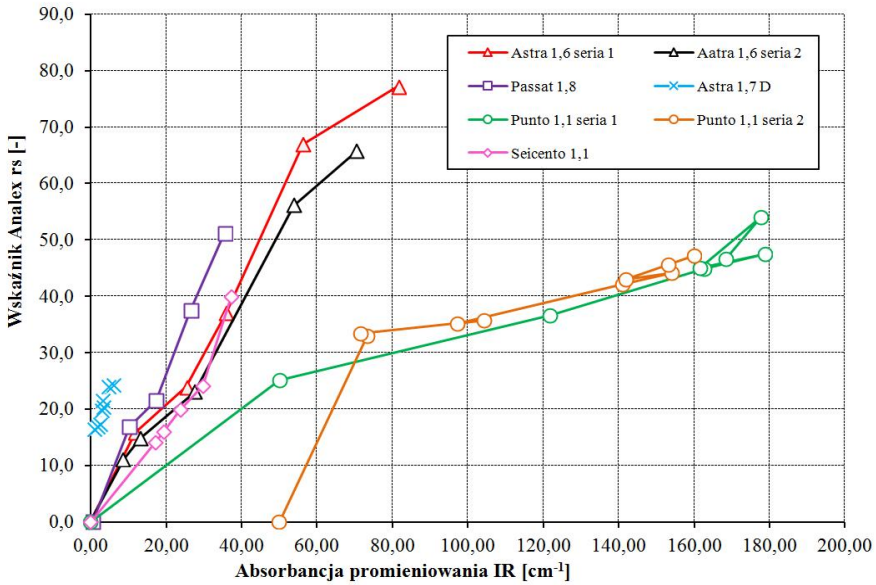


Rys.7. Zmiana stałej dielektrycznej w zestawieniu z liczbą zasadową, mierzonej za pomocą: a) Lubri Sensor b) Analexrs

a)



b)



Rys.8. Zmiana absorbancji promieniowania IR w zestawieniu z liczbą zasadową, mierzonej za pomocą: a) Lubri Sensor b) Analexrs

Współczynnik korelacji liniowej Pearsona to współczynnik określający poziom zależności liniowej między zmiennymi losowymi. Wykorzystano korelację do zbadania czy zachodzi związek pomiędzy dwiema zmiennymi (właściwościami, cechami). Współczynnik korelacji mówi nam o sile związku. Jest ona określana jako wartość w przedziale od 0 do 1. Im wartość współczynnika dąży do wartości 1 tym siła związku jest większa. Siła związków korelacyjnych:

poniżej 0,2 – korelacja słaba (praktycznie brak związku),
0,2 ÷ 0,4 – korelacja niska (zależność wyraźna),
0,4 ÷ 0,6 – korelacja umiarkowana (zależność istotna),
0,6 ÷ 0,8 – korelacja wysoka (zależność znaczna),
0,8 ÷ 0,9 – korelacja bardzo wysoka (zależność bardzo duża),
0,9 ÷ 1,0 – zależność praktycznie pełna.

Współczynnik korelacji Pearsona r_{xy}, zastosowano do badania związków prostoliniowych badanych zmiennych, w których zwiększenie wartości jednej z cech powoduje proporcjonalne zmiany średnich wartości drugiej cechy (wzrost lub spadek).

Współczynnik ten obliczono na podstawie wzoru:

$$r_{xy} = \frac{cov(x,y)}{Sd_x \cdot Sd_y} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\overline{x_x})^2 \cdot \sum(y_i-\overline{x_y})^2}}. \quad (3)$$

W tabelach 3÷8 przedstawiono zestawienie współczynników korelacji liniowej Pearsona dla poszczególnych pojazdów.

Tabela 3. Zestawienie współczynników korelacji dla Fiata Punto 1,1 - seria 1

Fiat Punto 1,1 (seria 1)	Wskaźnik Lubri-Sensor [-]	Wskaźnik Analex rs [-]
Wskaźnik Analex rs [-]	0,814	1,000
Liczba zasadowa [mgKOH/g]	0,823	0,715
Absorbancja promieniowania IR [cm^{-1}]	0,804	0,949

Tabela 4. Zestawienie współczynników korelacji dla Fiata Punto 1,1 - seria 2

Fiat Punto 1,1 (seria 2)	Wskaźnik Lubri-Sensor [-]	Wskaźnik Analex rs [-]
Wskaźnik Analex rs [-]	0,888	1,000
Liczba zasadowa [mgKOH/g]	0,72	0,897
Absorbancja promieniowania IR [cm^{-1}]	0,731	0,692

Tabela 5. Zestawienie współczynników korelacji dla Opla Astry 1,6 - seria 1

Opel Astra 1,6 (seria 1)	Wskaźnik Lubri-Sensor [-]	Wskaźnik Analex rs [-]
Wskaźnik Analex rs [-]	0,982	1,000
Liczba zasadowa [mgKOH/g]	0,966	0,974
Absorbancja promieniowania IR [cm^{-1}]	0,995	0,971

Tabela 6. Zestawienie współczynników korelacji dla Opla Astry 1,6 - seria 2

Opel Astra 1,6 (seria 2)	Wskaźnik Lubri-Sensor [-]	Wskaźnik Analex rs [-]
Wskaźnik Analex rs [-]	0,992	1,000
Liczba zasadowa [mgKOH/g]	0,960	0,953
Absorbancja promieniowania IR [cm^{-1}]	0,984	0,988

Tabela 7. Zestawienie współczynników korelacji dla Volkswagen'a Passat'a 1,8

Volkswagen Passat 1,8	Wskaźnik Lubri-Sensor [-]	Wskaźnik Analex rs [-]
Wskaźnik Analex rs [-]	0,926	1,000
Liczba zasadowa [mgKOH/g]	0,903	0,916
Absorbancja promieniowania IR [cm^{-1}]	0,974	0,990

Tabela 8. Zestawienie współczynników korelacji dla Opla Fiat Seicento 1,1

Fiat Seicento 1,1	Wskaźnik Lubri-Sensor [-]	Wskaźnik Analex rs [-]
Wskaźnik Analex rs [-]	0,982	1,000
Liczba zasadowa [mgKOH/g]	0,927	0,963
Absorbancja promieniowania IR [cm^{-1}]	0,962	0,946

Z przedstawionych w tabelach liczb wynika, że dominuje głównie korelacja wysoka i bardzo wysoka. Analiza szczegółowa uzyskanych wyników upoważnia do sformułowania następujących wniosków:

- pomiar stałej dielektrycznej był dokonywany przez dwa różniące się konstrukcją urządzenia, znanym od lat Lubri Sensorem który można traktować jak przenośny miernik oraz Analexrs, nowoczesnym czujnikiem mogącym w połączeniu z systemem akwizycji danych monitorować w czasie rzeczywistym względną przenikalność elektryczną. Pomimo różnej konstrukcji urządzeń i ich wskazań uzyskano bardzo wysoki współczynnik korelacji w przedziale 0,81÷0,99,
- spektroskopia w podczerwieni jako metodyka badań oleju pozwala na analizę jego składu i obserwację zmian jakie zachodzą w trakcie starzenia. Jednakże ustalenie jednoznacznego parametru charakteryzującego jakość oleju w wybranym czasie eksploatacji jest bardzo trudne. Mając na uwadze powyższe oraz to, że badanie wymaga użycia kosztownej aparatury eliminują tę metodę do zastosowania w pojeździe. Obliczona korelacja pomiędzy wartościami absorbancji promieniowania a stałej dielektrycznej zawiera się w przedziale 0,73÷0,99 dla Lubri Sensora i odpowiednio 0,69÷0,99 dla Analexrs.
- jak wspomniano wcześniej zmniejszanie się rezerwy alkalicznej w trakcie użytkowania oleju silnikowego jest procesem znanym z tego, że wpływa na degradację oleju.

W momencie określanym mianem punktu zrównoważenia dochodzi do stanu wyczerpania rezerwy zasadowej zapewniającej bieżącą neutralizację kwaśnych produktów spalania i wymiana oleju na świeży jest jak najbardziej wskazana. Ustalenie momentu wymiany oleju metodami diagnostyki pokładowej jest bardzo trudne. Obliczona korelacja pomiędzy wartościami liczby zasadowej a stałej dielektrycznej zawiera się w przedziale 0,72÷0,97 dla Lubri Sensora i odpowiednio 0,71÷0,97 dla Analexrs.

- pomiar stałej dielektrycznej czujnikiem Analexrs lub pochodnym, zamontowanym w strefie magistrali olejowej silnika spalinowego mógłby być źródłem informacji diagnostycznej – parametrem stanu oleju silnikowego.

7. Koncepcja autorskiej metody predykcji wymiany oleju

Etapem finalnym publikacji jest próba opracowania metodyki prognozowania wymiany oleju w ujęciu systemowym. Uznano, że w celu określenia optymalnego momentu wymiany oleju potrzebne są niezbędne informacje, takie jak:

- parametry oleju świeżego, zastosowanego lub dobranego (klasa jakości, klasa lepkości, wartość liczby zasadowej, itp.) - określającego punkt bazowy,
- warunki użytkowania pojazdu i profil kierowcy (zakres prędkości obrotowej silnika, obciążenie silnika, temperatura płynu chłodzącego, temperatura i poziom oleju, temperatura otoczenia, ciśnienie oleju w rożnych miejscach układu smarowania, liczba zimnych rozruchów, pokonywane dystanse, itp.),
- pomiary z dedykowanych czujników stanu oleju (stałej dielektrycznej, poziomu liczby zasadowej TBN, lepkości lub inne, które wchodzą na rynek).

Za typowanie momentu wymiany odpowiedzialny byłby algorytm sterowania, do którego na bieżąco podawane byłyby wszystkie dostępne informacje z magistrali danych (sieci komunikacyjnej) pojazdów i maszyn. Byłyby to dane określające degradację oleju (tworzenie charakterystyki degradacji oleju), takie jak np.:

- wartość procentowa utraty rezerwy alkalicznej,
- wartość lub poziom liczby zasadowej,
- stała dielektryczna jako wartość lub zmiana procentowa,
- czasookres pracy (kilometry, motogodziny, itp).

Układ uwzględniałby korekty pochodzące z pomiarów typu on-line, jak również podawane jako zewnętrzne parametry odniesienia (potocznie nazywanymi trudnymi warunkami użytkowania), za pomocą których byłby wyznaczany indywidualny czas wymiany, np.:

- rodzaj użytkowania,
- typ kierowcy lub operatora,
- tendencja zmian wartości mierzonych,
- warunki otoczenia, itp..

Całość prognozowania oparta byłaby na działającym w czasie rzeczywistym systemie, który bardzo łatwo można wprowadzić w obliczenia sieci neuronowych tworząc sztuczną inteligencję. Należy jednak podkreślić iż decydującą rolę odgrywa pomiar stałej dielektrycznej, która jako wielkość uniwersalna uwzględnia większość czynników wpływających na degradacje oleju (paliwo, sadza, szlamy, laki, żele, ścier ferromagnetyczny, woda, itp.) oraz mocno koreluje z wielkościami, które wykorzystuje się do precyzyjnych pomiarów laboratoryjnych (spektrometria, liczba zasadowa, ilość zanieczyszczeń, itp). W logikę układu można wprowadzać dowolną charakterystykę zmian poszczególnych własności fizykochemicznych oleju wyznaczaną poprzez korelację z wybranym, mierzonym parametrem.

W dobie nowoczesnych trendów i technologii w zakresie budowy układów smarowania pojazdów i maszyn, takich jak:

- nanotechnologia olejowa - nowoczesne dodatki do oleju poprawiające własności nie tylko smarne lecz utrzymujące na oczekiwanym poziomie poszczególne parametry pracy oraz odbudowujące strukturę par trących,
- downsizing - mała ilość oleju w misce olejowej, wysokie temperatury pracy, duże wysilenie silników,
- modułowa budowa układu olejowego, gdzie wymienia się zużyty olej wraz z filtrem na cały nowy moduł,
- układy pokładowego oczyszczania oleju,

precyzyjna diagnostyka zyskuje na znaczeniu i odgrywa ważną rolę. Degradacja oleju zależy od wielu czynników a zaproponowana koncepcyjna metoda opiera się na analizie wielu zmiennych, dlatego jest ujęta w "inteligentny" system. Do pracy algorytmu potrzebna jest jak największa liczba danych. Niewystarczające jest monitorowanie tylko jednej własności oleju i typowanie według niej momentu wymiany oleju.

Jest to nowa kategoria użytkowania, którą można wykorzystać w nowoczesnych pojazdach hybrydowych i autonomicznych. Zapewni to łatwość operacji eksploatacyjnych, efektywny recykling i minimalizację odpadów oraz spełnienie norm ekologicznych.

8. Wnioski

Przebieg jak i czasookres pracy pojazdu nie są właściwymi parametrami decydującymi o wymianie oleju na nowy. Konieczne jest wprowadzenie łatwej i taniej metody diagnostycznej oleju silnikowego, która pozwalałaby na pełne wykorzystanie potencjału jego właściwości fizykochemicznych przy jednoczesnym wyeliminowaniu oleju, który utracił swe właściwości i nie zapewnia już odpowiedniej ochrony silnika. W celu szczegółowego poznania wybranych zjawisk towarzyszących starzeniu się oleju przebadano olej w różnych stadiach jego użytkowania. Szczegółowe wnioski sformułowano następująco:

- Badanie widma IR potwierdziły charakter zmian termooksydacyjnych, które wpływają na degradację oleju w czasie eksploatacji.
- Badanie rezerwy alkalicznej wykazało tendencję malejącą wskazując na intensyfikację kwaśnego środowiska.
- Starzeniu się oleju towarzyszy zmiana względnej przenikalności elektrycznej. Stała dielektryczna jest łatwo mierzalnym parametrem. Pomiar stałej dielektrycznej czujnikiem zamontowanym w strefie magistrali olejowej silnika spalinowego mógłby być źródłem informacji diagnostycznej.
- Badanie korelacji zmian stałej dielektrycznej w odniesieniu do zmiennych jakimi są liczba zasadowa i absorbancja promieniowania IR wykazało wysoki poziom zależności co upoważnia wykorzystać stałą dielektryczną w charakterze parametru diagnostycznego do monitorowania stanu oleju silnikowego.
- Określenie optymalnego momentu wymiany oleju wymaga opracowania metodyki prognozowania wymiany oleju w ujęciu systemowym.

Literatura

1. Al-Ghouti M A, Al-Toum L. Virgin and recycled engine oil differentiation: A spectroscopic study, Journal of Environmental Management 2009; 90.
2. Bolognesi P, Bruno O, Landi A, Sani L, Taponecco L. Electric Machines and Drives for X-by-Wire System in Ground Vehicles 2003; 10th European Conference on Power Electronics and Application: 1-16.
3. Carey A A, Hayzen A J. The dielectric constant and oil analysis, Practicing Oil Analysis Magazine 2001; 9: .
4. Cheng H. Autonomous Inteligent Vehicles: Theory, Algorithms and Implementation. London Limited: Springer-Verlag, 2011.
5. Gomółka L. Ocena oleju silnikowego w eksploatacji. Praca doktorska, Wydział Mechaniczny, Politechnika Opolska, 2017.
6. Gomółka L, Augustynowicz A, Maciąg A. Analysis of the rank spreading the decline of oil in internal combustion engines, Comustion Engines 2011; 1 (144): 1-4.
7. Guan L, Feng X L, Xiong G. Engine lubricating oil classification by SAE grade and source based on dielectric spectroscopy data, Analytica Chimica Acta 2008; 628: 117-120.
8. Han Z, Wang YQing X. Characteristics Study of In-Situ Capacitive Sensor for Monitoring Lubrication Oil Debris, Sensors 2017; 17 (12) 2851:1-13.
9. Idzior M. The development of combustion engines in the aspect of their manufacturing, Comustion Engines 2006; 1 (124): 60-70.
10. Laber A, Adamczuk K. Improving the working conditions of friction by using operation additives to lubricating oil, Tribologia 2014; 1: 27-39.
11. Laber S, Laber A. Property assessment lubricants and tribological lotos dynamic engine oil, Tribologia 2015; 3: 89-97.
12. Serrano J R. Imagining the Future of the Internal Combustion Engine for Ground Transport in the Current Context, Applied Sciences 2017; 7 (10) 1001: 1-5.
13. Stevan Junior S L, Paiter L, Galvao J R, Roque D V, Chaves E S. Sensor and Methodology for Dielectric Analysis of Vegetal Oils Submitted to Thermal Stress, Sensors 2015; 15 (10): 26457-26477.
14. Torrents J M, Pallas-Areny R. Sensing oil condition through temperature coefficient of dielectric constant 2003, XVII IMEKO World Congress - Metrology in the 3rd millennium, Dubrovnik, Croatia: 917-919.
15. Patch R, Kamenka D, Menzel J. Return voltage measurements Diagnostic interpretations based on the dielectric time constants, Materials Science-Poland 2009; 27 (4): 1157-1169.
16. Yang Y, Yang D, Hu Z, Zhang X. Oil Contamination Monitoring Based on Dielectric Constant Measurement 2009, International Conference on Measuring Technology and Mechatronics Automation(ICMTMA), Zhangjiajie, Hunan, China: 245-252.
17. Zhmud B. TBN, What`s in It for me?, Lube Magazine online 2018; 4.
18. Zhu J, Yoon J M, He D W, Qu Y, Bechhoefer E. Lubrication oil condition monitoring and remaining useful life prediction with particle filtering, International Journal of Prognostics and Health Management 2013; 4 Special issue 2: 1-15.
19. kittiwake.com/online-sensors
20. pmlubricants.com.au.
21. specord-service.de

Dr inż. Andrzej Urbaś

Katedra Podstaw Budowy Maszyn
Akademia Techniczno-Humanistyczna w Bielsku-Białej
Willowa 2, 43-309 Bielsko-Biała
E-mail: aurbas@ath.bielsko.pl

Dr hab. inż. Marek Szczotka, Prof. ATH

Katedra Transportu
Akademia Techniczno-Humanistyczna w Bielsku-Białej
Willowa 2, 43-309 Bielsko-Biała
E-mail: mszczotka@ath.bielsko.pl

The influence of the friction phenomenon on a forest crane operator's level of discomfort

Wpływ zjawiska tarcia na poziom dyskomfortu pracy operatora żurawia leśnego

Keywords: *crane dynamics, friction, seat vibrations*

Abstract: A mathematical model of a forest crane that is suitable for dynamics analysis of its operation cycle is presented in this paper. The flexibility of the operator's seat, drives and supports is taken into account. Joint coordinates are applied to describe the motion of the links together with the homogeneous transformations technique. Lagrange equations of the second order are used when deriving the equations of motions. Joint forces and torques are determined based on recursive Newton-Euler algorithms. These joint forces are then used in the LuGre friction model, which allows to calculate the friction coefficients and friction forces. Numerical analyses performed here show the influence of various friction forces on the vibration level as perceived by the operator of the crane. The level of discomfort is discussed based on standards commonly used in the vehicle and transportation industry for evaluations of vibration comfort.

Słowa kluczowe: *dynamika żurawia, tarcie, drgania siedzenia*

Streszczenie: W niniejszym artykule przedstawiono model matematyczny żurawia leśnego, który jest stosowany do analizy dynamiki cyklu jego pracy. Uwzględniono podatność podparcia fotela operatora, napędów oraz podpór. Do opisu ruchu członów stosuje się współrzędne złączowe i macierze przekształceń jednorodnych. Do wyprowadzenia równań ruchu modelu żurawia zastosowano podejście bazujące na formalizmie równań Lagrange'a drugiego rodzaju. Siły i momenty węzłowe są określane na podstawie rekurencyjnego algorytmu Newtona-Eulera. Siły te są następnie wykorzystywane w modelu tarcia LuGre, który pozwala obliczyć współczynniki i siły tarcia. Przeprowadzone analizy numeryczne pokazują wpływ różnych sił tarcia na poziom drgań odczuwany przez operatora żurawia. Poziom dyskomfortu operatora wywołany przez drgania maszyny został oszacowany w oparciu o standardy często stosowane w przemyśle samochodowym i transportowym.

Nomenclature

b, c, s, d, j	–	indeks podatnie posadowionej bazy, żurawia, siedzenia, napędu, połączenia
$(c,l)\|_{l=1,\dots,7}$	–	indeks członu
g	–	przyspieszenie ziemskie
$l^{(c,l)}$	–	długość członu
$m^{(b)}, m^{(c,l)}, m^{(s)}$	–	masa członu
n_b	–	liczba członów
n_{dof}	–	liczba współrzędnych uogólnionych opisujących ruchu układu
$\tilde{n}_{dof}^{(c,l)}$	–	liczba współrzędnych uogólnionych opisujących ruchu członu (c,l) względem członu $(c,l-1)$
$n_{dof}^{(b)}, n_{dof}^{(c,l)}, n_{dof}^{(s)}$	–	liczba współrzędnych uogólnionych opisujących ruchu członu względem układu odniesienia $n_{dof}^{(c,l)} = n_{dof}^{(b)} + n_{dof}^{(c,l-1)} + \tilde{n}_{dof}^{(c,l)}, n_{dof}^{(c,l)} = 0$
$s_\beta^{(sup_\alpha,i)}, d_\beta^{(sup_\alpha,i)}\Big\|_{\substack{\alpha\in\{b,s\}\\ \beta\in\{x,y,z\}}}$	–	współczynniki sztywności i tłumienia posadowienia
$s^{(d,i)}, d^{(d,i)}$	–	współczynniki sztywności i tłumienia napędu
$t_f^{(j,i)}, f_f^{(j,i)}$	–	moment tarcia w połączeniu obrotowym, siła tarcia w połączeniu przesuwnym
$sup_\alpha\|_{\alpha\in\{b,s\}}$	–	indeks podpory
$t^{(d,i)}, f^{(d,i)}$	–	momenty napędowe, siła napędowa
$\mathbf{r}_A^{(a)} = \begin{bmatrix} x_A^{(a)} & y_A^{(a)} & z_A^{(a)} & 1 \end{bmatrix}^T$	–	wektor pozycji punktu A zdefiniowany w lokalnym układzie członu a
$\mathbf{H}^{(b)}, \mathbf{H}^{(c,l)}, \mathbf{H}^{(s)}$	–	macierze pseudo-bezwładnościowe
$\tilde{\mathbf{T}}^{(c,l)}$		macierz przekształceń jednorodnych z lokalnego układu współrzędnych członu (c,l) do układu członu $(c,l-1)$
$\mathbf{T}^{(b)}, \mathbf{T}^{(c,l)}, \mathbf{T}^{(s)}$	–	macierze przekształceń jednorodnych z lokalnych układów do układu odniesienia $\mathbf{T}^{(c,l)} = \mathbf{T}^{(b)}\mathbf{T}^{(c,l-1)}\tilde{\mathbf{T}}^{(c,l)}, \mathbf{T}^{(c,0)} = \mathbf{I}$ $\mathbf{T}_i = \dfrac{\partial \mathbf{T}}{\partial q_i}, \mathbf{T}_{i,j} = \dfrac{\partial^2 \mathbf{T}}{\partial q_i \partial q_j}$

Parametry tarcia

$\boldsymbol{\sigma}_0^{(j)} = \left(\sigma_{0,i}^{(j)}\right)_{i=1,\dots,7}, \boldsymbol{\sigma}_1^{(j)} = \left(\sigma_{1,i}^{(j)}\right)_{i=1,\dots,7}, \boldsymbol{\sigma}_2^{(j)} = \left(\sigma_{2,i}^{(j)}\right)_{i=1,\dots,7}$	–	wektory współczynników sztywności, tłumienia oraz tłumienia wiskotycznego włoska modelu szczotkowego
$\mathbf{z}^{(j)} = \left(z_i^{(j)}\right)_{i=1,\dots,7}$	–	wektor odkształceń włoska modelu szczotkowego
$\boldsymbol{\mu}^{(j)} = \left(\mu_i^{(j)}\right)_{i=1,\dots,7}$	–	wektor współczynników tarcia
$\boldsymbol{\mu}_s^{(j)} = \left(\mu_{s,i}^{(j)}\right)_{i=1,\dots,7}, \boldsymbol{\mu}_k^{(j)} = \left(\mu_{k,i}^{(j)}\right)_{i=1,\dots,7}$	–	wektory współczynników tarcia statycznego i kinetycznego
$\dot{\mathbf{q}}_S^{(c)} = \left(\dot{q}_{S,i}^{(c)}\right)_{i=1,\dots,7}$	–	wektor prędkości Stribecka

Parametry stosowane do oceny komfortu (BS 6841, 1987, ISO 2631-1, 1997)

$y(t)$	Przebieg czasowy sygnału (dyskretny, obliczany przez symulację numeryczną). Sygnał $y(t)$ powinien być filtrowany przy użyciu filtrów częstotliwości.
$RMS = \sqrt{\frac{1}{T}\int_0^T y^2(t)\mathrm{d}t}$	Root-Mean-Square opisuje energetyczną zawartość sygnału drgań, T jest czasem drgań.
$C_f = \frac{\max(\|y(t)\|)}{RMS}$	Współczynnik szczytu, do stosowania w obecności wstrząsów (krótkotrwałe zdarzenia przejściowe o dużej wartości).
$RMQ = \sqrt[4]{\frac{1}{T}\int_0^T y^4(t)\mathrm{d}t}$	Quad-Mean-Square, podobnie jak RMS, ale lepiej opisuje wpływ dyskomfortu drgań, gdy $C_f > 9$.

$$K = \frac{1}{n\sigma^4}\sum_{i=1}^{n}(y_i - \tilde{y})^4$$

Kurtoza, stosowana do wysoce impulsywnych sygnałów w domenie czasu, gdzie n jest liczbą dyskretnych danych, σ jest odchyleniem standardowym, $\tilde{y}$ jest średnią wartością analizowanego sygnału.

$$VDV = \sqrt[4]{\int_0^T y^4(t)\mathrm{d}t}$$

Wartość dawki drgań, daje pomiar skumulowanego poziomu drgań otrzymywanych w danym okresie czasu (często 8 godzin lub 16 godzin). Należy stosować filtry zależne od lokalizacji i kierunku analizowanych drgań.

$$T_{15} = \left(\frac{15}{VDV}\right)^4 t$$

Czas trwania w sekundach, wymagany do osiągnięcia wartości $VDV = 15\mathrm{ms}^{-1.75}$, który jest zdefiniowany jako poważny dyskomfort. Parametr t jest czasem trwania mierzonego lub obliczanego sygnału.

1. Wprowadzenia

Problemy ze sterowaniem żurawiem stają się coraz ważniejsze dla projektantów i operatorów. Zwiększenie efektywności oraz bezpieczeństwa prac przeładunkowych, jak również poziomu komfortu dla całego zaangażowanego personelu oraz operatorów żurawi są głównymi czynnikami napędzającymi rozwój w tej dziedzinie. W tym celu nowoczesne maszyny są wyposażone w dość zaawansowane i drogie czujniki oraz inne urządzenia sterujące. Symulacje komputerowe we wczesnym etapie projektowania są bardzo przydatnym podejściem, które ma na celu zmniejszenie ogólnych kosztów i wyeliminowanie błędów konstrukcyjnych, które można przewidzieć bez budowania rzeczywistych prototypów żurawia. Co więcej, można łatwo symulować złożone zachowanie żurawia bez ryzyka uszkodzenia lub obrażeń. W tym kontekście opracowano wiele modeli żurawi o różnym stopniu złożoności i szczegółowości.

Model matematyczny żurawia chwytakowego przedstawiono w pracach [12, 13, 18, 22-24, 19, 20, 26-30]. Model ten został zastosowany do badania dynamiki żurawia podczas różnych cykli pracy i scenariuszy obsługi, w tym szacowania obciążeń, ruchu ładunku, sterowania układem napędowym i innych. W modelu matematycznym uwzględniono podatność układu podporowego [22-24, 26-30], który jest modelowany w postaci jednokierunkowych elementów sprężysto-tłumiących. W podobny sposób zbudowany jest układ reprezentujący połączenie fotela operatora z kolumną żurawia. W opracowanym w niniejszym artykule modelu żurawia wszystkie układy napędowe modelowane są również w postaci podatnej [26-30]. Do opisu geometrii członów żurawia stosuje się macierze przekształceń jednorodnych oraz współrzędne złączowe [7, 14, 15]. Równania ruchu wyprowadzono na podstawie metod i algorytmów przedstawionych w [11, 32]. Do całkowania równań ruchu stosuje się metodę Runge-Kutty IV rzędu ze stałym krokiem czasu. W celu określenia sił i momentów w połączeniach, które są konieczne do obliczenia sił i momentów tarcia, w każdym kroku całkowania równań ruchu rozwiązuje się odwrotne zadanie dynamiki Newtona-Eulera [5]. Współczynniki tarcia dla każdej pary kinematycznej oblicza się stosując model tarcia LuGre [1, 2, 17, 21], który uwzględnia zjawisko przesunięcia wstępnego [4], a także efekt Stribeck'a [25].

Wpływ tarcia na dynamikę żurawi leśnych został omówiony w poprzednich artykułach. Model tarcia Dahla zbadano w [27], natomiast prace [29, 30] koncentrowały się na modelu tarcia LuGre.

Przeprowadzone w niniejszym artykule analizy koncentrowały się na dynamicznych właściwościach żurawia ze szczególnym uwzględnieniem właściwości fotela operatora podczas wybranych cykli pracy. Zastosowano model tarcia LuGre, z

dwoma różnymi poziomami tarcia w połączeniach. Przewidywano poziom dyskomfortu, określony przez drgania przekazywane z kolumny na siedzenie i operatora, z uwzględnieniem podejścia bazującego na normach [6, 10] stosowanych w analizach N.V.H. (hałas, drgania i wpływ na użytkownika). Analiza poziomu komfortu w różnych układach, w tym we wszystkich rodzajach pojazdów, budynkach i innych konstrukcjach, jest bardzo ważna i wymagana przez urzędy certyfikujące [9, 16]. Przeprowadzono wiele testów w celu określenia, jak ludzkie ciało odczuwa dyskomfort [8]. Powszechną praktyką jest symulowanie i testowanie parametrów komfortu również w maszynach specjalnych lub budowlanych, jak na przykład w [3], ale w szczególności w wielu gałęziach przemysłu pojazdów lądowych.

2. Model matematyczny żurawia leśnego

Na rys. 1. przedstawiono model żurawia leśnego, który składa się z ośmiu sztywnych członów. Człony te są napędzane przez podatne napędy, generujące momenty $\mathbf{t}^{(d,i)}\big|_{i=1,2,3,7}$ oraz siłę $\mathbf{f}^{(d,4)}$. Cały żuraw (jego platforma) jest podparty na ośmiu podatnych podporach. Połączenie między siedziskiem i kolumną żurawia modelowane jest za pomocą bezmasowych elementów sprężysto-tłumiących.

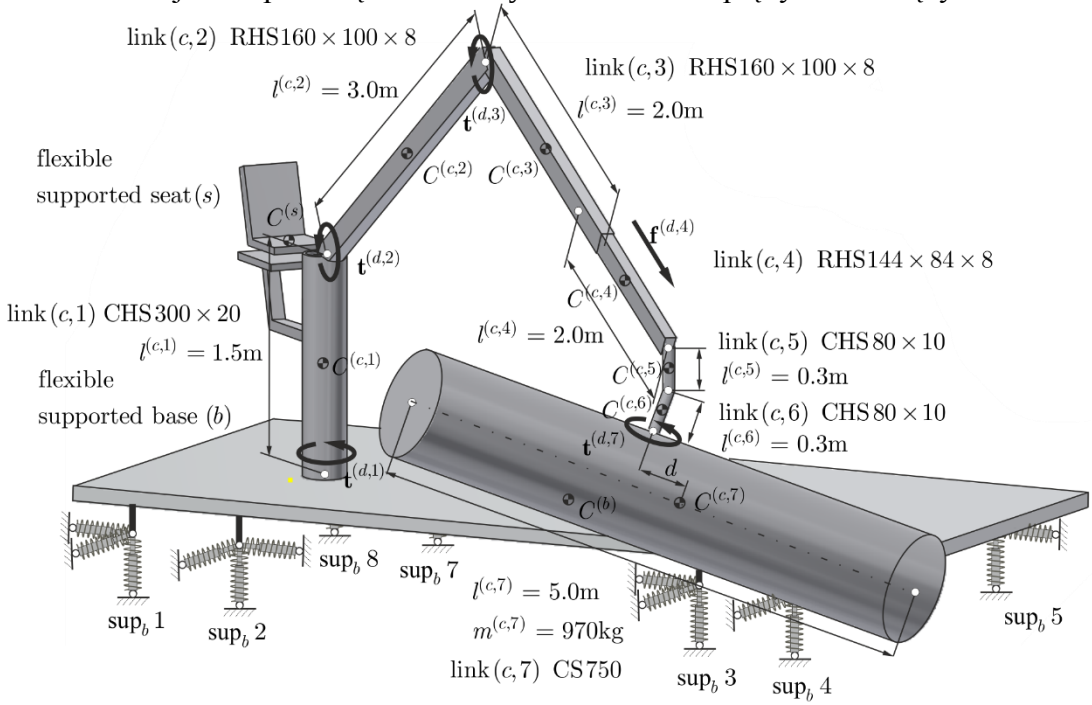


Fig. 1. Model of the forest crane
Rys.1. Model żurawia leśnego

Współrzędne złączowe oraz macierze przekształceń jednorodnych są stosowane do opisu geometrii żurawia leśnego. Na rys. 2 przedstawiono lokalne układy współrzędnych i numerację niektórych elementów żurawia.

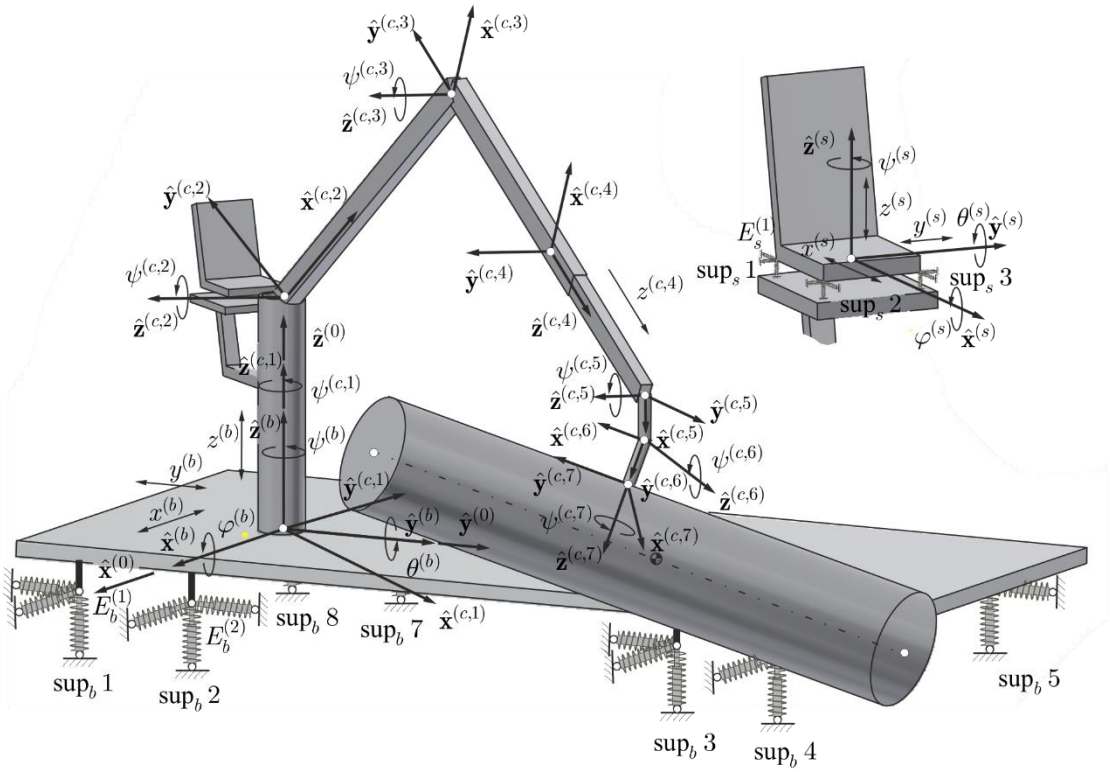


Fig. 2. Coordinate systems and notation applied to crane links
Rys. 2. Układy współrzędne i notacja stosowana do członów żurawia

Wektor współrzędnych uogólnionych ma następującą formę:

$$\mathbf{q} = (q_k)_{k=1,\ldots,n_{dof}} = \left[\left(\mathbf{q}^{(b)}\right)^T \quad \left(\mathbf{q}^{(c)}\right)^T \quad \left(\mathbf{q}^{(s)}\right)^T\right]^T, \tag{1}$$

gdzie:

$$\mathbf{q}^{(b)} = \left(q_j^{(b)}\right)_{j=1,\ldots,6} = [x^{(b)} \quad y^{(b)} \quad z^{(b)} \quad \psi^{(b)} \quad \theta^{(b)} \quad \varphi^{(b)}]^T,$$

$$\mathbf{q}^{(c)} = \left(q_j^{(c)}\right)_{j=1,\ldots,7} = [\psi^{(c,1)} \quad \psi^{(c,2)} \quad \psi^{(c,3)} \quad z^{(c,4)} \quad \psi^{(c,5)} \quad \psi^{(c,6)} \quad \psi^{(c,7)}]^T,$$

$$\mathbf{q}^{(s)} = \left(q_j^{(s)}\right)_{j=1,\ldots,6} = [x^{(s)} \quad y^{(s)} \quad z^{(s)} \quad \psi^{(s)} \quad \theta^{(s)} \quad \varphi^{(s)}]^T.$$

Macierze przekształceń jednorodnych są zdefiniowane następująco:

$$\mathbf{T}^{(\alpha)}\Big|_{\alpha\in\{b,s\}} = \begin{bmatrix} \mathrm{c}\psi^{(\alpha)}\mathrm{c}\theta^{(\alpha)} & \mathrm{c}\psi^{(\alpha)}\mathrm{s}\theta^{(\alpha)}\mathrm{s}\varphi^{(\alpha)} - \mathrm{s}\psi^{(\alpha)}\mathrm{c}\varphi^{(\alpha)} & \mathrm{c}\psi^{(\alpha)}\mathrm{s}\theta^{(\alpha)}\mathrm{c}\varphi^{(\alpha)} + \mathrm{s}\psi^{(\alpha)}\mathrm{s}\varphi^{(\alpha)} & x^{(\alpha)} \\ \mathrm{s}\psi^{(\alpha)}\mathrm{c}\theta^{(\alpha)} & \mathrm{s}\psi^{(\alpha)}\mathrm{s}\theta^{(\alpha)}\mathrm{s}\varphi^{(\alpha)} + \mathrm{c}\psi^{(\alpha)}\mathrm{c}\varphi^{(\alpha)} & \mathrm{s}\psi^{(\alpha)}\mathrm{s}\theta^{(\alpha)}\mathrm{c}\varphi^{(\alpha)} - \mathrm{c}\psi^{(\alpha)}\mathrm{s}\varphi^{(\alpha)} & y^{(\alpha)} \\ -\mathrm{s}\theta^{(\alpha)} & \mathrm{c}\theta^{(\alpha)}\mathrm{s}\varphi^{(\alpha)} & \mathrm{c}\theta^{(\alpha)}\mathrm{c}\varphi^{(\alpha)} & z^{(\alpha)} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\widetilde{\mathbf{T}}^{(c,1)} = \begin{bmatrix} \mathrm{c}\psi^{(c,1)} & -\mathrm{s}\psi^{(c,1)} & 0 & 0 \\ \mathrm{s}\psi^{(c,1)} & \mathrm{c}\psi^{(c,1)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \widetilde{\mathbf{T}}^{(c,2)} = \begin{bmatrix} \mathrm{c}\psi^{(c,2)} & -\mathrm{s}\psi^{(c,2)} & 0 & 0 \\ 0 & 0 & -1 & 0 \\ \mathrm{s}\psi^{(c,2)} & \mathrm{c}\psi^{(c,2)} & 0 & l^{(c,1)} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\widetilde{\mathbf{T}}^{(c,3)} = \begin{bmatrix} \mathrm{c}\psi^{(c,3)} & -\mathrm{s}\psi^{(c,3)} & 0 & l^{(c,2)} \\ \mathrm{s}\psi^{(c,3)} & \mathrm{c}\psi^{(c,3)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \widetilde{\mathbf{T}}^{(c,4)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -z^{(c,4)} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\widetilde{\mathbf{T}}^{(c,5)} = \begin{bmatrix} \mathrm{c}\psi^{(c,5)} & -\mathrm{s}\psi^{(c,5)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\mathrm{s}\psi^{(c,5)} & -\mathrm{c}\psi^{(c,5)} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \widetilde{\mathbf{T}}^{(c,6)} = \begin{bmatrix} \mathrm{c}\psi^{(c,6)} & -\mathrm{s}\psi^{(c,6)} & 0 & l^{(c,5)} \\ 0 & 0 & 1 & 0 \\ -\mathrm{s}\psi^{(c,6)} & -\mathrm{c}\psi^{(c,6)} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\widetilde{\mathbf{T}}^{(c,7)} = \begin{bmatrix} \mathrm{c}\psi^{(c,7)} & -\mathrm{s}\psi^{(c,7)} & 0 & 0 \\ 0 & 0 & 1 & l^{(c,6)} \\ -\mathrm{s}\psi^{(c,7)} & -\mathrm{c}\psi^{(c,7)} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathrm{s}\alpha^{(\beta)} = \sin\alpha^{(\beta)}, \mathrm{c}\alpha^{(\beta)} = \cos\alpha^{(\beta)}.$$

Współczynniki tarcia wyznaczane są w oparciu o model tarcia LuGre. Równania ruchu są wyprowadzane stosując równania Lagrange'a II rodzaju. Ogólną ich postać można zapisać następująco:

$$\dot{\mathbf{z}}^{(j)} = \mathbf{LuGre}\left(t, \mathbf{q}^{(c)}, \mathbf{z}^{(j)}\right), \tag{2.1}$$

$$\mathbf{M}\ddot{\mathbf{q}} = \mathbf{e}(t, \mathbf{q}, \dot{\mathbf{q}}) + \mathbf{s}(\mathbf{q}, \dot{\mathbf{q}}) + \mathbf{d}\left(t, \mathbf{q}^{(c)}, \dot{\mathbf{q}}^{(c)}\right) - \mathbf{f}\left(t, \mathbf{q}^{(c)}, \dot{\mathbf{q}}^{(c)}\right), \tag{2.2}$$

gdzie

$$(LuGre_i)_{i=1,\ldots,7} = \dot{q}_i^{(c)} \left(1 - \frac{\sigma_{0,i}^{(j)} z_i^{(j)} \mathrm{sgn}\left(\dot{q}_i^{(c)}\right)}{\mu_{k,i}^{(j)} + \left(\mu_{s,i}^{(j)} - \mu_{k,i}^{(j)}\right) \exp\left(-\left(\frac{\dot{q}_i^{(c)}}{\dot{q}_{S,i}^{(c)}} \right)^2 \right)} \right),$$

$$\boldsymbol{\mu}^{(j)} = \boldsymbol{\sigma}_0^{(j)} \mathbf{z}^{(j)} + \boldsymbol{\sigma}_1^{(j)} \dot{\mathbf{z}}^{(j)} + \boldsymbol{\sigma}_2^{(j)} \dot{\mathbf{q}}^{(c)},$$

$$\mathbf{M} = \sum_{\alpha \in \{b,c,s\}} \mathbf{M}^{(\alpha)},$$

$$\mathbf{M}^{(b)} = \begin{bmatrix} \mathbf{M}_{1,1}^{(b)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \mathbf{M}^{(c)} = \begin{bmatrix} \mathbf{M}_{1,1}^{(c)} & \cdots & \mathbf{M}_{1,j}^{(c)} & \cdots & \mathbf{M}_{1,n_b-1}^{(c)} & \mathbf{0} \\ \vdots & \ddots & \vdots & \iddots & \vdots & \vdots \\ \mathbf{M}_{i,1}^{(c)} & \cdots & \mathbf{M}_{i,j}^{(c)} & \cdots & \mathbf{M}_{i,n_b-1}^{(c)} & \mathbf{0} \\ \vdots & \iddots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{M}_{n_{b-1},1}^{(c)} & \cdots & \mathbf{M}_{n_{b-1},j}^{(c)} & \cdots & \mathbf{M}_{n_{b-1},n_{b-1}}^{(c)} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$$\mathbf{M}^{(s)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{1,1}^{(s)} \end{bmatrix},$$

$$\mathbf{M}_{1,1}^{(\alpha)}\Big|_{\alpha \in \{b,s\}} = \left(m_{i,j}^{(\alpha)}\right)_{i,j=1,\ldots,n_{dof}^{(\alpha)}}, m_{i,j}^{(\alpha)} = \mathrm{tr}\left\{ \mathbf{T}_i^{(\alpha)} \mathbf{H}^{(\alpha)} \left(\mathbf{T}_j^{(\alpha)} \right)^T \right\},$$

$$\mathbf{M}_{i,j}^{(c)} = \sum_{l=\max\{i,j\}}^{n_b-1} \mathbf{M}_{i,j}^{(c,l)}, m_{i,j}^{(c,l)} = \mathrm{tr}\left\{ \mathbf{T}_i^{(c,l)} \mathbf{H}^{(c,l)} \left(\mathbf{T}_j^{(c,l)} \right)^T \right\},$$

$$\mathbf{e} = \sum_{\alpha \in \{b,c,s\}} \mathbf{e}^{(\alpha)},$$

$$\mathbf{e}^{(b)} = \begin{bmatrix} \mathbf{e}_1^{(b)} \\ \mathbf{0} \end{bmatrix}, \mathbf{e}^{(c)} = \begin{bmatrix} \mathbf{e}_1^{(c)} \\ \vdots \\ \mathbf{e}_i^{(c)} \\ \vdots \\ \mathbf{e}_{n_{b-1}}^{(c)} \\ \mathbf{0} \end{bmatrix}, \mathbf{e}^{(s)} = \begin{bmatrix} \mathbf{0} \\ \mathbf{e}_1^{(s)} \end{bmatrix},$$

$$\mathbf{e}_1^{(\alpha)}\Big|_{\alpha \in \{b,s\}} = -\left(\mathbf{h}^{(\alpha)} + \mathbf{g}^{(\alpha)} \right),$$

$$\mathbf{h}^{(\alpha)} = \left(h_i^{(\alpha)}\right)_{i=1,\ldots,n_{dof}^{(\alpha)}}, h_i^{(\alpha)} = \sum_{m=1}^{n_{dof}^{(\alpha)}} \sum_{n=1}^{n_{dof}^{(\alpha)}} \mathrm{tr}\left\{ \mathbf{T}_i^{(\alpha)} \mathbf{H}^{(\alpha)} \left(\mathbf{T}_{m,n}^{(\alpha)} \right)^T \right\} \dot{q}_m^{(\alpha)} \dot{q}_n^{(\alpha)},$$

$$\mathbf{g}^{(\alpha)} = \left(g_i^{(\alpha)}\right)_{i=1,\ldots,n_{dof}^{(\alpha)}}, g_i^{(\alpha)} = m^{(\alpha)} g \mathbf{j}_3 \mathbf{T}_i^{(\alpha)} \mathbf{r}_{C^{(\alpha)}}^{(\alpha)},$$

$$\mathbf{e}_i^{(c)} = -\sum_{l=i}^{n_b-1} \left(\mathbf{h}_i^{(c,l)} + \mathbf{g}_i^{(c,l)} \right),$$

$$\mathbf{h}_i^{(c,l)} = \left(h^{(c,l)}_{n_{dof}^{(c,i-1)}+k} \right)_{k=1,\ldots,n_{dof}^{(c,i)}},$$

$$h_i^{(c,l)} = \sum_{m=1}^{n_{dof}^{(c,l)}} \sum_{n=1}^{n_{dof}^{(c,l)}} \mathrm{tr}\left\{ \mathbf{T}_i^{(c,l)} \mathbf{H}^{(c,l)} \left(\mathbf{T}_{m,n}^{(c,l)} \right)^T \right\} \dot{q}_m^{(c,l)} \, \dot{q}_n^{(c,l)},$$

$$\mathbf{g}_i^{(c,l)} = \left(g^{(c,l)}_{n_{dof}^{(c,i-1)}+k} \right)_{k=1,\ldots,n_{dof}^{(c,i)}}, \; g_i^{(c,l)} = m^{(c,l)} g \mathbf{j}_3 \mathbf{T}_i^{(c,l)} \mathbf{r}_{C^{(c,l)}}^{(c,l)},$$

$$\mathbf{s} = \sum_{\alpha \in \{b,s\}} \mathbf{s}^{(sup_\alpha)}, \; \mathbf{s}^{(sup_\alpha)} = \left(s_k^{(sup_\alpha)} \right)_{k=1,\ldots,n_{nof}},$$

$$s_k^{(sup_b)} = \begin{cases} -\sum_{i=1}^{n_{sup_b}} \left(\left(\frac{\partial \mathbf{e}^{(sup_b,i)}}{\partial q_k} \right)^T \mathbf{S}^{(sup_b,i)} \mathbf{e}^{(sup_b,i)} + \left(\frac{\partial \dot{\mathbf{e}}^{(sup_b,i)}}{\partial \dot{q}_k} \right)^T \mathbf{D}^{(sup_b,i)} \dot{\mathbf{e}}^{(sup_b,i)} \right) & \text{if } q_k \in \{\mathbf{q}^{(b)}\} \\ 0 & \text{otherwise} \end{cases}$$

$$s_k^{(sup_s)} = \begin{cases} \sum_{i=1}^{n_{sup_s}} \left(\left(\frac{\partial \mathbf{e}^{(sup_s,i)}}{\partial q_k} \right)^T \mathbf{S}^{(sup_s,i)} \mathbf{e}^{(sup_s,i)} + \left(\frac{\partial \dot{\mathbf{e}}^{(sup_s,i)}}{\partial \dot{q}_k} \right)^T \mathbf{D}^{(sup_s,i)} \dot{\mathbf{e}}^{(sup_s,i)} \right) & \text{if } q_k \in \{\mathbf{q}^{(b)}, \psi^{(c,1)}\} \\ -\sum_{i=1}^{n_{sup_s}} \left(\left(\frac{\partial \mathbf{e}^{(sup_s,i)}}{\partial q_k} \right)^T \mathbf{S}^{(sup_s,i)} \mathbf{e}^{(sup_s,i)} + \left(\frac{\partial \dot{\mathbf{e}}^{(sup_s,i)}}{\partial \dot{q}_k} \right)^T \mathbf{D}^{(sup_s,i)} \dot{\mathbf{e}}^{(sup_s,i)} \right) & \text{if } q_k \in \{\mathbf{q}^{(s)}\} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{e}^{(sup_b,i)} = \mathbf{J} \mathbf{T}^{(b)} \mathbf{r}_{E_b^{(i)}}^{(b)}, \; \mathbf{e}^{(sup_s,i)} = \mathbf{J} \left(\mathbf{T}^{(s)} \mathbf{r}_{E_s^{(i)}}^{(s)} - \mathbf{T}^{(c,1)} \mathbf{r}_{E_s^{(i)}}^{(c,1)} \right),$$

$$\mathbf{J} = \begin{bmatrix} \mathbf{j}_1 \\ \mathbf{j}_2 \\ \mathbf{j}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{S}^{(sup_\alpha,i)}\big|_{\alpha \in \{b,s\}} = \mathrm{diag}\left\{ s_x^{(sup_\alpha,i)}, s_y^{(sup_\alpha,i)}, s_z^{(sup_\alpha,i)} \right\},$$

$$\mathbf{D}^{(sup_\alpha,i)}\big|_{\alpha \in \{b,s\}} = \mathrm{diag}\left\{ d_x^{(sup_\alpha,i)}, d_y^{(sup_\alpha,i)}, d_z^{(sup_\alpha,i)} \right\},$$

$$\mathbf{d}^{(c)} = \left(d_k^{(c)} \right)_{k=1,\ldots,n_{dof}} = \begin{bmatrix} \mathbf{0} & t_{dr}^{(c,1)} & t_{dr}^{(c,2)} & t_{dr}^{(c,3)} & f_{dr}^{(c,4)} & \mathbf{0} & t_{dr}^{(c,7)} & \mathbf{0} \end{bmatrix}^T,$$

$$t_{dr}^{(c,i)}\Big|_{i=1,2,3,7} = -\left(s_{dr}^{(c,i)} \left(\psi_{dr}^{(c,i)} - \psi^{(c,i)} \right) + d_{dr}^{(c,i)} \left(\dot{\psi}_{dr}^{(c,i)} - \dot{\psi}^{(c,i)} \right) \right),$$

$$f_{dr}^{(c,4)} = -\left(s_{dr}^{(c,4)} \left(z_{dr}^{(c,4)} - z^{(c,4)} \right) + d_{dr}^{(c,4)} \left(\dot{z}_{dr}^{(c,4)} - \dot{z}^{(c,4)} \right) \right),$$

$$\mathbf{f}^{(j)} = \left(f_k^{(j)} \right)_{k=1,\ldots,n_{dof}} =$$

$$\begin{bmatrix} \mathbf{0} & t_f^{(j,1)} & t_f^{(j,2)} & t_f^{(j,3)} & f_f^{(j,4)} & t_f^{(j,5)} & t_f^{(j,6)} & t_f^{(j,7)} & \mathbf{0} \end{bmatrix}^T.$$

Wartości momentów tarcia $t_f^{(j,i)}\big|_{i=(1\div7)\setminus\{4\}}$ w połączeniach obrotowych oraz siły tarcia $f_f^{(j,4)}$ w połączeniu przesuwnym [29] są obliczane na podstawie znajomości sił i momentów węzłowych, wyznaczanych za pomocą rekurencyjnego algorytmu Newtona-Eulera.

W oparciu o powyższe sformułowanie opracowano program komputerowy w środowisku Visual C ++. Zastosowano standardową metodę Runge-Kutta IV rzędu ze stałym krokiem czasu 10^{-4}s.

3.1. Scenariusze pracy żurawia oraz przypadki jego obciążenia

Na rys. 3 przedstawiono sekwencję ruchu żurawia. W czasie $t = 0$ s ładunek spoczywa na platformie. Po dwóch sekundach ładunek jest podnoszony poprzez wzrost wartości kąta wysięgnika. Następnie kolumna żurawia obraca się (osiągając 90° w czasie $t = 5$ s), a jednocześnie rozpoczyna się ruch teleskopu w czasie $t = 5$ s. Dla końcowego kąta obrotu kolumny, $\psi_{dr}^{(c,1)} = 180°$, teleskop zatrzymuje się osiągając minimalną długość w czasie $t = 9$s. Cykl kończy się, gdy ładunek jest ustawiony na platformie po przeciwnej stronie.

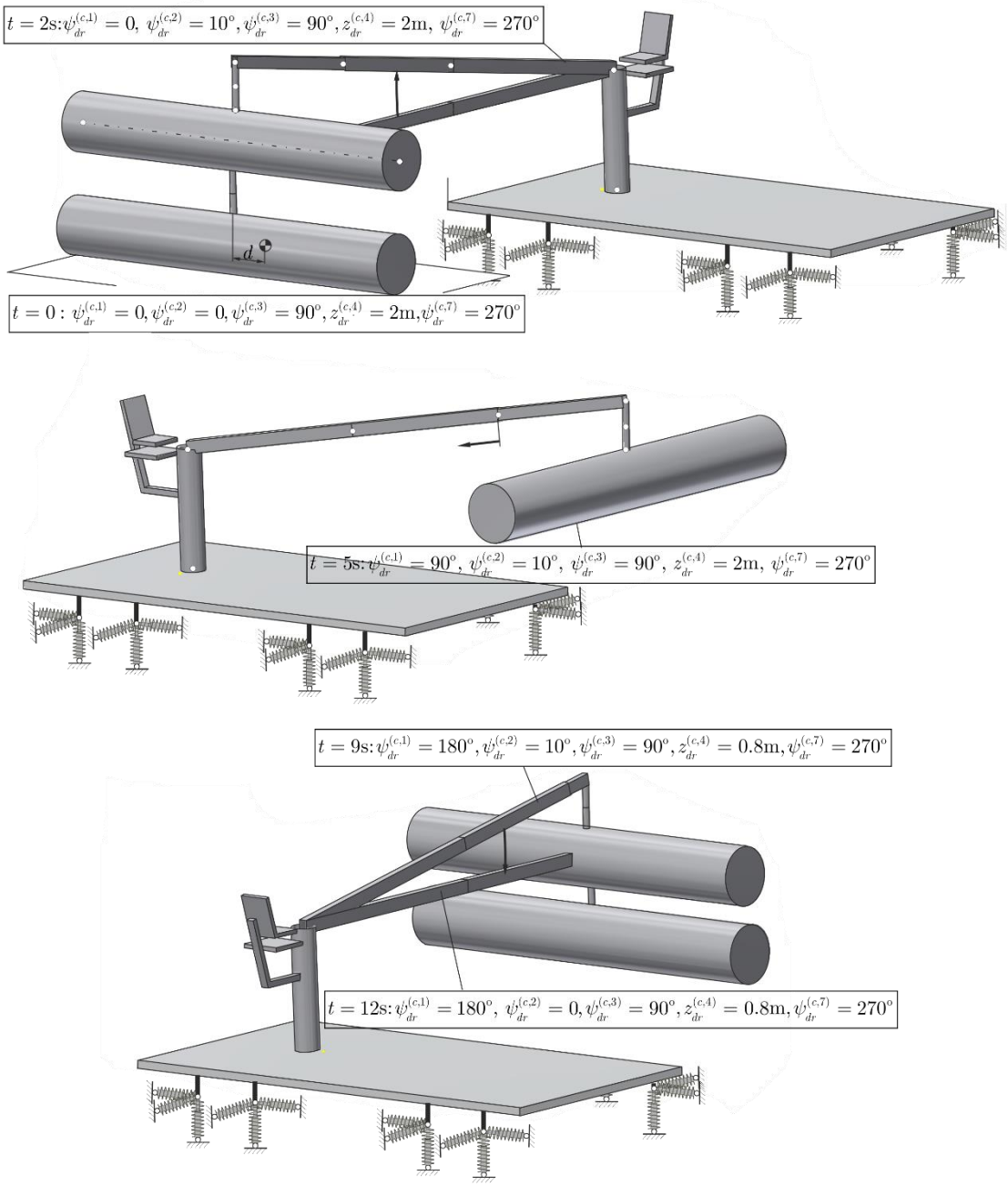


Fig. 3. Crane operation sequences
Rys. 3. Sekwencje ruchu żurawia

Warunki obciążenia żurawia:
1) pusty chwytak (E) - praca z rozładowanym żurawiem,
2) żurawia z obciążeniem (F) - praca z masą ładunku $m^{(c,7)}$.
Scenariusz w przypadku bez ładunku ma identyczne funkcje jazdy, z odwrotnymi rzędami / wartościami - wracając do pozycji odbioru tak samo jak w $t = 0$ s.
Jak pokazano na rys. 1, odległość d oznacza przesunięcie między środkiem ciężkości ładunku a osią połączenia i napędu $\mathbf{t}^{(d,7)}$. Wpływ tej odległości został uznany za jeden z ważnych parametrów w analizie. Każdy cykl roboczy, w stanie praktycznym, będzie charakteryzował się zmienną wartością d, spowodowaną nieidealnym rozkładem masy pni i niewspółosiowością długości, początkową pozycją ładunku na platformie magazynowej oraz z wielu innych powodów. Dla przeprowadzonych badań przedstawionych w niniejszej pracy przyjmuje się, że zakres $d = \pm 20\text{cm}$ jest typowy.
Przypadki obciążeń analizowane w tej pracy są wymienione na rys. 4. "Empty" oznacza przypadki ruchu dźwigu zdefiniowane dla tej samej sekwencji jak z obciążeniem - tylko bez obciążenia dołączonego do chwytaka. "Loaded" to cykle z masą pnia drzewa $m^{(c,7)} = 970\text{kg}$.

$$\begin{array}{l|l} x = 0\,(\text{no damping}) & L = F\ (\text{loaded crane}) \\ x = 1\,(\text{nominal damping}) & L = E\ (\text{empty hook}) \end{array}$$

$$Dx - Fy - dz - L$$

$$\begin{array}{l|l} y = 1\,(\text{friction coefs. "Set-1"}) & z = 0\ \ (\text{no load CoG offset}) \\ y = 2\,(\text{friction coefs."Set-2"}) & z = 10\,(10\text{cm offset of load CoG}) \\ & z = 20\,(20\text{cm offset of load CoG}) \end{array}$$

Fig. 4. Analyzed cases and symbol assignation
Rys. 4. Analizowane przypadki i ich symbole

Przypadki bez tłumienia w mocowaniu siedzenia są analizowane tylko w celu porównania wpływu tłumienia na poziom drgań operatora. Analiza wszystkich możliwych przypadków, prowadząca do bardziej ogólnej oceny konkretnego projektu żurawia, jest dużym zadaniem i wykracza poza rozmiar tej pracy.

3.2. Główne parametry układu

Wszystkie parametry masowe żurawia mają przypisane właściwości zgodnie z właściwościami geometrii (przekroje), jak pokazano na rys. 1. Masę siedzenia operatora przyjmuje się jako łączną masę operatora (jedna pojedyncza masa 80kg) i masę własną siedzenia równą 25kg; stąd zakłada się, że łączna masa siedzenie-operator wynosi $m^{(s)} = 105\text{kg}$.
Parametry geometryczne podpór platformy zawarte są w tab.1. Przedstawiono również założone współczynniki sztywności i tłumienia.

Table 1. Parameters of the crane supports
Tabela 1. Parametry podpór żurawia

	1	2	3	4	5	6	7	8
$x_{E_b^{(i)}}^{(b)}$[m]	1.5	1.5	1.5	1.5	–1.5	–1.5	–1.5	–1.5
$y_{E_b^{(i)}}^{(b)}$[m]	0	1.0	8.0	9.0	9.0	8.0	1.0	0
$z_{E_b^{(i)}}^{(b)}$[m]					–0.57			
$s_\alpha^{(sup_b,i)}\big\vert_{\alpha\in\{x,y\}}$[Nm^{-1}]					$3\cdot10^6$			
$s_z^{(sup_b,i)}$[Nm^{-1}]					$1\cdot10^7$			
$d_\alpha^{(sup_b,i)}\big\vert_{\alpha\in\{x,y\}}$[Nsm^{-1}]					$5\cdot10^4$			
$d_z^{(sup_b,i)}$[Nsm^{-1}]					$9\cdot10^4$			

Rozmieszczenie punktów mocowania oraz elementy sprężysto-tłumiące łączące kolumnę żurawia i siedzenie określono w tab. 2.

Table 2. Parameters of the seat supporting elements
Tabela 2. Parametry podparcia siedzenia $sup_s i$

	1	2	3	4
$x_{E_s^{(i)}}^{(s)}$ [m]	–0.25	0.25	0.25	–0.25
$y_{E_s^{(i)}}^{(s)}$ [m]	0.25	0.25	–0.25	–0.25
$z_{E_s^{(i)}}^{(s)}$ [m]			–0.05	
$x_{E_s^{(i)}}^{(c,1)}$ [m]	–0.85	–0.35	–0.35	–0.85
$y_{E_s^{(i)}}^{(c,1)}$ [m]	0.25	0.25	–0.25	–0.25
$z_{E_s^{(i)}}^{(c,1)}$ [m]			1.4	
$s_\alpha^{(sup_s,i)}\big\|_{\alpha\in\{x,y\}}$ [Nm^{-1}]			10^3	
$s_Z^{(sup_s,i)}$ [Nm^{-1}]			$9\cdot10^3$	
$d_\alpha^{(sup_s,i)}\big\|_{\alpha\in\{x,y\}}$ [Nsm^{-1}]			40	
$d_Z^{(sup_s,i)}$ [Nsm^{-1}]			120	

Parametry tarcia zawarto w tab. 3. Zdefiniowano dwa zestawy w celu rozróżnienia różnych warunków połączeń, tj. normalnego (**Set-1**) dobrze nasmarowanego i słabo nasmarowanego (**Set-2**).

Table 3. Friction parameters
Tabela 3. Parametry tarcia

		1	2	3	4	5	6	7
Set-1	$\mu_{k,i}^{(j)}$	0.15	0.10	0.10	0.10	0.15	0.15	0.15
	$\mu_{s,i}^{(j)}$	0.20	0.15	0.20	0.20	0.20	0.20	0.20
Set-2	$\mu_{k,i}^{(j)}$	0.07	0.07	0.05	0.20	0.20	0.20	0.10
	$\mu_{s,i}^{(j)}$	0.10	0.10	0.15	0.30	0.35	0.35	0.15
	$\dot{q}_{S,i}^{(c)}$ [rad s^{-1}, ms^{-1}]				0.005			
	$\sigma_{0,i}^{(j)}$ [Nmrad^{-1}, Nm^{-1}]		10^5		10^7		10^5	
	$\sigma_{1,i}^{(j)}$ [Nmsrad^{-1}, Nsm^{-1}]				5			
	$\sigma_{2,i}^{(j)}$ [Nmsrad^{-1}, Nsm^{-1}]				0			

3.3. Ocena drgań

Odczuwanie dyskomfortu nie jest jednakowe dla każdej osoby, tj. odczuwany komfort zależy od wielu czynników. Standardy często stosowane w przemyśle to np. BS 6841 i ISO 2631-1 - rys. 5. W niniejszej pracy wykorzystano te standardy do oceny poziomu drgań i dyskomfortu odczuwanego przez operatora. Podobne podejście można zastosować do dynamiki pojazdów i ich podukładów w celu oszacowania komfortu jazdy, który zastosowano również w procedurach optymalizacyjnych, dających pożądany minimalny dyskomfort [31].

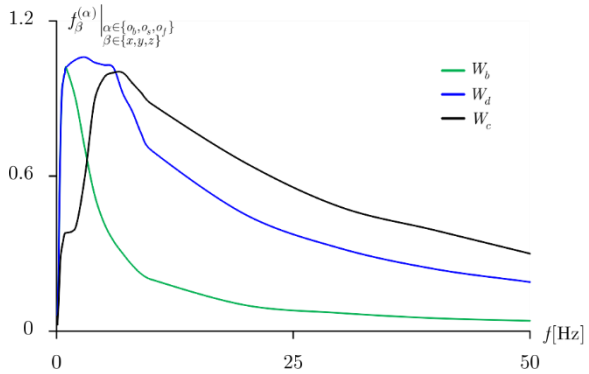


Fig. 5. Frequency-weighting filters to be applied in comfort assessment
Rys. 5. Filtry częstotliwościowe stosowane w ocenie komfortu

W przypadku ogólnym, człowiek narażony jest na drgania w kombinacji wszystkich 6 kierunków (translacji i obrotów), ale w przedstawionej pracy rozważane są tylko sygnały w osiach x, y, z. Operator żurawia odczuje drgania na plechach i na stopy (drgania na rękach i głowie mogą mieć również znaczenie). Na rys. 6 przedstawiono przypisane badane lokalizacje i odpowiednie filtry.

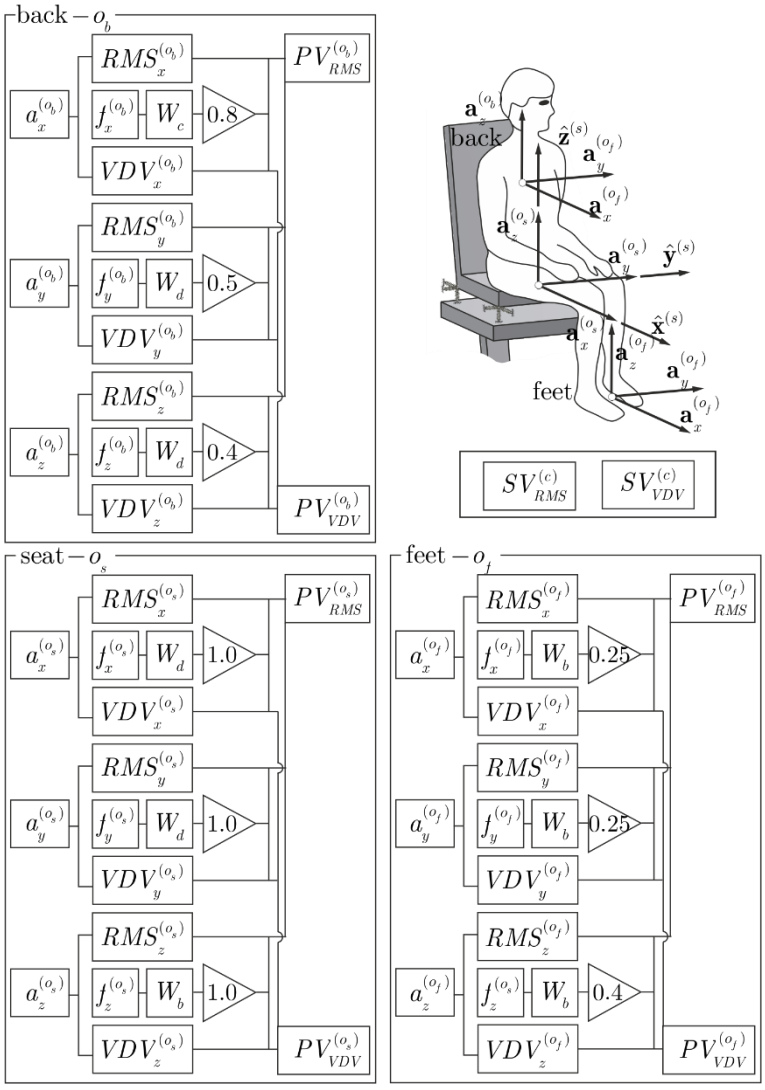


Fig. 6. Locations of interest and frequency-weighting filters
Rys. 6. Analizowane położenia oraz filtry częstotliwościowe

Każde rozważane miejsce oceny komfortu charakteryzuje się następującymi zależnościami:

$$PV_{RMS}^{(\alpha)}\Big|_{\alpha\in\{O_b,O_s,O_f\}} = \sqrt{\sum_{\beta\in\{x,y,z\}}\left(f_\beta^{(\alpha)}RMS_\beta^{(\alpha)}\right)^2}, \tag{3.1}$$

$$PV_{VDV}^{(\alpha)}\Big|_{\alpha\in\{O_b,O_s,O_f\}} = \sqrt[4]{\sum_{\beta\in\{x,y,z\}}\left(f_\beta^{(\alpha)}VDV_\beta^{(\alpha)}\right)^4}. \tag{3.2}$$

Pojedyncze wartości określające poziom komfortu dla całej konstrukcji są zdefiniowane jako suma wszystkich wartości lokalizacji. Proponowane podejście pozwoli podsumować komfort jako jedną wartość, którą można porównać między różnymi projektami lub różnymi operacjami.
Obowiązują następujące definicje:

$$SV_{RMS}^{(c)} = \sqrt{\sum_{\alpha\in\{O_b,O_s,O_f\}}\left(PV_{RMS}^{(\alpha)}\right)^2}, \tag{4.1}$$

,

$$SV_{VDV}^{(c)} = \sqrt{\sum_{\alpha\in\{O_b,O_s,O_f\}}\left(PV_{VDV}^{(\alpha)}\right)^2}. \tag{4.2}$$

Inne definicje (takie jak "running *RMS*" (BS 6841) i „peak-to-peak” (wartość międzyszczytowa) mogłyby również zostać zastosowane do oceny wpływu pracy żurawia na dyskomfort operatora [3]. Wybór najbardziej odpowiedniego parametru będzie wyborem projektowym i powinien być wykonany w zależności od rodzaju operacji wykonywanej przez żuraw. Można również opracować ogólne procedury optymalizacyjne, skutkujące minimalizacją dyskomfortu, zmieniając na przykład sposób mocowania lub położenie punktów podparcia.

4. Wyniki symulacji

4.1. Przykładowe przebiegi czasowe

Na rys. 7 przedstawiono przykładowe wyniki - wszystkie serie przedstawiają przebiegi czasowe i częstotliwościowe obliczone dla składowej pionowej przyspieszenia siedzenia (pokazano niefiltrowane wyniki). Badany jest efekt tłumienia w punktach mocowania siedzenia. Zazwyczaj przyspieszenie dla układu bez tłumienia byłoby zbyt zachowawcze, nawet w przypadku sztywnego układu siedzenia w niektórych prostych konstrukcjach, w których nie występują elementy tłumiące (występuje tylko tłumienie strukturalne).

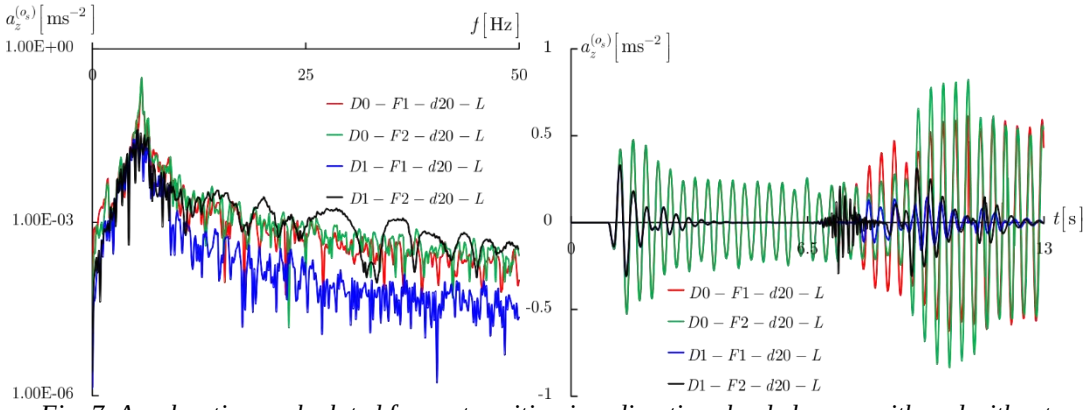


Fig. 7. Accelerations calculated for seat position in z direction; loaded crane with and without damping in seat mount elements. Frequency plots are on the left; time histories are on the right.

Rys. 7. Przyspieszenia siedzenia w kierunku osi z; żuraw obciążony z i bez tłumienia w posadowieniu siedzenia - charakterystyka częstotliwościowa (z lewej), czasowa (z prawej)

Na rys. 8 przedstawiono przyspieszenia obliczone dla parametrów tarcia **Set-2** i pracy z pustym chwytakiem (nieobciążony żuraw).
Podobny zestaw wyników dla podparcia nóg operatora przedstawiono na rys. 9. Efekt tarcia na ruch translacyjny wysięgnika (napęd uruchamiany podczas obrotu) jest wyraźnie widoczny w miejscu oparcia stóp operatora.

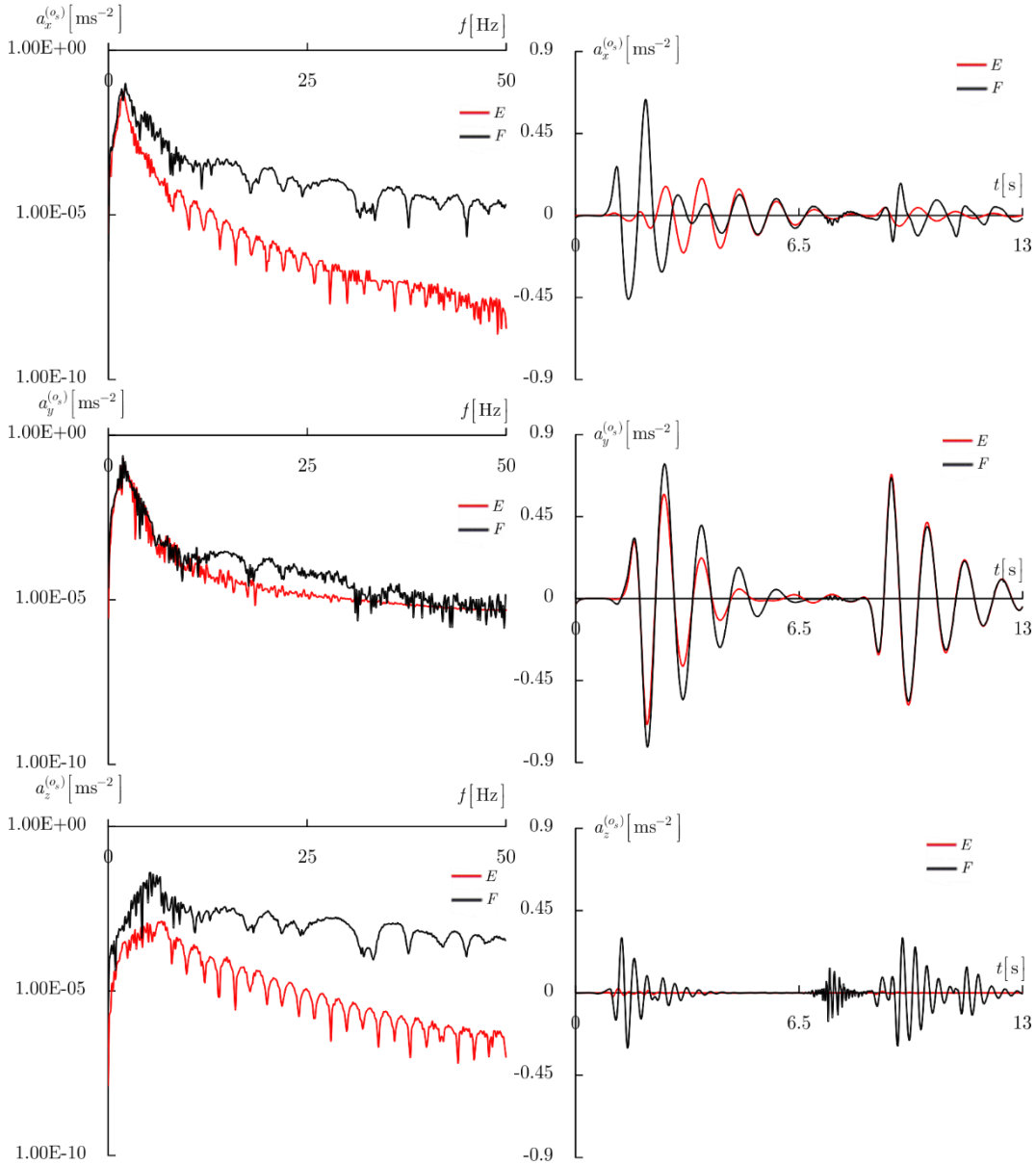


Fig. 8. Accelerations (filtered) calculated for seat position in x, y, z direction operation; with empty and loaded crane (load cases $D1-F2-d0-E$ and $D1-F2-d20-F$). Frequency (filtered) plots on the left; time histories on right.

Rys. 8. Przyspieszenia (filtrowane) siedzenia w kierunku osi x, y, z; żuraw bez ładunku i obciążony (przypadku obciążenia $D1-F2-d0-E$ and $D1-F2-d20-F$). Charakterystyka częstotliwościowa (filtrowana) (z lewej), czasowa (z prawej)

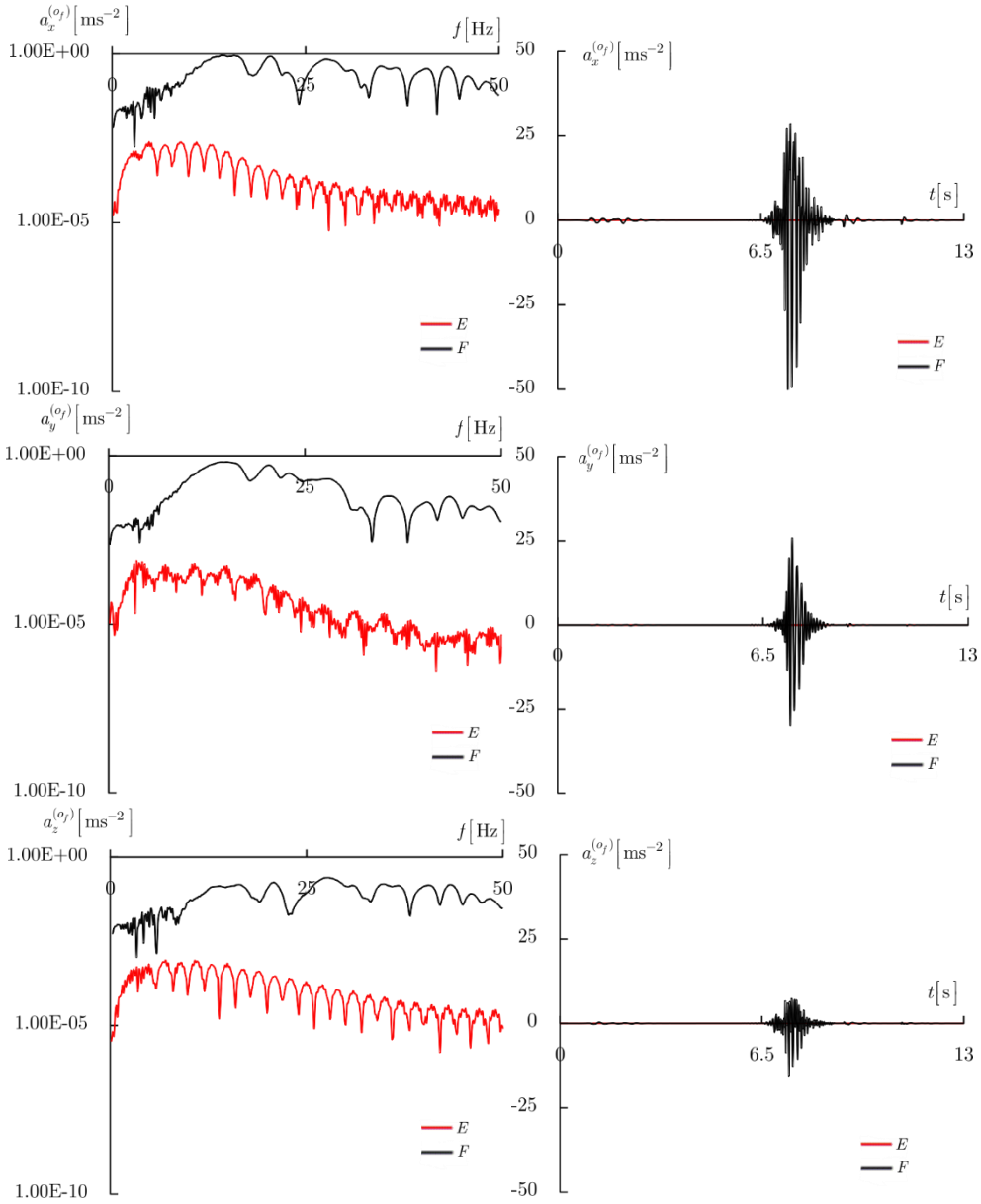


Fig. 9. Accelerations (filtered) calculated for feet rest position in x, y, z direction; operation with empty and loaded crane (load cases $D1 - F2 - d0 - E$ and $D1 - F2 - d20 - F$). Frequency (filtered) plots on the left; time histories on the right.

Rys. 9. Przyspieszenia (filtrowane) w położeniu stóp operatora w kierunku osi x, y, z; żuraw bez ładunku i obciążony (przypadki obciążenia $D1 - F2 - d0 - E$ i $D1 - F2 - d20 - F$). Charakterystyka częstotliwościowa (filtrowana) (z lewej), czasowa (z prawej)

Wpływ położenia środka ciężkości ładunku ($d = 0\text{cm},\ d = 10\text{cm}, d = 20\text{cm}$) przedstawiono na rys. 10. Wyniki pokazują znaczący wpływ środka ciężkości ładunku (i wywoływanych tym momentów) na dynamikę żurawia. Wysoka wartość siły tarcia powstaje w wysięgniku (podczas fazy ruchu teleskopu) i impulsy są wyraźnie widoczne. Większość z nich jest przenoszona na podłogę i podstawę siedzenia. Zawieszenie siedzenia zapewnia jednak dobrą izolację, a impulsy widoczne między 6÷9s nie są przenoszone w inne miejsca (takie jak siedzenie i oparcie).

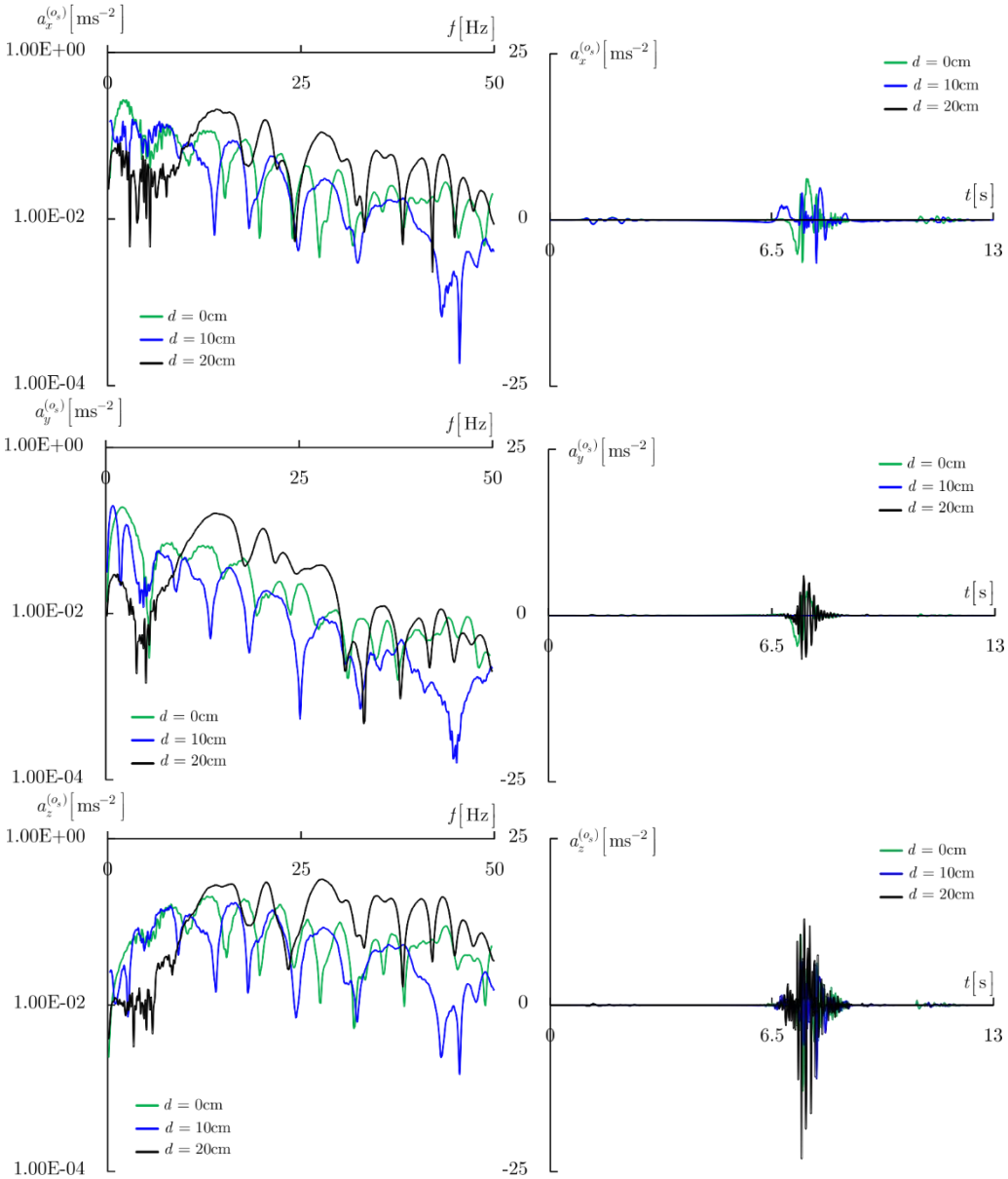


Fig. 10. Accelerations (filtered) calculated for seat base position in x, y, z direction; operation with loaded crane and different d value. Frequency (filtered) plots are on the left; time histories are on the right. Load cases: $D1 - F2 - d0 - F$, $D1 - F2 - d10 - F$ and $D1 - F2 - d20 - F$
Rys. 10. Przyspieszenia (filtrowane) siedzenia w kierunku osi x, y, z; żuraw obciążony i zmienna wartość d. Charakterystyka częstotliwościowa (filtrowana) (z lewej), czasowa (z prawej). Przypadki obciążenia: $D1 - F2 - d0 - F$, $D1 - F2 - d10 - F$ oraz $D1 - F2 - d20 - F$

4.2. Poziom drgań - indeksy

W tym rozdziale przedstawiono różne wskaźniki związane z odczuwanym dyskomfortem dla przeanalizowanych scenariuszy operacji żurawia.

Niektóre indeksy dotyczące scenariuszy obciążenia i parametrów projektowych wymienionych w sekcjach 3.2 i 3.1 przedstawiono w tab. 6. Obliczenia przeprowadzono z zestawem tarcia **Set-2** (przypadek $D1 - F2 - d0 - E$).

Table 6. Vibration discomfort parameters, case $D2 - F2 - d0 - E$
Tabela 6. Parametry dyskomfortu drgań: przypadek $D2 - F2 - d0 - E$

Parameter	rigid seat			flexible seat			back			feet		
	x	y	z	x	y	z	x	y	z	x	y	z
RMS	0.01	0.00	0.00	0.06	0.21	0.00	0.07	0.22	0.00	0.01	0.00	0.00
PV_{RMS}		0.01			0.22			0.12			0.00	
SV_{RMS}					0.25							
RMQ	0.01	0.00	0.01	0.10	0.32	0.01	0.10	0.33	0.01	0.03	0.01	0.01
VDV	0.02	0.00	0.01	0.19	0.61	0.01	0.20	0.63	0.01	0.05	0.01	0.02
PV_{VDV}		0.02			0.61			0.63			0.05	
K	20.51	12.68	62.06	5.44	5.21	20.25	5.44	5.03	5.68	54.81	24.95	68.61
C_f	6.90	5.32	11.66	3.21	3.28	6.99	3.20	3.18	3.09	10.64	8.37	11.17

Podobnie, dla żurawia obciążonego ładunkiem oraz charakteryzującego się wartościami współczynników tarcia **Set-1** oraz **Set-2**, wyniki dla parametrów charakterystycznych pokazano w Tab. 7 and 8.

Table 7. Vibration discomfort parameters, case $D1 - F1 - d20 - F$
Tabela 7. Parametry dyskomfortu drgań: przypadek $D1 - F1 - d20 - F$

Parameter	rigid seat			flexible seat			back			feet		
	x	y	z	x	y	z	x	y	z	x	y	z
RMS	0.19	0.04	0.19	0.13	0.24	0.06	0.16	0.25	0.03	0.50	0.09	0.15
PV_{RMS}		0.27			0.28			0.18			0.14	
SV_{RMS}					0.36							
RMQ	0.38	0.09	0.47	0.23	0.36	0.10	0.30	0.37	0.06	1.21	0.19	0.35
VDV	0.72	0.17	0.89	0.44	0.68	0.20	0.56	0.70	0.11	2.30	0.36	0.66
PV_{VDV}		0.97			0.71			0.48			0.58	
K	13.90	20.27	37.11	10.39	4.95	9.98	11.23	4.75	13.14	34.62	22.47	31.63
C_f	6.06	7.10	10.09	4.94	3.40	5.21	5.36	3.23	5.63	9.34	7.72	9.90

Table 8. Vibration discomfort parameters, case $D1 - F2 - d20 - F$
Tabela 8. Parametry dyskomfortu drgań: przypadek $D1 - F2 - d20 - F$

Parameter	rigid seat			flexible seat			back			feet		
	x	y	z	x	y	z	x	y	z	x	y	z
RMS	1.04	0.73	2.01	0.13	0.24	0.07	0.17	0.25	0.04	5.11	3.10	1.38
PV_{RMS}		2.38			0.28			0.18			1.59	
SV_{RMS}					1.63							
RMQ	1.67	1.07	2.45	0.24	0.37	0.15	0.31	0.37	0.07	6.53	2.28	1.64
VDV	3.18	2.04	4.65	0.45	0.69	0.29	0.59	0.71	0.13	12.39	4.33	3.11
PV_{VDV}		4.92			0.73			0.51			3.13	
K	29.38	29.10	53.16	11.69	4.88	9.10	12.21	4.69	8.28	64.23	65.46	44.08
C_f	8.91	8.56	11.87	5.10	3.36	4.59	5.47	3.20	4.30	13.02	12.46	11.09

Niektóre parametry zdefiniowane w sekcji 3.3, obliczone dla założonych parametrów tarcia w połączeniach żurawia, określonych przez **Set-1** i **Set-2**, przedstawiono na rys. 11. Wartości filtrowane RMS, zredukowane do jednej wartości, $SV_{RMS}^{(c)}$, są pokazane (obliczone) jak pokazano na rys. 6. Wyniki są obliczane dla komfortu siedzenia żurawia, gdy zawieszenie siedzenia jest podatne. Dla porównania, wyniki wskazują również poziom RMS, kiedy założono sztywne podparcie.
Czas (w godzinach dla **Set-1** i minutach dla **Set-2**) wymagany do kumulacji pożądanego poziomu wartości energii drgań (tj. $15\text{ms}^{-1.75}$) jest zestawiony na rys.

11. Jest to czas obliczony z uwzględnieniem całkowitej wartości energii skojarzonej z drgań wyznaczony zależnością:

$$T_{15}^{(c)} = \left(\frac{15}{\sqrt{\sum_{\alpha\in\{o_b,o_s,o_f\}} \left(VDV_c^{(\alpha)}\right)^2}} \right)^4 t, \tag{5}$$

gdzie $t = 13\text{s}$.

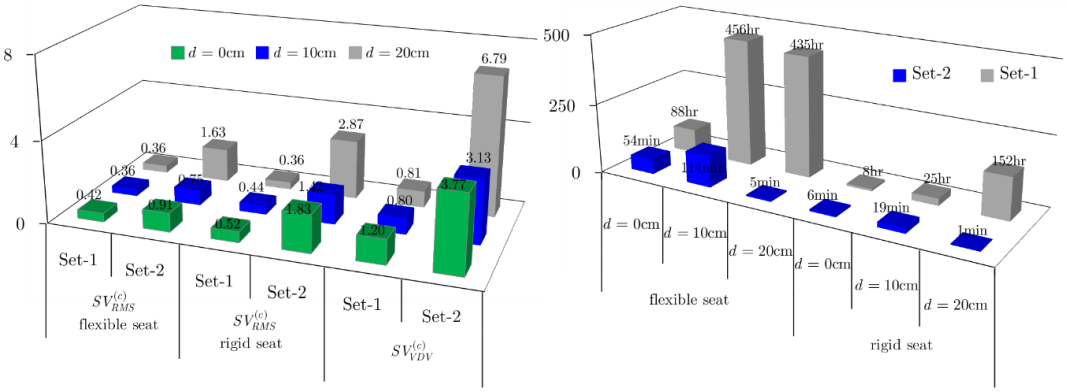


*Fig. 11. $SV_{RMS}^{(c)}$ and $SV_{VDV}^{(c)}$ values (left), $T_{15}^{(c)}$ time (right). Loaded crane operation, friction coefficients: **Set-1** and **Set-2**.*
*Rys. 11. Wartości wskaźników $SV_{RMS}^{(c)}$ i $SV_{VDV}^{(c)}$ (z lewej), $T_{15}^{(c)}$ (z prawej). Żuraw obciążony, współczynniki tarcia: **Set-1** i **Set-2**.*

Dla współczynników tarcia założonych w **Set-1**, ogólny poziom dyskomfortu można oszacować jako "lekko uciążliwy" podczas pracy z podatnym i sztywnym siedzeniem. Również czas ekspozycji jest duży, szczególnie w przypadku żurawia wyposażonego w podatne siedzenie. Wartości $T_{15}^{(c)}$ pokazują znacznie większą różnicę pomiędzy izolowanymi i sztywnymi siedzeniami. Ponieważ ta miara lepiej pasuje do charakterystyki sygnałów, powinna być ważniejsza niż indeks $SV_{RMS}^{(c)}$. Stąd oczekiwane są znacznie gorsze warunki dla operatora, gdy siedzenie jest sztywno połączone z kolumną. W najgorszym przypadku praca 8 godzin w takich warunkach będzie postrzegana jako poważny / ogromny dyskomfort i potencjalnie niebezpieczny dla zdrowia.
Różną tendencję uzyskuje się dla współczynników tarcia **Set-2**: znacznie gorsze wyniki (większy dyskomfort) zostały obliczone przy manipulowaniu ładunkiem z większym przesunięciem między osią chwytaka a położeniem środka ciężkości ($d = 20\text{cm}$). Warunki te generowały podwójną wartość $SV_{RMS}^{(c)}$ i zmniejszały czas $T_{15}^{(c)}$ do prawie minimum (kilka minut). W przypadku siedzenia z elementami tłumiącymi wartość RMS na poziomie $0.8 \div 1.6$ można sklasyfikować jako umiarkowany dyskomfort, natomiast w przypadku bez takich elementów (sztywne siedzenie) ta sama operacja doprowadzi do bardzo niewygodnego poziomu.
Rozważając czas trwania takiego poziomu drgań, operator nie powinien pracować na "sztywnym siedzeniu", zwłaszcza gdy wystąpi większe tarcie. Zakładając, że pnie nie zawsze są chwytane przy odległości $d = 20\text{cm}$, ale najczęściej wokół $d = (0 \div 10)\text{cm}$, operator może pracować na żurawiu w ograniczonym czasie $1 \div 2$ h a wskaźnik VDV uzyskałby wartość nieakceptowalną. Biorąc pod uwagę typowe wykorzystanie żurawia leśnego (czas załadunku / rozładunku nie jest przekraczany zazwyczaj w ciągu dnia), szacunkowe wyniki pokazują, w jaki sposób obecny projekt maszyny może mieć wpływ na aspekty zdrowotne.

Różne wyniki uzyskuje się w przypadku pracy z nieobciążonym żurawiem; w takich warunkach drgania wywołane operowaniem pustym żurawiem nie powodują znacznego dyskomfortu operatora, a poziom tarcia również nie ma istotniejszego znaczenia (dla wskaźników używanych w ocenie dyskomfortu). Dominującym efektem, w przypadku nieobciążonego żurawia, jest ruch spowodowany napędami, a ponieważ nie ma obciążenia (małe siły bezwładności), odczuwany dyskomfort jest praktycznie identyczny. Połączoną wartość dla siedzenia: $SV_{RMS}^{(O_S)} = 0.25$ (oba zestawy tarcia) można sklasyfikować jako "niekomfortową".
Przedstawione wyniki obejmują tylko jedną konfigurację zawieszenia siedzenia. Zwykle wykonuje się szereg iteracji w celu znalezienia właściwej równowagi pomiędzy pożądanym komfortem i ograniczeniami konstrukcyjnymi.

4. Wnioski

Przedstawiony model matematyczny oraz oprogramowanie komputerowe posłużyły do analizy wpływu tarcia na pracę operatorów żurawi. Obliczone odpowiedzi dynamiczne mogą być przydatne dla projektantów we wczesnej fazie projektowania urządzeń tego typu. Przedstawione wyniki potwierdzają, że wysokie wartości współczynników tarcia mogą mieć znaczący wpływ na dyskomfort pracy operatora. Dyskomfort pracy operatorów żurawi należy wziąć pod uwagę już w początkowej fazie projektowania.
Model przedstawiony w niniejszym pracy można zastosować do wielu innych aspektów typowych scenariuszy roboczych żurawi. W niniejszej pracy bardziej szczegółowo analizowane są tylko przyspieszenia wybranych punktów konstrukcji, jednak model komputerowy pozwala badać znacznie więcej parametrów, takich jak obciążenia w określonych podzespołach, wpływ funkcji napędowej na te obciążenia, możliwa jest zatem optymalizacja parametrów geometrycznych i sztywności. Istotna jest także prostota i skuteczność modelu, zwłaszcza w przypadku badania konkretnych aspektów układu, w przypadku wykonywania obliczeń wariantowych.
Bardziej zaawansowane modele, uwzględniające na przykład podatność elementów nośnych, mogą również być właściwym kierunkiem dla bardziej szczegółowej analizy drgań i komfortu. Można to rozwiązać w podobny sposób, jak przedstawiono w tej pracy, tj. poprzez rozszerzenie modelu. Wadą będzie czas potrzebny na analizy oraz fakt, że wymagane są bardziej złożone dane wejściowe.

Literatura

1. Augustynek K, Urbaś A. Comparison of bristles' friction models in dynamics analysis of spatial linkages. Mechanics Research Communications 2017; doi.org/10.1016/j.mechrescom.2017.01.003.
2. Åström K J, Canudas-de-Witt C. Revisiting the LuGre model. IEEE Control Systems Magazine. Institute of Electrical and Electronics Magazine 2008; 28(6): 101-114, doi: 10.1109/MCS.2008.929425
3. Cann A P, Salmoni A W, Vi P, Eger T R. An exploratory study of whole-body vibration exposure and dose while operating heavy equipment in the construction industry. Applied Occupational and Environmental Hygiene 2003; 18: 999-1005.
4. Courtney-Pratt J S, Eisner E. The effect of a tangential force on the contact of metallic bodies. Proceedings of the Royal Society 1957; A: 529-550.
5. Craig J J. Introduction to robotics. Mechanics and control. Addison-Wesley Publishing Company, Inc., 1989.

6. BS 6841. Guide to measurement and evaluation of human exposure to whole-body mechanical vibration and repeated shock. British Standard, 1987.
7. Denavit J, Hartenberg R S. A kinematic notation for lower-pair mechanisms based on matrices. Journal of Applied Mechanics 1995; 23: 215-221.
8. Giacomin M, Hacaambwa T M. Performance of ISO2631 and BS6841 Comfort Criteria for Evaluating Automobile Road Vibrations, O1A1083, ATA 7th International Conference on the Role of Experimentation in the Modern Automotive Product Development Process, Florence, Italy, May 23-25, 2001.
9. Griffin M J. Handbook of Human Vibration. Elsevier Academic Press, 1990.
10. ISO 2631-1. Mechanical vibration and shock - Evaluation of human exposure to whole body vibration. Part 1: General requirements. International Organization for Standardization, 1997.
11. Jurevič E I (ed.). Dynamics of robot control. Nauka. Moscow: 1984. (in Russian)
12. La Hera P X, Morales D O. Non-linear dynamics modelling description for simulating the behavior of forestry cranes. International Journal of Modelling Identification and Control 2014; 21(2): 125-138.
13. La Hera P, Morales D O. Model-based development of control systems for forestry cranes. Journal of Control and Science and Engineering 2015; vol., no. ID 256951: doi:10.1155/2015/256951.
14. Legnani G, Casalo F, Righettini P, Zappa B. A homogeneous matrix approach to 3D kinematics and dynamics – I. Theory, Mechanism and Machine Theory 1996; 31(5): 586-605.
15. Legnani G, Casalo F, Righettini P, Zappa B. A homogeneous matrix approach to 3D kinematics and dynamics – II. Applications to chains of rigid bodies and serial manipulators. Mechanism and Machine Theory 1996; 31(5): 573-587.
16. Mansfield N J. Human response to vibration. CRC Press LLC, Boca Raton, Florida, 2004.
17. Marques F, Flores P, Pimenta Claro J C, Lankarani H M. A survey and comparison of several friction force for dynamic analysis of multibody mechanical systems. Nonlinear Dynamics 2016; 86(3): 1407-1443, doi:10.1007/s11071-016-2999-3.
18. Morales D O, Westerberg S, La Hera P X, Mettin U, Freidovich L, Shiriaev A S. Increasing the level of automation in the forestry logging process with crane trajectory planning and control. Journal of Field Robotics 2014; 31(3): 343-363.
19. Papadopoulos E, Sarkar S. On the dynamic modeling of an articulated electrohydraulic forestry machine. in: Proceedings of the 1996 AIAA Forum on Advanced Developments in Space Robotics, WI, 1-2 August, 1996.
20. Papadopoulos E, Frenette R, Mu B, Gonthier Y. On the modeling and control of an experimental harvester machine manipulator. in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, Grenoble, France, 8-12 September, 1997.
21. Pennestri E, Rossi V, Salvini P, Valentini P P. Review and comparison of dry friction force models. Nonlinear Dynamics 2016; 83(4): 1785-1801.
22. Posiadała B. Influence of crane support system on motion of the lifted load. Mechanism and Machine Theory 1997; 32(1): 9-20.
23. Posiadała B. Modeling and analysis of the dynamics of load carrying system. in: Proc. of World Congress on Engineering and Computer Science, San Francisco, USA, 2012.

24. Posiadała B, Waryś P, Cekus D, Tomala M. The dynamics of the forest crane during the load carrying. International Journal of Structural Stability and Dynamics 2013; 13(7), doi: 10.1142/S0219455413400130.
25. Stribeck R. Die wesentlichen Eigenschaften der Gleit- und Rollenlager, Zeitschrift des Vereines Deutscher Ingenieure 2013; 46(36).
26. Urbaś A. Analysis of flexibility of the support and its influence on dynamics of the grab crane. Latin American Journal of Solids and Structures 2013; 10(1): 109-121.
27. Urbaś A. Application of the Dahl friction model in the dynamics analysis of grab cranes, MATEC Web of Conferences 8347, 03008. doi: 10.1051/matecconf/201 CSNDD 2016 68303008, 2016.
28. Urbaś A. Computational implementation of the rigid finite element method in the statics and dynamics analysis of forest cranes. Applied Mathematical Modelling 2017; 46: 750-762, doi: 10.1016/j.apm.2016.08.006.
29. Urbaś A, Harlecki A. Application of the rigid finite element method and the LuGre friction model in the dynamic analysis of the grab crane. in: Proceedings of 4th Joint International Conference on Multibody System Dynamics, Montreal, Canada, May 29-June 2, 2016.
30. Urbaś A, Szczotka M. Modelling friction phenomena in the dynamics analysis of forest cranes. Engineering Transactions 2016; 64(4): 393-400.
31. Szczotka M. Simulation and optimisation of the steering kickback performance. Journal of Theoretical and Applied Mechanics 2011; 49(1): 187-208.
32. Wittbrodt E, Szczotka M, Maczyński A, Wojciech S. Rigid Finite Element Method in Analysis of Dynamics of Offshore Structures. Ocean Engineering & Oceanography. Springer. Berlin-Heidelberg: 2013.

dr inż. Agnieszka Chudzik

Department of Dynamics
Lodz University of Technology
Stefanowskiego 1/15, 90-537 Lodz, Poland
E-mail: Agnieszka.Chudzik@p.lodz.pl
phone: +48 426312420

dr hab. inż. Bogdan Warda

Department of Vehicles and Fundamentals of Machine Design
Lodz University of Technology
Stefanowskiego 1/15, 90-537 Lodz, Poland
E-mail:Bogdan.Warda@p.lodz.pl
phone: +48 426312251

Wpływ wewnętrznego luzu promieniowego na trwałość zmęczeniową promieniowego łożyska walcowego

Effect of radial internal clearance on the fatigue life of the typical radial cylindrical roller bearing

Słowa kluczowe: *łożysko walcowe, rozkład naprężeń, trwałość zmęczeniowa, metoda elementów skończonych, luz promieniowy*

Keywords*: rolling bearing; stress distribution; fatigue life; finite element method; radial clearance*

Streszczenie: W pracy przedstawiono wyniki badań wpływu wewnętrznego luzu promieniowego w promieniowym łożysku walcowym na jego prognozowaną trwałość zmęczeniową. Wykorzystując zagadnienie Boussinesqa dla półprzestrzeni sprężystej i metodę elementów skończonych, określono rozkłady naprężeń podpowierzchniowych niezbędne do oszacowania prognozowanej trwałości zmęczeniowej łożyska. W obliczeniach uwzględniono geometryczne parametry łożyska: jego luz promieniowy i kształt części tocznych. Do określenia wartości luzu promieniowego wykorzystano model Lundberga i Palmgrena. Zastosowanie programu ANSYS pozwoliło na przeprowadzenie analizy rozkładu naprężeń von Misesa w dowolnej płaszczyźnie współpracujących ze sobą elementów tocznych łożyska. Analiza obliczeń wykazała, że promieniowe łożysko walcowe będzie miało najwyższą wytrzymałość z niewielkim luzem promieniowym.

Abstract: The paper presents result of the research on influence of internal radial clearance in radial cylindrical roller bearing on its fatigue durability. By solving the Boussinesq problem for the elastic half-space and finite elements method, stress distributions were determined, necessary to estimate predicted fatigue life of the bearing. The calculations took into account geometrical parameters of the bearing: its radial clearance and shape of rolling parts. Predicted radical clearance was computed by using the Lundberg and Palmgren model. ANSYS program allowed to introduce the analysis of von Mises stress distribution in any plane of cooperating components. The outcome revealed, radial cylindrical roller bearing will have highest endurance with slight radial clearance.

Oznaczenia

A	stała materiałowa
B	szerokość łożyska
C	nośność dynamiczna łożyska
D	zewnętrzna średnica łożyska
D_r	średnica wałeczka
E	moduł Younga
F_r	obciążenie promieniowe łożyska
INT	funkcja integer
L	trwałość zmęczeniowa
L_r	długość wałeczka
Q_r	wypadkowa siła normalna w styku wałeczka z bieżnią główną
S_t	stała Stribeck'a
U_z	lokalne przemieszczenie wewnątrz elementów łożyska
Z	głębokość występowania naprężeń maksymalnych naprężeń podpowierzchniowych wzdłuż osi x
Z_r	liczba wałeczków w łożysku
b	połowa szerokości pola styku
c	wykładnik w równaniu prawdopodobieństwa trwałości
d	średnica otworu łożyska
d_{bi}	średnica bieżni pierścienia wewnętrznego
d_{bo}	średnica bieżni pierścienia zewnętrznego
e	wykładnik Weibulla
g	luz promieniowy w łożysku
h	wykładnik w równaniu prawdopodobieństwa trwałości
i	bieżnia pierścienia wewnętrzego (indeks)
j	numer wałeczka
l	długość pola styku wałeczka z bieżnią główną
o	bieżnia pierścienia zewnętrzego (indeks)
n	liczba par wałeczków przenoszących obciążenie
r_b	promień bieżni głównej
r_c	sfazowanie technologiczne
u	liczby cykli obciążenia przypadająca na jeden obrót
δ_{max}	odkształcenie w styku najbardziej obciążonego wałeczka
δ_c	całkowite przemieszczenie osi łożyska
v	współczynnik Poissona
σ	maksymalne zastępcze naprężenie von Misesa wzdłuż osi x
φ	prawdopodobieństwo trwałości elementu pierścienia łożyska
ψ	kąt mierzony po obwodzie łożyska
ψ_j	kąt pomiędzy wałeczkami
ψ_ε	kąt opisujący wielkość strefy łożyska przenoszącej obciążenie

1. Wstęp

Od wielu lat prowadzone są badania łożysk tocznych i na podstawie otrzymanych wyników przewiduje się ich trwałości zmęczeniową. Czynnikami decydującymi, mającymi bezpośredni wpływ na określenie trwałości zmęczeniowej są zjawiska zachodzące pomiędzy elementami tocznymi łożyska a bieżniami. Analityczne metody obliczeniowe trwałości zmęczeniowej nie pozwalają na uwzględnienie wielu czynników, np.:

– luzu promieniowego w łożysku,
– kształtu tworzących elementów łożyska tocznego,
– niewspółosiowości pierścieni łożyska,
– obciążenia osiowego promieniowego łożyska walcowego.

Ponadto analityczne metody obliczeniowe nie pozwalają przewidzieć wyniku niedokładności produkcji elementów tocznych, niewystarczającej sztywności zabudowy i ugięcia rolki nośnej. MES staje się powszechnie używanym narzędziem w wielu dziedzinach nauki i technologii. Zastosowanie MES pozwala na wprowadzenie parametrów, które do tej pory w obliczeniach analitycznych nie można było uwzględnić. Celem analizy numerycznej i analitycznej było zbadanie zjawisk zachodzących na powierzchni kontaktu elementów tocznych z bieżniami łożyska. Ze względu na złożoność badanego modelu do obliczeń numerycznych autorzy opracowali model bryłowy fragmentu łożyska uwzględniający połowę wałeczka i odpowiednich fragmentów obu pierścieni.

Powierzchnia kontaktu współpracujących elementów o niepłaskich powierzchniach styku wszystkich typów urządzeń jest bardzo mała. Przy stosunkowo niewielkich siłach obciążenia generowane są znaczne wartości nacisków powierzchniowych, co powoduje naprężenia podpowierzchniowe. Naprężenia podpowierzchniowe mają znaczący wpływ na trwałość elementów konstrukcyjnych. Ten rodzaj problemu występuje w przekładniach zębatych i wszelkiego rodzaju łożyskach tocznych. H. Hertz [7] opublikował badania dotyczące powierzchni kontaktu pod ciśnieniem. Przedstawił ogólne rozwiązanie kontaktu dwóch ciał sprężystych poddanych poprzecznemu obciążeniu. Artykuł ten był podstawą badań Lundberga [14], który przedstawił rozwiązanie analityczne modelu, w którym walec jest ściskany pomiędzy dwiema płaszczyznami. W 1907 r. Stribeck [23] wykazał, że obciążenie zewnętrzne łożyska jest przenoszone przez część elementu tocznego, a obciążenie poszczególnych elementów tocznych zależy od ich położenia i wewnętrznego luzu w łożysku. Im większy luz pomiędzy elementami tocznymi, tym mniejsza liczba wałeczków, które przenoszą obciążenie i tym większa siła działa na wałeczek leżący na linii działania siły zewnętrznej. Stribeck [23] stwierdził, że wartość siły działającej na wałeczek w przypadku łożysk o zerowym luzie wewnętrznym wynosi 4,37 średniego obciążenia łożyska walcowego. Zasugerował, żeby wartość tej siły była 5 razy większa niż średnia siła dla łożyska z niezerowym luzem wewnętrznym. Palmgren [17] zaproponował, że w przypadku łożyska walcowego z zerowym luzem wewnętrznym wartość ta powinna wynosić 4,08. W przypadku innych łożysk o typowym luzie wewnętrznym, w tym w promieniowych łożyskach walcowych, wartość ta powinna być równa 5. Stała ta była nazywana stałą Stribeck'a.

Łożyska toczne są powszechnie stosowane w konstrukcjach inżynierskich. Są to elementy, które pozycjonują oś obrotu wału, który przenosi zarówno obciążenia poprzeczne, jak i wzdłużne. Łożysko toczne powinno charakteryzować się wysoką trwałością zapewniającą długie godziny pracy. Przez wiele lat łożyska toczne były przedmiotem zainteresowania badaczy, którzy badali czas działania łożysk aż do ich awarii w różnych warunkach pracy. Umożliwiło to określenie czasu bezawaryjnej pracy łożyska. Lundberg i Palmgren [15, 16] rozpoczęli badania nad wpływem obciążenia elementów tocznych na trwałość łożysk tocznych. Przeprowadzono również wiele kompleksowych badań różnych rodzajów styku powierzchni [9] i przedstawiono propozycje obliczania rozkładu obciążenia elementów tocznych różnych rodzajów łożysk [2, 3, 4, 5, 6, 8]. Wygodnym i tanim narzędziem do przeprowadzenia takich badań są symulacje przeprowadzane na modelach numerycznych łożysk. W tym celu budowane są modele numeryczne łożysk obciążane siłami zewnętrznymi (poprzecznymi, wzdłużnymi lub ich kombinacją). Badając model łożyska wieńcowego Kania [10] przedstawił charakterystyki: odkształcenie – obciążenie otrzymane na drodze obliczeń numerycznych.

Ricci [18] przeprowadził obliczenia numeryczne rozkładu obciążenia między elementami łożyska walcowego. Przedstawił wyniki zmian geometrii łożyska pod określonym obciążeniem. W 2011 Tang i wsp. [24] opracowali numeryczny model łożyska walcowego wykorzystując pakiet oprogramowania ANSYS. Obliczone zostały wartości nacisków powierzchniowych i maksymalnych naprężeń w styku elementów. W tym samym roku Laniado-Jacome [13] przeprowadził badania w celu ustalenia rozkładu obciążenia zewnętrznego na elementy toczne łożysk tocznych obciążonych siłą poprzeczną przy zmiennej prędkości obrotowej. Analizę przeprowadzono na modelu numerycznym i analitycznym. W Shah Maulik i in. [19] przeprowadzili badania nad wpływem modyfikacji kształtu tworzących wałeczka na rozkład ciśnienia w obszarze styku. Trwałość łożyska walcowego badano za pomocą pakietu ANSYS. Shingala i wsp. [21] przeprowadzili analizę numeryczną styku elementów tocznych w łożyskach igiełkowych stosowanych w skrzyni biegów. Badania, które rozszerzyły zakres analiz numerycznych o wpływ defektów na współpracę elementów tocznych, zostały przedstawione przez Deshapande i wsp. [1] oraz przez Shaha Rohit [20].

Jednym z najczęściej używanych programów do obliczeń numerycznych jest ANSYS. Badania prowadzone przy użyciu tego pakietu stale się rozwijają, a modele numeryczne są udoskonalane. Prezentowana praca jest próbą przybliżenia modelu numerycznego do geometrii i warunków obciążenia rzeczywistych łożysk. Głównym celem badań była ocena wpływu luzu promieniowego w łożysku tocznym na zjawiska zachodzące w styku elementów tocznych z bieżniami. W artykule przedstawiono metodologię postępowania przy określaniu naprężenia podpowierzchniowego niezbędnego do obliczenia trwałości zmęczeniowej metodą numeryczną.

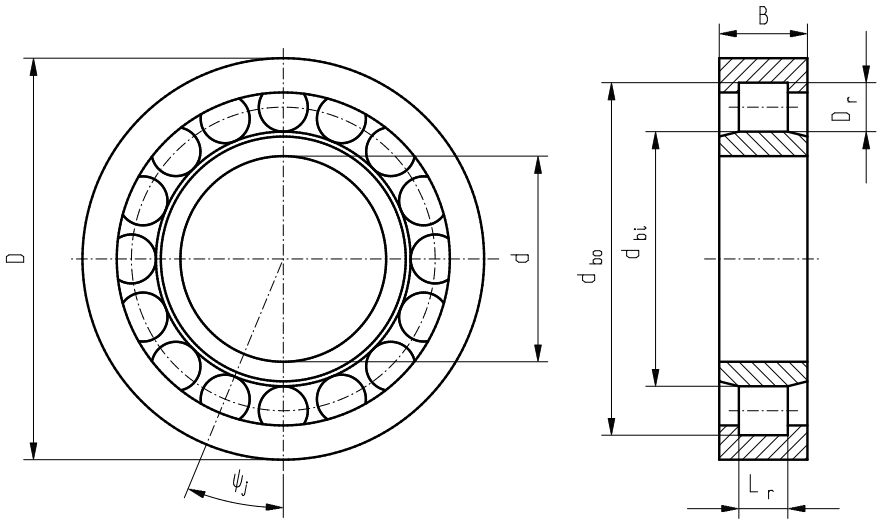
2. Rozkład obciążenia wewnętrznego

Łożyska walcowe są ważnym elementem w produkcji urządzeń mechanicznych i są szeroko stosowane w różnych dziedzinach. W przypadku większości zastosowań łożyska walcowe przenoszą obciążenie promieniowe. Promieniowe łożysko walcowe typu N lub NU umożliwia swobodny ruch wału w kierunku osiowym w stosunku do obudowy. Pozwala to przezwyciężyć różnice w rozszerzalności cieplnej materiałów stosowanych w konstrukcji. Łożyska typu NJ umożliwiają przenoszenie lekkich osiowych obciążeń wału w jednym kierunku. Łożyska walcowe umożliwiają przenoszenie stosunkowo wysokich obciążeń promieniowych i mogą pracować z dużymi prędkościami obrotowymi. Ich dopuszczalne prędkości przekraczają dopuszczalne prędkości dla łożysk baryłkowych i stożkowych. Wałeczki w łożyskach tocznych rzadko są idealnymi walcami. W celu zmniejszenia spiętrzenia naprężeń na końcach wałeczka, spowodowanego możliwym przemieszczeniem wału i obudowy, końce wałeczków są zaokrąglone, a profil tworzących skorygowany. Niewłaściwe ustawienie czołowych powierzchni pierścieni łożyska, odkształcenie wału, nachylenie wewnętrznego lub zewnętrznego pierścienia może spowodować pochylenie się rolek względem pierścieni.

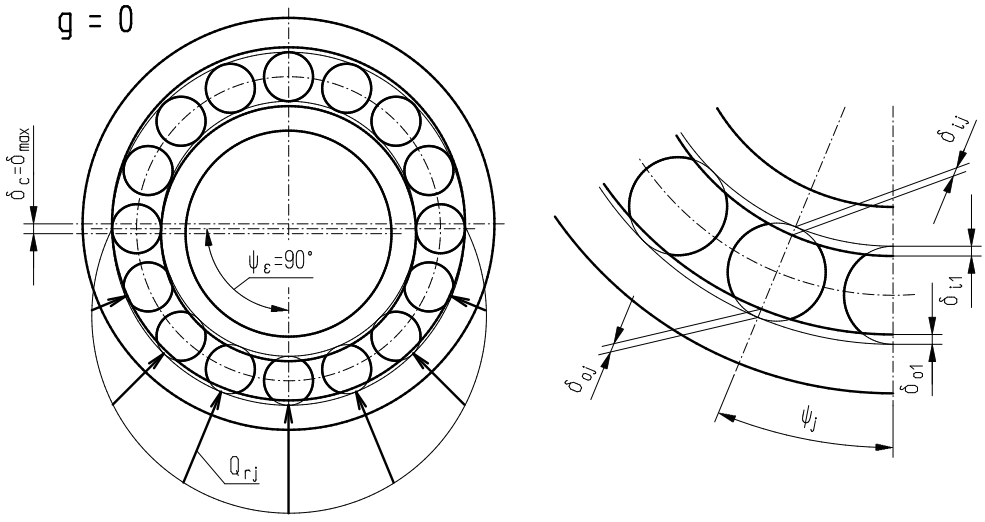
2.1. Rozkład obciążenia w promieniowym łożysku walcowym obciążonym siłą promieniową

Łożysko składa się z dwóch pierścieni (wewnętrznego i zewnętrznego) i od kilku do kilkunastu wałeczków umieszczonych między nimi w zależności od wielkości łożyska. Ich równomierny rozkład zapewnia koszyk. Konstrukcja pierścienia zewnętrznego łożyska typu NU lub pierścienia wewnętrznego łożyska typu N zapobiega przemieszczaniu się wałeczków w kierunku osiowym. W łożyskach typu NJ, konstrukcja wewnętrznego pierścienia

umożliwia jednostronne przemieszczenie wałeczków. Łożyska walcowe są produkowane z różnymi luzami promieniowymi od C1 do C5. Łożyska C1 mają najmniejszy luz promieniowy, a C5 największy. Większość łożysk produkowana jest z normalnym luzem promieniowym CN. Niezależnie od tego, wybierając odpowiednie pasowanie, można uzyskać wstępne obciążenie łożyska. Wtedy łożysko działa z luzem promieniowym mniejszym od zera ($g < 0$). Stribeck [23] wykazał, że rozkład obciążenia w łożysku tocznym obciążonym siłą promieniową jest nierównomierny. Obszar obciążenia łożyska z dodatnim luzem promieniowym jest mniejszy niż 180°. Rysunek 1 pokazuje nieobciążone łożysko z luzem $g = 0$, natomiast na rysunku 2 pokazano łożysko obciążone siłą promieniową F_r.



Rys. 1. Łożysko z luzem promieniowym $g = 0$ przed obciążeniem siłą F_r
Fig. 1. Bearing with radial clearance $g = 0$ before loading with force F_r



Rys. 2. Łożysko z luzem promieniowym $g = 0$ po obciążeniu siłą F_r
Fig. 2. Bearing with radial clearance $g = 0$ after loading with force F_r

Liczbę par elementów tocznych przenoszących obciążenie przy $g = 0$ obliczono wg wzoru przedstawionego poniżej:

$$n_{g=0} = INT\left(\frac{Z_r - 1}{4}\right). \quad (1)$$

Liczba elementów tocznych przenoszących obciążenie w sytuacjach pokazanych na rysunku 2 wynosi $2n_{g=0} + 1$. Kąt między rolkami obliczono korzystając ze wzoru:

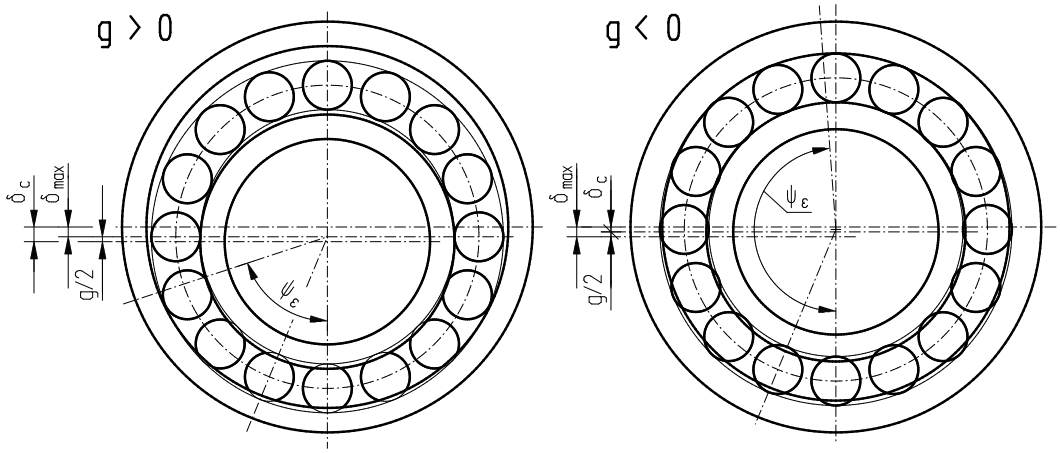
$$\psi_j = \frac{2\cdot\pi}{Z_r}. \tag{2}$$

Strefę łożyska przenoszącą obciążenie w przypadku dodatniego *(g> 0)* i ujemnego *(g <0)* luzu przedstawiono na rysunku 3. Wielkość tej strefy opisana jest przez kąt:

$$\psi_\varepsilon = arccos\left(\frac{g}{2\cdot\delta_c}\right), \tag{3}$$

$$\delta_c = \delta_{max} + g/2. \tag{4}$$

W przypadku, gdy luz promieniowy $g > 0$, liczba wałeczków poddanych obciążeniu jest mniejsza niż w przypadku $g = 0$. Natomiast, gdy $g < 0$ liczba wałeczków może osiągnąć wartość maksymalną równą liczbie wałeczków w łożysku Z_r.



Rys. 3. Łożysko z dodatnim i ujemnym luzem promieniowym po obciążeniu siłą F_r
Fig. 3. Bearing with positive and negative radial clearance after loading with force F_r

Równanie równowagi łożyska obciążonego siłą promieniową F_r ma postać:

$$F_r = Q_{r1} + 2Q_{r2}cos\psi_2 + \cdots + 2Q_{rj}cos\psi_j + \cdots + 2Q_{rn}\,cos\psi_n. \tag{5}$$

Obciążenie wałeczka nr 1 siłą Q_{r1} powoduje całkowite odkształcenie wałeczka δ_{max}, będące sumą odkształceń pomiędzy wałeczkiem i pierścieniem wewnętrznym δ_{i1} oraz wałeczkiem i pierścieniem zewnętrznym δ_{o1}:

$$\delta_{max} = \delta_{i1} + \delta_{o1}. \tag{6}$$

Wartość odkształcenia pozostałych wałeczków można obliczyć ze wzoru:

$$\delta_{rj} = \delta_c \cdot cos\psi_j - g/2, \tag{7}$$

gdzie:

$$\delta_{rj} = \delta_{ij} + \delta_{oj}. \tag{8}$$

Znając wartość odkształcenia, siłę Q_{rj} w styku wałeczka z bieżnią można wyznaczyć według wzoru Palmgrena [15, 16]:

$$Q_{rj} = 78000\ \delta_{i,oj}^{10/9} \cdot l^{8/9}\,, \tag{9}$$

gdzie długość pola styku wałeczka z bieżnią określona jest przez:

$$l = L_r - 2r_c\,. \tag{10}$$

Równanie (5) zostało rozwiązane numerycznie przy użyciu programu komputerowego ROLL1, zbudowanego według metodyki opisanej w pracach [25, 26]. Uzyskane wyniki umożliwiają określenie stałej Stribeck'a:

$$S_t = Q_{r1} \cdot Z_r / F_r. \tag{11}$$

3. **Prognozowanie trwałości zmęczeniowej promieniowego łożyska**

Wielkość luzu promieniowego w promieniowym łożysku walcowym decyduje o rozkładzie obciążenia na elementy toczne, a zatem ma znaczny wpływ na jego trwałość zmęczeniową. Aby określić trwałość zmęczeniową łożyska, zastosowano metodykę opisaną w [25].

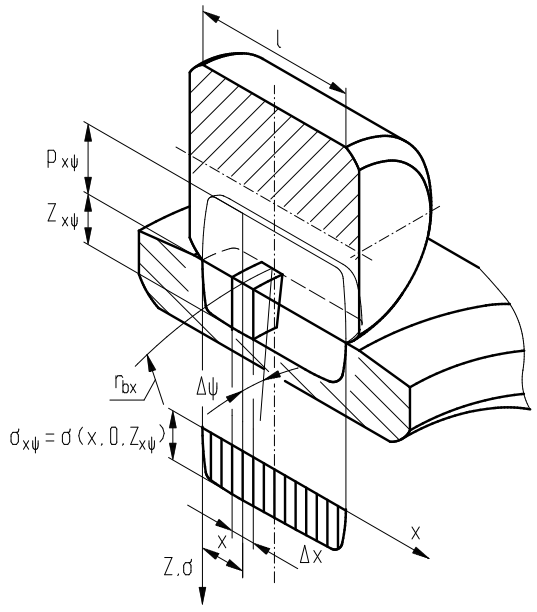
Prawdopodobieństwo trwałości stacjonarnego pierścienia łożyska przyjmuje postać:

$$ln\frac{1}{\varphi_o} = A \cdot u_o^e\, L_o^e \int_0^{2\pi} \int_0^l r_{box}\ \sigma_{ox\psi}^c\, Z_{ox\psi}^{1-h}\, dx\, d\psi. \tag{12}$$

Prawdopodobieństwo trwałości pierścienia łożyska wirującego względem obciążenia jest następujące:

$$ln\frac{1}{\varphi_i} = A \cdot u_i^e\, L_i^e\, 2\pi \left[\frac{1}{2\pi}\int_0^{2\pi}\left(\int_0^l r_{bix}\ \sigma_{ix\psi}^c\, Z_{ix\psi}^{1-h}\, dx\right)^{1/e} d\psi\right]^e. \tag{13}$$

W powyższych wzorach L jest liczbą obrotów, u liczbą cykli obciążenia na jeden obrót, $\sigma_{x\psi}$ jest maksymalnym naprężeniem podpowierzchniowym von Misesa, a $Z_{x\psi}$ jest głębokością, na której występują te naprężenia (Rys. 4). A jest stałą materiałową, która w przypadku łożyska będącego przedmiotem rozważań jest równa $4.5 \cdot 10^{-40}$. Wartości wykładników występujących w równaniu (12, 13) przyjmuje się: c = 31/3, h = 7/3, e = 9/8.



Rys. 4. Rozkład nacisków powierzchniowych i naprężeń podpowierzchniowych decydujących o trwałości zmęczeniowej promieniowego cylindrycznego łożyska walcowego

Fig. 4. Distributions of pressure and subsurface stresses deciding on fatigue life of radial cylindrical roller bearing

Korzystając ze wzoru (14) określono trwałość zmęczeniową badanego łożyska.

$$L = (L_o^{-e} + L_i^{-e})^{-1/e}. \quad (14)$$

Przy prawdopodobieństwie trwałości $\varphi_o = \varphi_i = 0{,}9$, obliczona wartość L jest tożsama z trwałością L_{10}. Wartość L reprezentuje liczbę obrotów, które łożysko może wykonać, zanim ulegnie zużyciu.

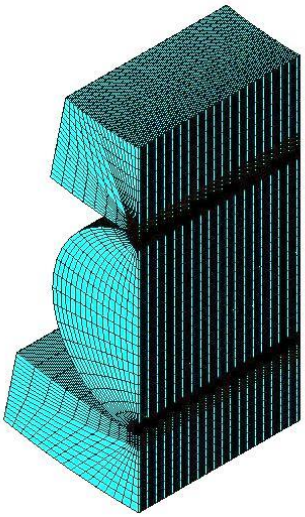
Równania (12) i (13) rozwiązano za pomocą programu ROLL2 [25]. Do wykonania obliczeń trwałości zmęczeniowej konieczne było określenie maksymalnych wartości zastępczych naprężeń podpowierzchniowych σ w strefie kontaktu między wałeczkami i bieżniami oraz głębokości ich występowania Z.

Do wyznaczenia naprężeń podpowierzchniowych zastosowano dwie metody obliczeń:

– metodę „half space" („półprzestrzeni sprężystej") wykorzystującą rozwiązanie Boussinesq opisujące odkształcenie półprzestrzeni sprężystej poddanej działaniu nacisków powierzchniowych - program komputerowy ROLL4 [25, 26],
– metodę elementów skończonych (MES).

Bryłowy model numeryczny MES został opracowany przy użyciu pakietu programu ANSYS - Augmented Lagrange. Wykorzystując warunki symetrii, obliczenia numeryczne wykonano dla połowy wałeczka z odpowiadającymi fragmentami bieżniami, co pokazano na rysunku 5. W modelu numerycznym użyto *8*-węzłowe elementy bryłowe typu *SOLID185*, w strefie kontaktu wałeczka z bieżniami - *CONTA174* i *TARGE170*. Dodatkowo w obliczeniach uwzględniono warunki symetrii i stopnie swobody wynikające z rzeczywistych warunków pracy.

W celu zwiększenia dokładności obliczeń zwiększono gęstość siatki elementów w obszarze przewidywanego kontaktu i zastosowano nierównomierny podział elementów stykowych. Odległość między węzłami siatki elementów powierzchni styku rolki wynosi 0,05 mm (przy średnicy wałeczka wynoszącej 15 mm). Odległości między węzłem pozostałych bieżni i walcem są regulowane pod kątem, eliminując błędy kształtu elementu. W rezultacie model numeryczny zawiera 200363 węzłów w 88910 elementach. Biorąc pod uwagę charakter styku łożyska tocznego z bieżniami, powierzchnia wałeczka była potraktowana jako powierzchnia styku. Powierzchnie bieżni przyjęto jako powierzchnie celu. Powierzchni zewnętrznej pierścienia odebrano możliwość przemieszczenia we wszystkich kierunkach. Wewnętrzny pierścień mógł poruszać się tylko w kierunku osi "z". Do obliczeń przyjęto współczynnik sztywności FKN = 1,5. W warunkach symetrii model został obciążony siłą F_r / 2.



Rys. 5. Siatka modelu numerycznego MES wałeczka z bieżniami
Fig. 5. FEM numerical model grid of the roller-raceways contact

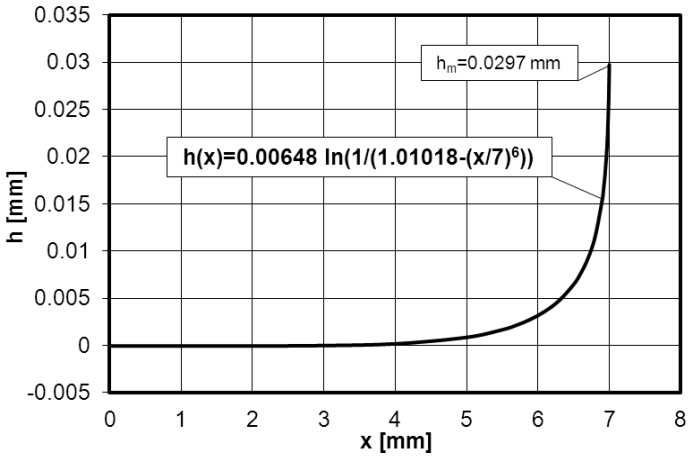
4. Przedmiot badań

Przedmiotem badań było cylindryczne łożysko walcowe NU 213 ECP [22]. Parametry badanego łożyska przedstawiono w tabeli 1.

Tabela 1. Parametry badanego łożyska NU 213 ECP [22]
Table 1. Parameters of the NU 213 ECP cylindrical roller bearing [22]

Średnica otworu łożyska	d = 65 mm
Zewnętrzna średnica łożyska	D = 120 mm
Szerokość łożyska	B = 23 mm
Średnica bieżni pierścienia zewnętrznego	d_{bo} = 108,5 mm
Średnica bieżni pierścienia wewnętrznego	d_{bi} = 78,5 mm
Średnica wałeczka	D_r = 15 mm
Długość wałeczka	L_r = 15 mm
Sfazowanie technologiczne	r_c = 0.5 mm
Liczba wałeczków w łożysku	Z_r = 16
Nośność dynamiczna	C = 122000 N

Wałeczki zostały skorygowane zgodnie z modyfikowaną korekcją logarytmiczną zaproponowaną przez Krzemińskiego-Fredę [11]. Zastosowany profil opisano w pracy [26] (Rys. 6). Parametry profilu korekcji logarytmicznej rzeczywistego łożyska typu NJ 213 ECP nie są podawane przez producenta.



Rys. 6. Profil wałeczka ze zmodyfikowaną korekcją logarytmiczną [26]
Fig. 6. Roller generator profile with a modified logarithmic correction [26]

5. Wyniki

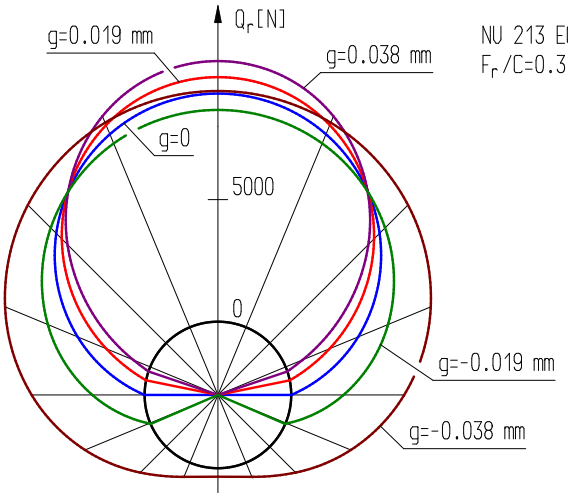
5.1. Obciążenie wałeczków

Według Krzemińskiego-Fredy [12] łożyska walcowe obciążone siłą promieniową F_r = 0,3 C pracują w trudnych warunkach roboczych. Normalne warunki pracy są uważane za takie, przy których obciążenie wynosi: F_r = (0.07−0.12) C. Badane łożysko obciążono siłą: F_r = 0.3 C = 36600 N i F_r = 0.1 C = 12200 N. Tabela 2 pokazuje rozkłady obciążenia

obliczone programem komputerowym ROLL1 dla wybranych luzów promieniowych. Rysunek 7 ilustruje rozkłady obciążenia promieniowego Q_r na rolkach przy F_r = 0,3 C.

Tabela 2. Rozkład obciążenia promieniowego na wałeczki łożyska NU 213 ECP
Table 2. Radial load distributions on NU 213 ECP bearing rollers

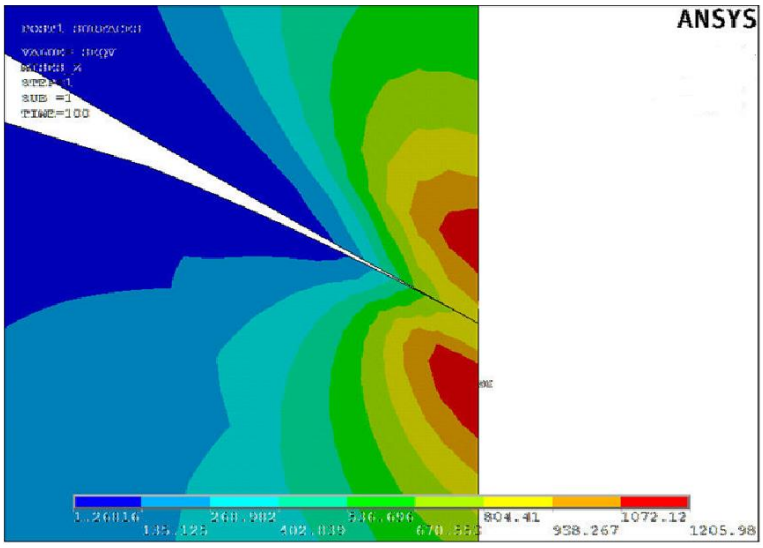
F_r = 36600 N						
g [mm]		-0.038	-0.019	0.0	0.019	0.038
ψ_ε [°]		180,0	113,36	90,00	78,49	71,32
Nr wałeczka	ψ_j [°]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]
1	0	9447	8672	9342	10015	10673
2	22.5	9069	8148	8556	8962	9353
3	45	8001	6676	6357	6036	5706
4	67.5	6429	4535	3213	1944	752
5	90	4624	2141	0	0	0
6	112.5	2888	51	0	0	0
7	135	1495	0	0	0	0
8	157.5	631	0	0	0	0
9	180	349	0	0	0	0
S_t		4.13	3.79	4.08	4.38	4.67
F_r = 12200 N						
g [mm]		-0.019	-0.0095	0.0	0.019	0.038
ψ_ε [°]		180,00	128,45	90,00	67,71	58,07
Nr wałeczka	ψ_j [°]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]	Q_{rj} [N]
1	0	3721	2875	3114	3701	4067
2	22.5	3598	2725	2852	3201	3344
3	45	3250	2304	2119	1821	1381
4	67.5	2735	1688	1071	12	0
5	90	2141	991	0	0	0
6	112.5	1562	343	0	0	0
7	135	1088	0	0	0	0
8	157.5	782	0	0	0	0
9	180	678	0	0	0	0
S_t		4.88	3.77	4.08	4.85	5.33



Rys. 7. Rozkłady obciążenia promieniowego na wałeczki łożyska NU 213 ECP
Fig. 7. The distributions of radial load on the rollers of NU 213 ECP bearing

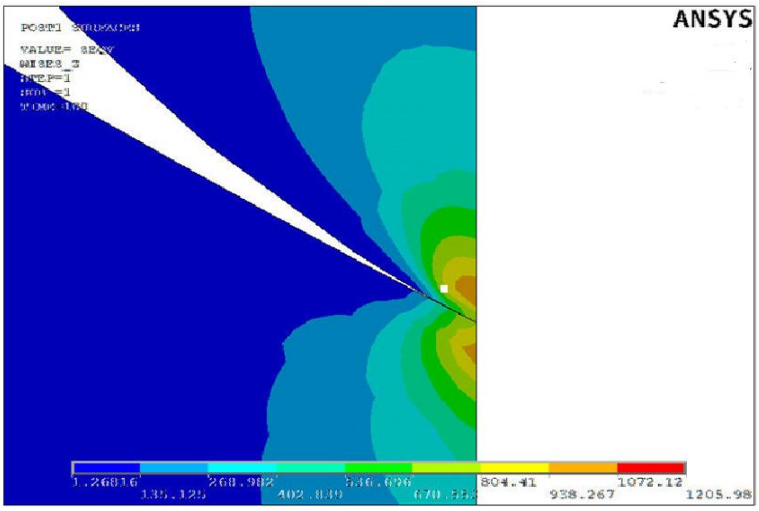
5.2. Rozkłady nacisków powierzchniowych i naprężeń von Misesa

Za pomocą obliczeń numerycznych uzyskano dane niezbędne do analizy zjawisk zachodzących w strefie kontaktu pary kinematycznej badanego modelu. Uzyskane wyniki obliczeń łożysk NU 213 ECP przedstawiono w postaci wykresów i map. Badane łożysko było obciążone siłą promieniową $F_r = 36600$ N przy założeniu, że osie elementów tocznych i bieżnie łożyska są równoległe, luz promieniowy wynosił $g = 0.038$ mm. Do obliczeń nacisków powierzchniowych i naprężeń von Misesa metodą „half space" przyjęto, że elementy toczne wykonane są z materiału sprężysto-idealnie plastycznego. Własności materiału zostały określone przez moduł Younga $E = 208$ GPa, współczynnik Poissona $\nu = 0{,}3$ i granicę plastyczności $\sigma_o = 1950$ MPa. Te same założenia dotyczące cech materiałowych zostały dokonane podczas obliczeń nacisków powierzchniowych i naprężeń von Misesa za pomocą MES. Rysunek 8 pokazuje graficzną mapę rozkładu maksymalnych naprężeń von Misesa pod powierzchnią styku wałeczka z bieżnią pierścienia wewnętrznego.

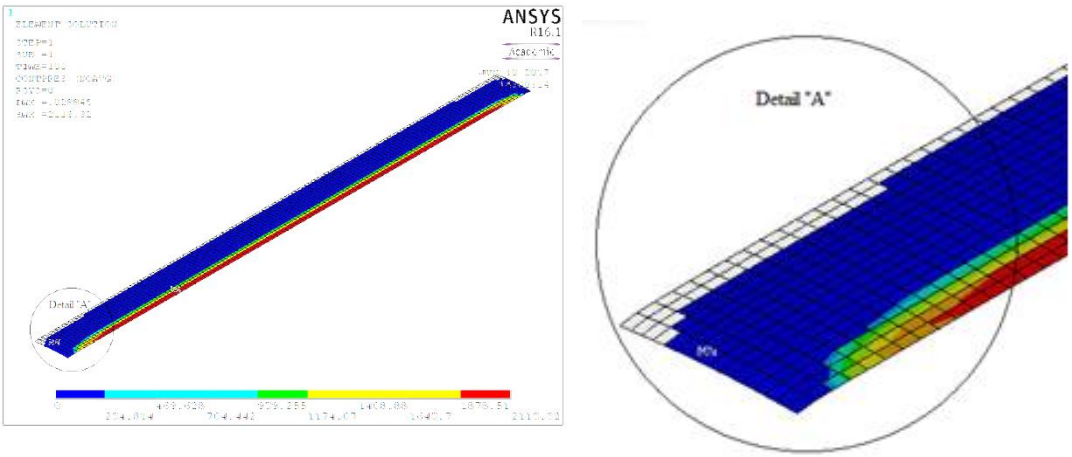


Rys. 8. Rozkład naprężeń von Misesa pod powierzchnią styku między wałeczkiem a bieżnią pierścienia wewnętrznego
Fig. 8. Distributions of von Mises stresses below the roller-inner ring raceway contact surface

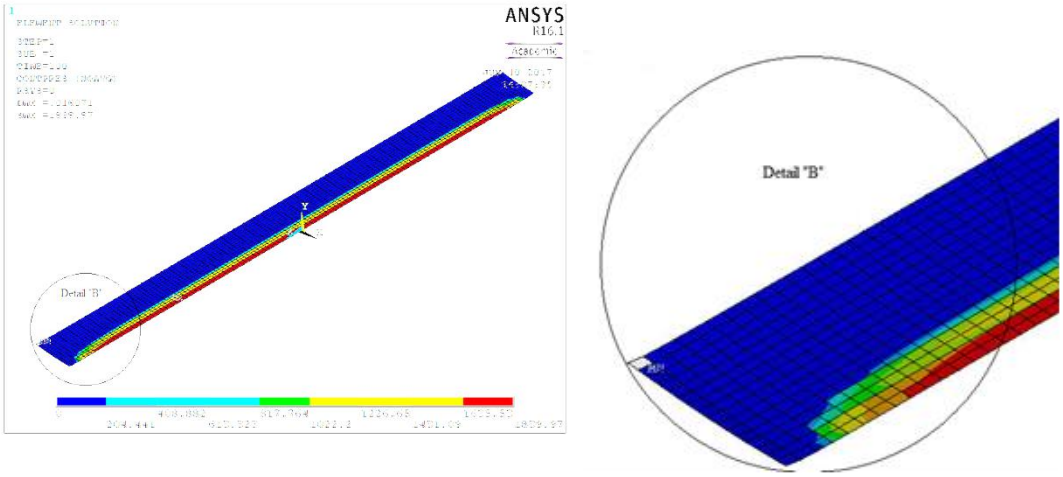
Rysunek 9 pokazuje rozkład maksymalnych naprężeń von Misesa poniżej powierzchni styku wałeczka z bieżnią zewnętrzną. Rysunki 10 i 11 pokazują powierzchnię styku wałeczka z wewnętrzną i zewnętrzną bieżnią wraz z powiększeniem szczegółów.



Rys. 9. Rozkład naprężeń von Misesa pod powierzchnią styku wałeczeka z bieżnią pierścienia zewnętrznego
Fig. 9. Distributions of von Mises stresses below the roller-outer ring raceway contact surface

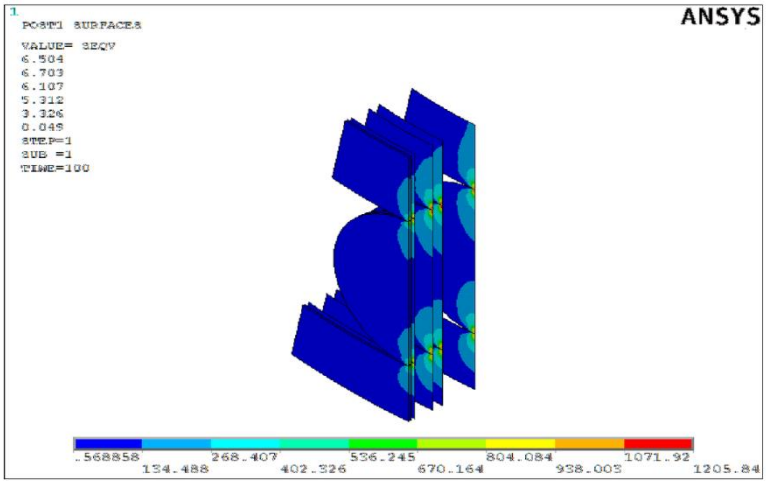


Rys. 10. Powierzchnia styku wałeczka z bieżnią pierścienia wewnętrznego
Fig. 10. Contact surface of the roller with the inner ring raceway

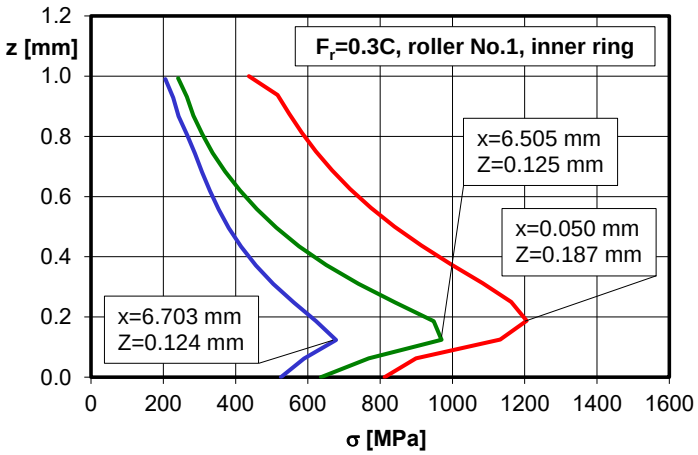


Rys. 11. Powierzchnia styku wałeczka z bieżnią pierścienia zewnętrznego
Fig. 11. Contact surface of the roller with the outer ring raceway

W wersji podstawowej ANSYS podaje tylko jedną wartość maksymalną i jedną wartość minimalną naprężenia von Misesa i współrzędne *(x, z)* miejsca wystąpienia tych naprężeń w całej strefie kontaktu. W niniejszej pracy autorzy przedstawiają możliwość zdefiniowania rozkładu naprężenia von Misesa w przekroju wyznaczonym przez dowolną wartość współrzędnej *x*, co pozwala na określenie maksymalnych wartości naprężeń σ i głębokości ich występowania *Z* w przekroju poprzecznym. Pakiet ANSYS umożliwia bezpośrednie określenie rozkładu naprężeń von Misesa w płaszczyźnie wybranej przez użytkownika. Przykład wyników obliczeń przedstawiono na rysunkach 12, 13 i 14. Wyniki uzyskane tą metodą umożliwiły określenie maksymalnych rozkładów naprężeń von Misesa i głębokości ich występowania wzdłuż linii styku. Rozkłady te są niezbędne do określenia prognozowanej trwałości zmęczeniowej łożyska.

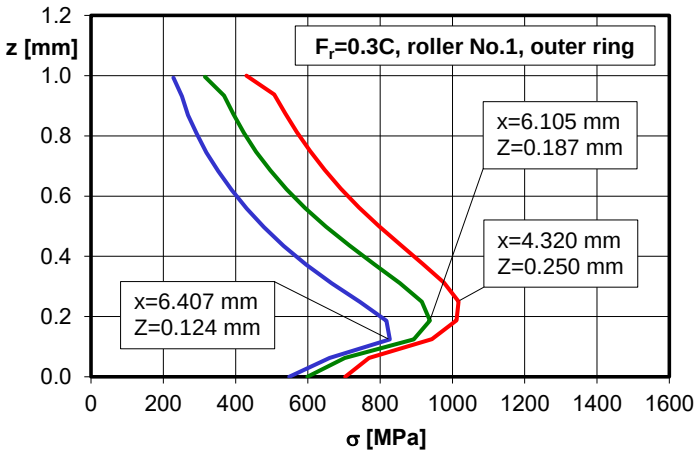


Rys. 12. Przykładowe wartości naprężenia von Misesa na głębokości materiału
Fig. 12. The example values of von Mises stresses at the material depth



Rys. 13. Rozkład naprężeń von Misesa w trzech przykładowych przekrojach styku wałeczka z bieżnią pierścienia wewnętrznego

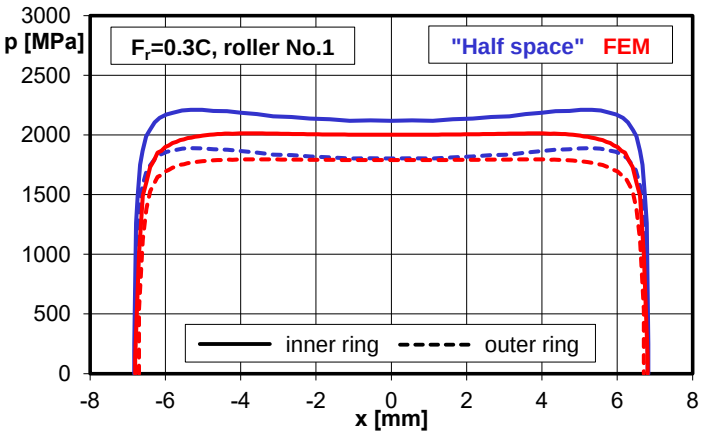
Fig. 13. Distribution of von Mises stresses in the three example cross-sections of the roller-inner ring raceway contact



Rys. 14. Rozkład naprężeń von Misesa w trzech przykładowych przekrojach styku wałeczka z bieżnią pierścienia zewnętrznego

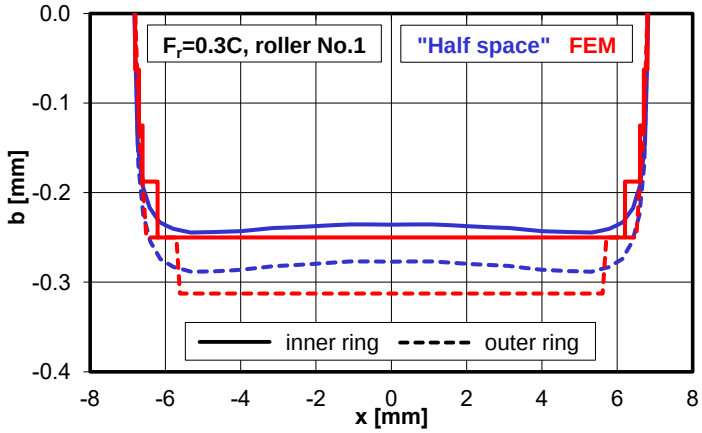
Fig. 14. Distribution of von Mises stresses in the three example cross-sections of the roller-outer ring raceway contact

Na rysunku 15 przedstawiono porównanie rozkładów maksymalnych nacisków powierzchniowych wzdłuż osi *x* najbardziej obciążonego wałeczka (nr 1) w kontakcie z wewnętrzną i zewnętrzną bieżnią obu pierścieni, uzyskanych za pomocą programu ROLL4 (metoda „half space") oraz MES. Rysunek 16 pokazuje granice obszaru styku wałeczka z obydwoma bieżniami.



Rys. 15. Rozkłady nacisków powierzchniowych wzdłuż osi *x* dla najbardziej obciążonego wałeczka w styku z bieżniami wewnętrznego i zewnętrznego pierścienia

Fig. 15. Pressure distributions along the *x* axis of the most heavily loaded roller in contact with the inner and outer ring raceway

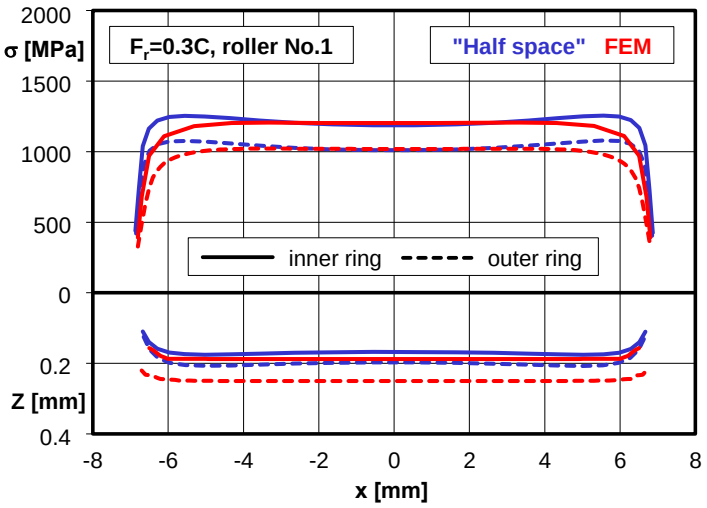


Rys. 16. Granice obszaru styku wałeczka z bieżniami

Fig. 16. Contact areas of a roller with raceways

Wyniki uzyskane przy użyciu metody „half space" i MES są wystarczająco zbieżne. Kształt i wielkość strefy kontaktu są podobne. Jedyną różnicę w rozkładach nacisków powierzchniowych obserwuje się tylko przy końcach pola styku. Wynika to z niewielkich nacisków powierzchniowych wywieranych w tym punkcie i ma niewielki wpływ na oszacowanie trwałości zmęczeniowej.

Rysunek 17 pokazuje naprężenia σ i głębokości Z w przypadku najbardziej obciążonego wałeczka w styku z bieżnią wewnętrzną i zewnętrzną pierścieni łożyska, uzyskane za pomocą obu metod.

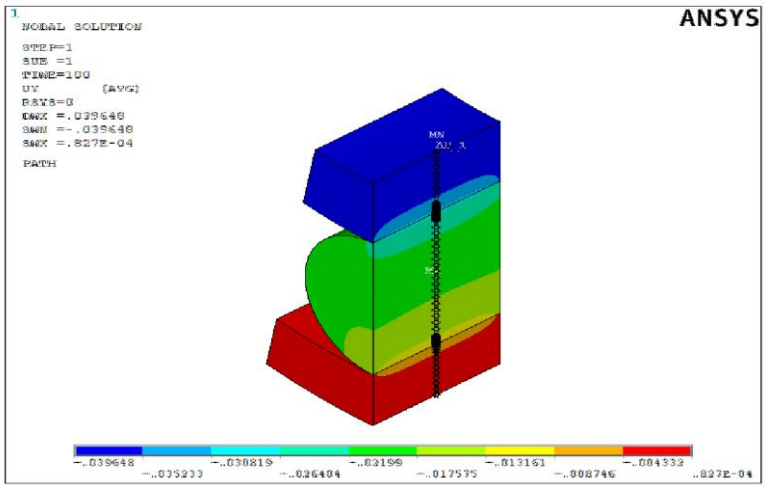


Rys. 17. Rozkład maksymalnych naprężeń von Misesa σ i głębokości ich występowania Z dla najbardziej obciążonego wałeczka w stykach z bieżniami pierścienia wewnętrznego i zewnętrznego

Fig. 17. Distributions of maximal von Mises stresses σ and the depth of their occurrence Z in the most heavily loaded roller in contact with the inner and outer ring raceway

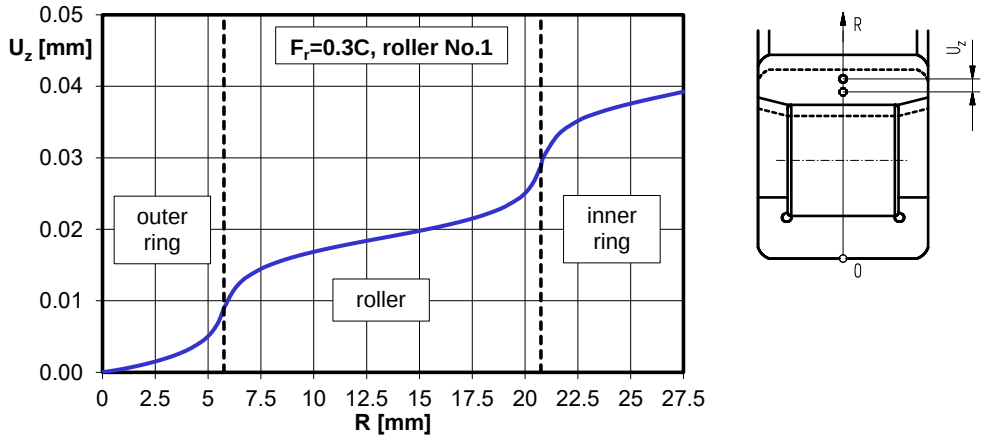
Krzywe opisujące głębokość, na której występują maksymalne naprężenia von Misesa, otrzymane zarówno metodą „half space", jak i przez MES, mają kształt wystarczająco zbliżony. Rozkłady głębokości Z otrzymane wg MES charakteryzują się nieco wyższymi wartościami w porównaniu z wynikami otrzymanymi metodą „half space". Różnice te wynikają z modelu zastosowanego w MES, który, w przeciwieństwie do metody „half space", bierze pod uwagę kontakt rolki z pierścieniem wewnętrznym i zewnętrznym.

Program ANSYS pozwala również uzyskać mapę przemieszczeń wewnątrz stykających się ciał. Na rysunkach 18 i 19 pokazano przemieszczenie współdziałających elementów łożyskowych pod wpływem przyłożonego obciążenia. Rysunek 19 pokazuje, że największe przemieszczenia występują w pobliżu styków wałeczka z bieżniami.



Rys. 18. Przemieszczenie elementów badanego łożyska

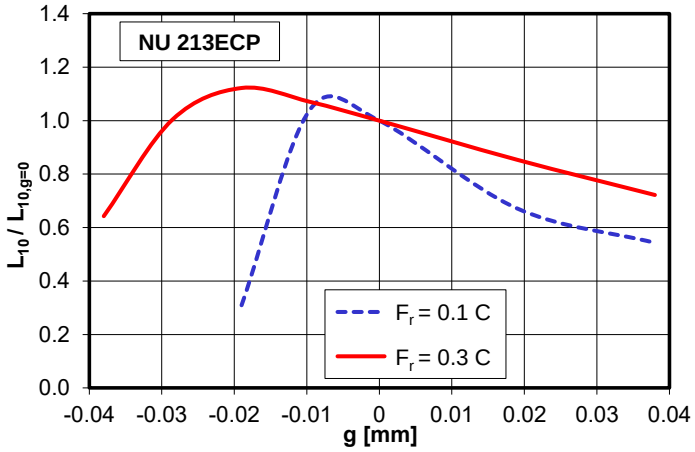
Fig. 18. Displacement of radial cylindrical roller bearing elements



Rys. 19. Lokalne przemieszczenie U_z wewnątrz elementów łożyska
Fig. 19. Local displacement U_z inside the bearing elements

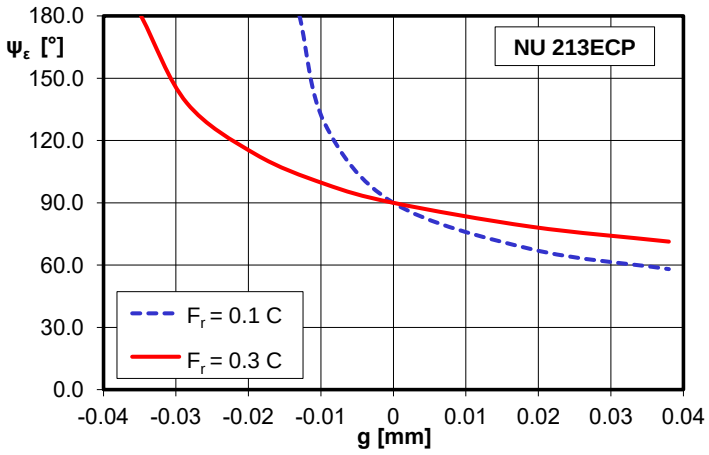
5.3. Trwałość zmęczeniowa

Ze względu na długi czas potrzebny do rozwiązania zagadnienia kontaktowego za pomocą MES, do obliczenia przewidywanej trwałości zmęczeniowej łożyska wykorzystano jedynie metodę „half space". Na rysunku 20 przedstawiono charakterystykę względnej trwałości zmęczeniowej łożyska, odniesionej do trwałości przy luzie promieniowym równym zeru, w funkcji luzu promieniowego. Przebieg krzywych jest potwierdzeniem znanego faktu, że łożysko osiąga najwyższą wytrzymałość przy niewielkim wcisku [12].



Rys. 20. Względna trwałość łożyska w funkcji luzu promieniowego
Fig. 20. Relative bearing life as a function of radial clearance

W przypadku obciążenia promieniowego $F_r = 0{,}3\,C$ trwałość zmęczeniowa badanego łożyska osiąga najwyższe wartości (tj. większe niż wytrzymałość dla $g = 0$) przy luzie promieniowym z zakresu od $g = 0$ do $g = -0{,}029$ mm. W przypadku obciążenia promieniowego $F_r = 0{,}1\,C$ największe trwałości zmęczeniowe osiąga łożysko z luzem promieniowym z zakresu od $g = 0$ do $g = -0{,}011$ mm. W obu przypadkach odpowiada to kątom strefy obciążenia z zakresu $90° \le \psi_\varepsilon \le 140°$ (Rys. 21).



Rys. 21. Kąt obciążonej strefy łożyska w funkcji luzu promieniowego
Fig. 21. Angle of the loaded zone as a function of radial clearance

Największy wzrost trwałości zmęczeniowej w stosunku do trwałości zmęczeniowej przy zerowym luzie promieniowym, który może osiągnąć łożysko, jest równy $L_{10} / L_{10,g=0} \approx 1{,}1$. Ten wzrost występuje przy kącie $\psi_\varepsilon \approx 115°$, niezależnie od wartości obciążenia promieniowego. Jednak wraz ze wzrostem obciążenia rośnie wielkość wcisku, przy którym uzyskuje się największy wzrost trwałości zmęczeniowej. W przypadku luzów promieniowych większych od zera i wcisków o wartościach większych niż wartości podane powyżej, trwałość zmęczeniowa badanego łożyska walcowego jest mniejsza niż trwałość zmęczeniowa przy $g = 0$. Przyczyny tego zjawiska wyjaśniono na rysunku 7, przedstawiającym rozkłady obciążenia dla różnych luzów promieniowych. Jak pokazano, w zakresie luzów od $g = 0$ do $g = -0{,}029$ mm, liczba wałeczków poddanych obciążeniu jest większa niż w przypadku luzu promieniowego $g > 0$, ale jednocześnie wartości sił działających na najbardziej obciążone wałeczki, które determinują trwałość zmęczeniową łożyska, są niższe. W przypadku luzów promieniowych mniejszych niż $g = -0{,}029$ mm i większych niż $g = 0$ wartości sił rosną, co skutkuje spadkiem trwałości zmęczeniowej.

6. Wnioski

Zastosowana metodyka pozwala przewidzieć trwałość dowolnego promieniowego łożyska walcowego dla zadanego obciążenia i luzu promieniowego poprzez wprowadzenie parametrów do obliczeń, które nie mogły być uwzględnione podczas wykonywania obliczeń analitycznych. Jednym z nich jest korekcja kształtu tworzących wałeczków łożyska.

Zaprezentowana metodyka umożliwiła określenie rozkładu obciążenia promieniowego na elementy toczne przy założonym luzie promieniowym. Pozwala również obliczyć stałą Stribeck'a.

W wyniku obliczeń numerycznych uzyskano dane niezbędne do analizy zjawisk zachodzących w strefie kontaktu pary kinematycznej badanego modelu. W przypadku badania łożysk tocznych konieczne było zbudowanie złożonych modeli bryłowych współpracujących elementów. Znajomość rozkładów naprężeń podpowierzchniowych jest konieczna do określenia trwałości zmęczeniowej łożysk tocznych.

Zastosowanie MES pozwoliło na otrzymanie dokładnych informacji o stanie naprężeń podpowierzchniowych w stykach. Podobny wynik można uzyskać rozwiązując zagadnienie Boussinesqa dla półprzestrzeni sprężystej.

Pakiet ANSYS pozwala określić rozkłady naprężeń von Misesa w płaszczyźnie wybranej przez użytkownika. Możliwość określenia rozkładu naprężenia von Misesa w przekroju zdefiniowanym przez dowolną współrzędną *x* jest szczególnie ważna dla przewidywania trwałości zmęczeniowej promieniowych łożysk walcowych poddanych złożonemu obciążeniu. Złożone obciążenie powoduje znaczne pochylenie elementów tocznych łożyska, co wymaga znajomości dokładnej pozycji maksymalnego naprężenia w materiale.

Przeprowadzone analizy wykazały, że promieniowe łożysko walcowe osiąga największą trwałość przy niewielkim ujemnym luzie promieniowym (wcisku). Największy wzrost trwałości zmęczeniowej w stosunku do trwałości zmęczeniowej przy zerowym luzie promieniowym łożysko osiąga przy luzie, którego kąt obciążonej strefy jest równy $\psi_\varepsilon \approx 115°$, niezależnie od wartości obciążenia promieniowego. Wzrost trwałości zmęczeniowej w tym przypadku jest równy $L_{10} / L_{10,g=0} \approx 1{,}1$.

Literatura

1. Deshpande H., Kulkarni S, Gandhare B S. Investigation on effect of defect on cylindrical roller bearing, by experimental and FEA approach. International Journal of Emerging Technology and Advanced Engineering 2014;4(6).

2. Hamrock B J, Anderson W J. Rolling-element bearings. NASA RP_1105/REV1; 1983.

3. Harris T A. Rolling bearing analysis. Wiley-Interscience, New York, USA; 1991.

4. Harris T A, Kotzalas M N. Advanced concepts of bearing technology,: Rolling bearing analysis, Fifth edition. CRC Press; 2006.

5. Harris T A, Kotzalas M N. Essential concepts of bearing technology, Fifth edition. CRC Press; 2006.

6. Harris T A, Kotzalas M N. Rolling Bearing Analysis, Fifth Edition - 2 Volume Set. CRC Press; 2007.

7. Hertz H. Uber die Beruhrung fester elastischer Korper. Gesammelte Werke (P. Lenard, ed.), Bd. 1, (J.A. Barth, Leipzig; 1895):155-173.

8. Jones A. A general theory for elastically constrained ball and roller bearing under arbitrary load and speedy conditions. Trans ASME 1960;105:591-595.

9. Jonson K L. Contact mechanics. Cambrige University Press, Cambrige; 1985.

10. Kania L. Modelling of rollers in calculation of slewing bearing with the use finite elements. Mech Mach Theory 2006;41:1359-76.

11. Krzemiński-Freda H. Correction of the generators of the main working surfaces of roller bearings. Arch Mech Eng 1990;37:115–132.

12. Krzemiński-Freda H. Roller Bearing. PWN, Warsaw; 1985 (in polish).

13. Laniado-Jacome E. Numerical model to study of contact force in a cylindrical roller bearing with technical mechanical event simulation. J Mech Eng Autom 2011;1:1–7.

14. Lundberg G. Cylinder compressed between two plane bodies. SKF Reg. 4134 1949.

15. Lundberg G, Palmgren A. Dynamic capacity of rolling bearings. Acta Polytech Scand Mech Eng 1947;1(3):1-52.

16. Lundberg G, Palmgren A. Dynamic capacity of roller bearings. Acta Polytech Scand Mech Eng 1952;2(4):96-127.

17. Palmgren A. Ball and roller bearing engineering. Third ed., SKF Industries, Philadelphia, PA 1959.

18. Ricci M C. Internal loading distribution in statically loaded ball bearings subjected to a centric thrust load: alternative approach. World Academy of Science, Engineering and Technology 2010;65:641-649.

19. Shah Maulik J, Darji P H. Fatigue life improvement through reduction of edge pressure in cylindrical roller bearing using FE analysis. International Journal For Technological Research in Engineering 2014;1(10):1069-74.

20. Shaha Rohit D, Kulkarni S S. Vibration analysis of deep groove ball bearing using finite element analysis. International Journal of Engineering Research and Applications 2015;5(5):44-50.

21. Shingala Niraj R, Sata Ankit V, Delvadiya Parth V, et al. Contact stress analysis of needle roller bearing used in synchromesh gear box. Trends in Machine Design 2018;5(1):5–20.

22. SKF General Catalogue; 2007.

23. Stribeck R. Ball bearings for various loads. Reports from the Central Laboratory for Scientific Technical Investigation, translation by K.W. Van Treuren et al., Trans ASME 1907;29:420-463.

24. Tang Zhaoping, Sun Jianping. The contact analysis for deep groove ball bearing based on Ansys. Procedia Engineering 2011;23:423-428.

25. Warda B, Chudzik A. Fatigue life prediction of the radial roller bearing with the correction of roller generators. International Journal of Mechanical Science 2014;89:299–310.

26. Warda B, Chudzik A. Effect of ring misalignment on the fatigue life of the radial cylindrical roller bearing. International Journal of Mechanical Science 2016;111-112: 1-11.

Dr hab. Leszek Knopik, prof. nadzw. UTP

Faculty of Management
UTP University of Science and Technology
Fordońska Street 430, 85-890 Bydgoszcz, Poland
e-mail: knopikl@utp.edu.pl

Dr hab. inż. Klaudiusz Migawa, prof. nadzw. UTP

Faculty of Mechanical Engineering
UTP University of Science and Technology
Al. prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland
e-mail: klaudiusz.migawa@utp.edu.pl

Semi-markowski model systemu obsługi z minimalną naprawą

Semi-Markov system model for minimal repair maintenance

Słowa kluczowe: *wymiana według wieku, naprawa minimalna, naprawa dokładna, zysk na jednostkę czasu, gotowość, procesy semi-Markowa, obsługa prewencyjna, obsługa korekcyjna*

Keywords*: age-replacement, minimal repair, perfect repair, profit per unit time, availability, semi-markov processes, preventive maintenance, corrective maintenance*

Streszczenie: W pracy bada się semimarkowski model wymian według wieku obiektów technicznych. W modelu uwzględnia się dwa rodzaje napraw: naprawy dokładne i naprawy minimalne. Naprawy minimalne w modelach semimarkowskich były badane w literaturze w niewielkim stopniu. Jako kryteria jakości pracy systemu rozważa się asymptotyczny współczynnik gotowości i zysk przypadający na jednostkę czasu. W pracy sformułowano różne warunki istnienia maksimum funkcji kryterialnych. Podane na końcu pracy dwa przykłady numeryczne ilustrują wyniki uzyskane w pracy.

Abstract: This paper analyzes the semi-Markov model of technical objects age-replacements. The model includes two types of repairs: perfect repairs and minimal repairs. Minimal repairs in semi-Markov models have been studied in the literature only to an extent. In this paper, the asymptotic availability coefficient and profit per unit of time are considered as criteria for the quality of the system operation. The paper formulates various conditions for the existence of the maximum of criteria functions. The two numerical examples given at the end of the paper illustrate the results obtained in the paper.

1. Wstęp

Systemy techniczne poddawane są procesom degradacji w czasie długiego okresu użytkowania. Powstające związku z tym uszkodzenia mają negatywny wpływ na bezpieczeństwo i przychody z pracy systemu. Uszkodzenia mogą powodować następne uszkodzenia niekorzystne dla systemu. W celu zmniejszenia liczby uszkodzeń obiektów technicznych do zarządzania systemami wprowadza się różne strategie działań profilaktycznych. Powstaje problem jak zmniejszyć koszty utrzymania systemów. Wymaga to opracowania skutecznych strategii napraw i wymian. Zarządzanie wymianami i naprawami w systemach przemysłowych wymaga wprowadzania do systemu różnych działań związanych z utrzymaniem odpowiedniego poziomu niezawodności i gotowości systemu. Działania te dzielą

się na dwa rodzaje: profilaktyka prewencyjna (preventive maintenance PM) i naprawy (corrective maintenance CM). Działanie corrective maintenace w praktyce jest prowadzone w dwóch wariantach: po naprawie system jest „dobry jak nowy" (perfect repair) oraz "zły jak stary" (minimal repair). Minimalna naprawa przywraca system do jego stanu niezawodnościowego tuż przed uszkodzeniem. W praktyce te działania przywracają system do stanu pośredniego pomiędzy tymi dwoma możliwymi ekstremalnymi przypadkami. Stan powstały w wyniku tej działalności jest określany jako niedokładne utrzymanie (imperfect maintenance). Różnorodne modele niedokładnego utrzymania szczegółowo przedstawiono w pracach przeglądowych [12, 13].

Zmniejszenie kosztów utrzymania systemów realizuje się przez wdrożenie różnych skutecznych strategii działań prewencyjnych i napraw. Do tych czynności wliczamy wymiany ważnych elementów systemu i ustalenie częstości przeglądów. Harmonogram tych czynności jest często ustalany przez projektanta sytemu lub producenta. Dział utrzymania ruchu też decyduje o wymianach zużytych elementów. Zawsze działania (CM) wymagają uprzedniej diagnozy uszkodzenia, jego identyfikacji, dlatego są kosztowne i wymagają wysokich umiejętności personelu. Koszty napraw (CM) są na ogół wyższe od kosztów profilaktyki prewencyjnej (PM). Podobnie średnie czasy napraw są większe od średnich czasów profilaktyki prewencyjnej. W przypadku niektórych systemów przemysłowych możliwe jest również naprawienie uszkodzonego komponentu bez jego wymiany. Ten rodzaj naprawy można uznać za minimalną naprawę (MR). Minimalna naprawa przywraca uszkodzony obiekt do stanu przed wystąpieniem uszkodzenia. Z tego punktu widzenia niektóre wymiany można uznać za minimalne naprawy. Opierając się na tej argumentacji w literaturze zaproponowano dużo praktycznych modeli wymian z minimalną naprawą. Z tych powodów opracowanie różnych strategii działań prewencyjnych, które proponują optymalne modele decyzyjne w celu zmniejszenia kosztów utrzymania systemu i zmniejszenia ryzyka zdarzeń niepożądanych jest ważnym tematem badawczym w inżynierii niezawodności. W ostatnich czterech dekadach modele utrzymania prewencyjnego wzbudziły rosnące zainteresowanie badaniami niezawodności systemów.

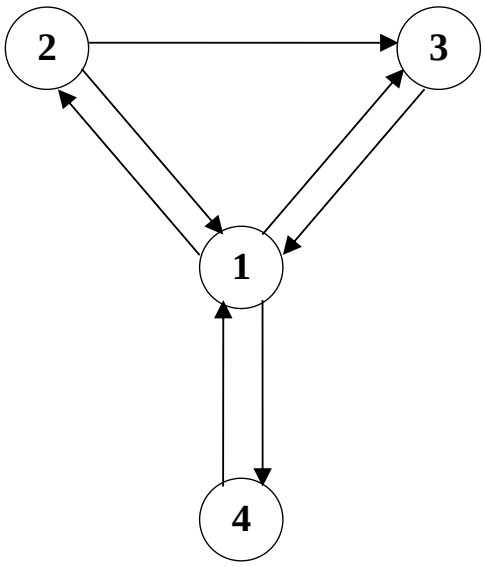
Pojęcie minimalnej naprawy wprowadzili Brown i Prochan w pracy [2]. W modelu minimalnej naprawy zakłada się, że gdy nastąpi uszkodzenie, to z prawdopodobieństwem p wykonuje się naprawę dokładną i z prawdopodobieństwem 1 – p naprawę minimalną. Naprawa dokładna przywraca obiekt techniczny do stanu „dobry jak nowy”. Jeśli p = 0, to naprawa jest zawsze minimalna, natomiast jeśli p = 1, to naprawa jest zawsze dokładna. Pham i Wang w pracy [13] nazwali taki mechanizm naprawy niedokładnym modelem utrzymania z regułą (p, q). W pracy [2] zakłada się, że prawdopodobieństwo naprawy dokładnej zależy od wieku obiektu technicznego w momencie awarii. W literaturze budowa modelu minimalnych napraw jest dokonywana różnymi metodami matematycznymi. Przegląd wykorzystywanych metod konstrukcji funkcji kryterialnych w modelach minimalnych napraw z utrzymaniem prewencyjnym według wieku jest w pracach [12, 13]. Jednak tylko jedna cytowana tam praca [3] wykorzystuje procesy semi-Markowa. Nowszy przegląd prac dotyczących minimalnych napraw zawiera książka [15]. Niedawno uzyskano nowe wyniki dotyczące minimalnych napraw w pracach [4, 5, 16, 17]. Ekonomiczny punkt widzenia na problem minimalnych napraw rozważa się w pracy [6]. Dodatkowo ta praca zawiera aktualny i obszerny przegląd literatury dotyczący minimalnych napraw i wymian prewencyjnych.

W tej pracy rozważa się strategię utrzymania systemu przy zastosowaniu (p, q) reguły wymian według wieku. Analizuje się możliwość zastosowania procesów semi-Markowa do budowy modelu wymian profilaktycznych w systemach z minimalną naprawą. Podstawą do budowania funkcji kryterialnej jest pewne twierdzenie graniczne dla procesów semi-Markowa [7, 8]. Takie podejście do budowy funkcji kryterialnej zastosowano w pracach [11, 12]. Wyniki pracy [3] są szczególnym przypadkiem wyników uzyskanych dla modelu 3-stanowego w pracy

[11] i tej pracy. W tej pracy w przeciwieństwie do większości prac z utrzymania systemów czasy napraw nie są pomijalne. Jako funkcję kryterialną w pracy bada się zysk na jednostkę czasu i współczynnik gotowości systemu. Sformułowano warunki dostateczne istnienia jednoznacznego maksimum obu funkcji kryterialnych. W rozdziale 2 zdefiniowano model 4-stanowy wymian z minimalnymi naprawami oraz określa się funkcję kryterialną jako zysk przypadający na jednostkę czasu. Rozdział 3 zawiera warunki dostateczne istnienia maksimum zysku na jednostkę czasu oraz maksimum współczynnika gotowości. W rozdziale 4 analizuje się dwa przykłady numeryczne pokazujące wyniki uzyskane w pracy. W pierwszym przykładzie maksymalizuje się współczynnik gotowości, natomiast w drugim zysk na jednostkę czasu. W obu przykładach założono, że czas do uszkodzenia ma rozkład Weibulla.

2. Funkcja kryterialna

W pracy bada się system, w którym obiekt techniczny może się znajdować w jednym z czterech stanów: S_1 – stan pracy bezawaryjnej, S_2 – stan minimalnej naprawy, S_3 – stan dokładnej naprawy, S_4 – stan wymiany prewencyjnej. Możliwe zmiany stanów pokazuje graf przedstawiony na rysunku 1.



Rys. 1. Graf skierowany zmian stanów S = { S_1, S_2, S_3, S_4 }
Fig. 1. Directed graph for changes of states S = { S_1, S_2, S_3, S_4 }

W przypadku, gdy znamy prawdopodobieństwa przejścia między stanami mamy określony łańcuch Markowa. Macierz przejść łańcucha Markowa ma postać

$$P = \begin{bmatrix} 0 & p_{12} & p_{13} & p_{14} \\ p_{21} & 0 & p_{23} & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Rozwiązując odpowiedni układ równań liniowych otrzymuje się prawdopodobieństwa graniczne dla łańcucha Markowa. Analizowany łańcuch posiada następujące prawdopodobieństwa graniczne:

$$
\begin{gathered}
p_1^* = 1 / M, \\
p_2^* = p_{12} / M, \\
p_3^* = (p_{13} + p_{12} p_{23}) / M, \\
p_4^* = p_{14} / M, \\
\text{gdzie } M = 2 + p_{12} p_{23}.
\end{gathered} \tag{1}
$$

W pracy bada się semimarkowski model napraw i wymian prewencyjnych według wieku. Rozważa się 4–stanowy proces semimarkowski X(t) z przestrzenią stanów S = {1, 2, 3, 4}. Jeśli X(t) = i, to rozważany obiekt techniczny w chwili t jest w stanie i. Przez z_i, i = 1, 2, 3, 4 oznacza się zysk na jednostkę czasu dla stanu i. W pracy zakłada się, że $z_1 > 0$, $z_i \le 0$ dla $2 \le i \le 4$. Jeśli obiekt techniczny jest w stanie 1, to przynosi zysk, natomiast jeśli obiekt techniczny jest w stanie i, gdzie $2 \le i \le 4$, to obiekt techniczny generuje stratę.

Jednostka jest wymieniana w wieku T lub gdy jest uszkodzona, cokolwiek nastąpi wcześniej. Przez $T_1(x)$ definiuje się czas do wymiany lub uszkodzenia. Zmienną $T_1(x)$ można zapisać jako

$$
T_1(x) = \begin{cases} T_1, & \text{if} \quad T_1 < x, \\ x, & \text{if} \quad T_1 \ge x. \end{cases} \tag{2}
$$

W pracy [11] udowodniono, że zysk przypadający na jednostkę czasu wyraża się wzorem

$$
L = \frac{\sum_{i=1}^{4} z_i p_i^* ET_i}{\sum_{i=1}^{4} p_i^* ET_i}, \tag{3}
$$

gdzie ET_i, i = 1, 2, 3, 4 jest średnim czasem przebywania obiektu technicznego w stanie S_i.

Zakłada się, że po upływie czasu x, jeśli obiekt nie uszkodził się, to przechodzi do stanu prewencji (wymiany). Proces zmian stanów s_i, i = 1, 2, 3, 4, przy uwzględnieniu wymiany prewencyjnej po czasie x jest nowym procesem semi-Markowa z macierzą P(x) prawdopodobieństw przejścia włożonego łańcucha Markowa. W stosunku do macierzy P opisanej wyżej zmienia się tylko pierwszy wiersz macierzy P. W szczególności na podstawie pracy [11] można napisać:

$$
\begin{gathered}
p_{12}(x) = p_{12} F_{12}(x), \\
p_{13}(x) = p_{13} F_{13}(x), \\
p_{14}(x) = p_{14} F_{14}(x) + R_1(x).
\end{gathered}
$$

gdzie:
$F_{1i}(x)$, i = 1, 2, 3, 4 są dystrybuantami warunkowymi czasu przebywania w stanie i, przed przejściem do stanu j, definiowane następująco:

$$
F_{ij}(t) = P\left\{ \tau_{k+1} - \tau_k < t \,\middle|\, X(\tau_{k+1}) = j, \;\; X(\tau_k) = i \right\}, \quad \text{dla} \;\; i, j = 1,2,3,4,
$$

$R_1(x) = 1 – F_1(x)$ jest funkcją niezawodności czasu T_1.

W celu uproszczenia obliczeń zakłada się, że prawdziwe są równości:

$$F_{12}(x) = F_{13}(x) = F_{14}(x) = F_1(x).$$

Na podstawie pracy [11] funkcja kryterialna ma postać

$$g(x) = \frac{z_1 ET_1(x) p_1^*(x) + z_2 ET_2 p_2^* + z_3 ET_3 p_3^* + z_4 ET_4 p_4^*}{ET_1(x) p_1^*(x) + ET_2 p_2^* + ET_3 p_3^* + ET_4 p_4^*}. \tag{4}$$

Wartości średnią $ET_1(x)$ oblicza się ze wzoru

$$ET_1(x) = \int_0^x dF_1(t) + xP\{T_1 \geq x\}.$$

Całkując przez części, otrzymuje się

$$ET_1(x) = \int_0^x R_1(t)dt.$$

Prawdopodobieństwa graniczne $p_1^*(x)$, $p_2^*(x)$, $p_3^*(x)$ są prawdopodobieństwami dla łańcucha Markowa z macierzą

$$P(x) = \begin{bmatrix} 0 & p_{12}(x) & p_{13}(x) & p_{14}(x) \\ p_{21} & 0 & p_{23} & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Na odstawie wzorów (1) można zapisać jako:

$$\begin{gathered} p_1^*(x) = 1 / M(x), \\ p_2^*(x) = p_{12}(x) / M(x), \\ p_3^*(x) = (p_{13}(x) + p_{12}(x)\, p_{23}) / M(x), \\ p_4^*(x) = p_{14}(x) / M(x), \\ \text{gdzie } M(x) = 2 + p_{12}(x)\, p_{23}. \end{gathered} \tag{5}$$

ET_2, ET_3 i ET_4 są wartościami średnimi czasów przebywania obiektu w stanach S_2, S_3 i S_4 systemu.

Na podstawie (5) funkcja kryterialna (4) wyraża się wzorem

$$g(x) = \frac{z_1 ET_1(x) + p_{12} F_1(x) z_2 ET_2 + [p_{13} F_1(x) + p_{12} p_{23} F_1(x)] z_3 ET_3 + [1 - p_{12} F_1(x) - p_{13} F_1(x)] z_4 ET_4}{ET_1(x) + p_{12} F_1(x) ET_2 + [p_{13} F_1(x) + p_{12} p_{23} F_1(x)] ET_3 + [1 - p_{12} F_1(x) - p_{13} F_1(x)] ET_4}.$$

Po przegrupowaniu można zapisać

$$g(x) = \frac{z_1 ET_1(x) + F_1(x)[p_{12} z_2 ET_2 + p_{12} p_{23} z_3 ET_3 + p_{13} z_3 ET_3 - p_{12} z_4 ET_4 - p_{13} z_4 ET_4] + z_4 ET_4}{ET_1(x) + F_1(x)[p_{12} ET_2 + p_{12} p_{23} ET_3 + p_{13} ET_3 - p_{12} ET_4 - p_{13} ET_4] + ET_4}.$$

Teraz licznik i mianownik funkcji kryterialnej można przedstawić jako:

$$L(x) = z_1\, ET(x) + F_1(x)\, B_1 + C_1,$$

$$M(x) = ET(x) + F_1(x)\,B + C.$$

Analogicznie

$$g(x) = \frac{z_1 ET_1(x) + F_1(x) B_1 + C_1}{ET_1(x) + F_1(x) B + C},$$

gdzie:

$$B_1 = p_{12}\, z_2\, ET_2 + p_{12}\, p_{23}\, z_3\, ET_3 - p_{12}\, z_4\, ET_4 + p_{13}\, ET_3\, z_3 - p_{13}\, z_4\, ET_4,$$
$$C_1 = z_4\, ET_4,$$
$$B = p_{12}\, ET_2 + p_{12}\, p_{23}\, ET_3 - p_{12}\, ET_4 + p_{13}\, ET_3 - p_{13}\, ET_4,$$
$$C = ET_4.$$

Po prostych przekształceniach jest:

$$B_1 = p_{12}\, z_2 ET_2 + z_3 ET_3 (p_{12}\, p_{23} + p_{13}) - z_4\, ET_4 (p_{12} + p_{13}),$$
$$B = p_{12}\, ET_2 + ET_3 (p_{12}\, p_{23} + p_{13}) - ET_4 (p_{12} + p_{13}).$$

Wprowadza się oznaczenia:

$$\alpha = -B\, z_1 + B_1,$$
$$\beta = C\, z_1 - C_1,$$
$$\gamma = CB_1 - C_1 B.$$

Współczynniki α, β i γ odgrywają ważną rolę przy formułowaniu warunków dostatecznych istnienia ekstremów funkcji kryterialnych.

3. Warunki istnienia maksimum funkcji kryterialnych

Niżej formułuje się założenia na wartości parametrów badanego systemu. Założenia te muszą oddawać rzeczywiste relacje między naprawami dokładnymi, naprawami minimalnymi i wymianami prewencyjnymi:

Z1. $z_1 > 0$, $z_2 < 0$, $z_3 < 0$, $z_4 < 0$. Ostanie oznacza, że obiekt techniczny przynosi zysk jedynie w stanie S_1, pozostałe stany wymagają nakładów.

Z2. $ET_3 > ET_4$, średni czas wymiany (prewencji) jest mniejszy od średniego czasu naprawy dokładnej.

Z3. $z_3 < z_4$, koszt jednostkowy ($-z_4$) wymiany (prewencji) jest większy od kosztu jednostkowego ($-z_3$) naprawy dokładnej.

Z4. $ET_3 > ET_2$, średni czas naprawy minimalnej jest mniejszy od średniego czasu naprawy dokładnej.

Z5. $z_3 < z_2$, koszt jednostkowy ($-z_2$) naprawy minimalnej jest większy od kosztu jednostkowego ($-z_3$) naprawy dokładnej.

Powyższe założenia nie zawierają relacji między stanem minimalnej naprawy i stanem wymiany prewencyjnej. W praktyce nie wiadomo jaka jest relacja między wartościami średnimi ET_2 i ET_4 oraz z_2 i z_4. Jednak jeśli $z_2 - z_4 \leq 0$, to na podstawie założenia Z3 jest $\gamma < 0$.

Niżej formułuje się warunki dostateczne na to, aby prawdziwe były nierówności $\alpha < 0$, $\beta > 0$, $\gamma < 0$.

Powyższe warunki będą formułowane w zależności od średnich czasów ET_i, kosztów z_i, i = 1, 2, 3, 4 oraz elementów macierzy $P = [p_{ij}]$, i, j = 1, 2, 3, 4.

Stosunkowo łatwo można obliczyć, że $\beta = ET_4 (z_1 - z_4)$.

Z założenia Z1 wynika, że $\beta > 0$.

Współczynnik α wyraża się za pomocą wzoru

$$\alpha = p_{12}\, ET_2\, (z_2 - z_1) + (p_{12}\, p_{23} + p_{13})\, ET_3\, (z_3 - z_1) + ET_4\, (p_{12} + p_{13})\, (z_1 - z_4). \quad (6)$$

Nierówność $\alpha < 0$ jest równoważna nierówności

$$(p_{12}\, p_{23} + p_{13})\, ET_3 > p_{12}\, ET_2\, (z_2 - z_1) / (z_1 - z_3) + ET_4\, (p_{12} + p_{13})\, (z_1 - z_4) / (z_1 - z_3). \quad (7)$$

Podobnie wyznacza się współczynnik γ

$$\gamma = ET_2[ET_4\, p_{14}\, (z_2 - z_4) + ET_3\, (p_{12}\, p_{23} + p_{13})\, (z_3 - z_4)]. \quad (8)$$

Nierówność $\gamma < 0$ jest równoważna nierówności

$$(p_{12}\, p_{23} + p_{13})\, ET_3 > ET_4\, p_{14}\, (z_2 - z_4) / (z_4 - z_3). \quad (9)$$

Oznaczamy prawe strony nierówności (7) i (9) odpowiednio przez δ_1 i δ_2. Niech $\delta = \max\{\delta_1, \delta_2\}$. Warunek $(p_{12}\, p_{23} + p_{13})\, ET_3 > \delta$ i (6), (7), (8) i (9) implikują nierówności $\alpha < 0$, $\gamma < 0$. Można teraz podać następujący wniosek:

Wniosek 1. Jeśli $p_{23} > (\delta / ET_3 - p_{13}) / p_{12}$, to prawdziwe są nierówności $\alpha < 0$, $\gamma < 0$.

W literaturze dotyczącej minimalnych napraw [2, 13, 14], zakłada się, że jeśli obiekt techniczny przechodzi ze stanu pracy do stanu uszkodzenia, to stan minimalnej naprawy osiąga z prawdopodobieństwem równym 1 – p, natomiast stan naprawy dokładnej z prawdopodobieństwem p. Na podstawie elementarnych własności prawdopodobieństwa warunkowego wiadomo, że prawdziwa jest równość

$$p_{12} / p_{13} = (1 - p) / p. \quad (10)$$

Wniosek 2. Jeśli Tϵ IFR, $\lambda(t)$ jest różniczkowalna, $\alpha < 0$, $\gamma < 0$, $\beta > 0$, $\beta + \gamma\, f(0^+) > 0$, $\lambda(\infty)\, \alpha\, ET + \beta - \alpha < 0$, to funkcja kryterialna g(x) osiąga wartość maksymalną.
Dowód.
Pochodna g'(x) funkcji ma postać

$$g'(x) = \{\alpha[f(x)\, ET(x) - R(x)\, F(x)] + \beta\, R(x) + \gamma\, f(x)\} / M^2(x),$$

gdzie M(x) jest mianownikiem funkcji kryterialnej g(x).
Wiadomo, że jeśli czas do uszkodzenia T należy do klasy rozkładów MTFR, to prawdziwa jest równość $H(x) = \lambda(x)\, ET(x) - F(x)] \geq 0$ dla $x \geq 0$. Klasa rozkładów MTFR została zbadana w pracach [9, 10]. Do klasy MTFR należą niektóre rozkłady czasów życia z jednomodalną funkcją intensywności uszkodzeń [9, 10]. Z tego, ze pochodna $H'(x)$ ma postać $H'(x) = \lambda'(x)\, ET(x)$ wynika, że jeśli funkcja intensywności uszkodzeń $\lambda(t)$ rośnie, to funkcja H(x) też rośnie. Klasa rozkładów z niemalejącą funkcją intensywności uszkodzeń (IFR) zawiera się w klasie MTFR. Znak pochodnej jest taki sam jak znak funkcji

$$h(x) = \alpha[\lambda(x)\, ET(x) - F(x)] + \beta + \gamma\, \lambda(x).$$

Wiadomo, że $H(0^+) = 0$, stąd $h(0^+) = \beta + \gamma\, f(0^+) > 0$. Z tego, że $\alpha < 0$, $\beta > 0$, $\gamma < 0$ i funkcja H(x) rośnie wynika, że funkcja h(x) maleje od wartości $h(0^+) = \beta + \gamma\, f(0^+) > 0$ do wartości $h(\infty) = \lambda(\infty)\alpha\, ET + \beta - \alpha < 0$. Wynika stąd, że pochodna g'(x) dokładnie jeden raz zmienia znak

z „+” na „– ”. Stąd wnioskuje się, że funkcja kryterialna g(x) osiąga dokładnie jedno maksimum. □

Jeśli $\lambda(\infty) = \infty$, to do istnienia maksimum funkcji kryterialnej g(x) wystarczają warunki: Tϵ IFR, różniczkowalność $\lambda(t)$, $\alpha < 0$, $\gamma < 0$, $\beta > 0$, $\beta + \gamma f(0^+) > 0$. Przykładem takiego rozkładu jest rozkład Weibulla z rosnącą funkcją intensywności uszkodzeń.

Z wniosków 1 i 2 wynika następujący warunek dostateczny istnienia maksimum funkcji kryterialnej:

Wniosek 3. Jeśli Tϵ IFR, $\lambda(t)$ jest różniczkowalna, $\beta + \gamma \lambda(0^+) > 0$, $p_{23} > (\delta / ET_3 - p_{13}) / p_{12}$, $\lambda(\infty)\, \alpha\, ET + \beta - \alpha < 0$, to funkcja kryterialna g(x) osiąga wartość maksymalną.

Niżej formułuje się warunek dostateczny istnienia maksimum asymptotycznego współczynnika gotowości. Aby otrzymać z funkcji kryterialnej g(x) współczynnik gotowości wystarczy przyjąć następujące warunki: $z_1 = 1$, $z_2 = z_3 = z_4 = 0$. Po uwzględnieniu tych warunków we wzorze (4) otrzymuje się $B_1 = 0$, $C_1 = 0$. Stąd na podstawie (2), (3) i (5) dla α, β, γ można obliczyć:

$$\alpha = -B = -p_{12} ET_2 - (p_{12} p_{23} + p_{13}) ET_3 + (p_{12} + p_{13}) ET_4,$$
$$\beta = ET_4,$$
$$\gamma = 0.$$

Nierówność $\alpha < 0$ jest równoważna nierówności

$$p_{23} > \{[ET_4(1 + p_{13} / p_{12}) - ET_2] / ET_3\} - p_{13} / p_{12}.$$

Ostatnią nierówność przy uwzględnieniu (7) można zapisać w postaci

$$p_{23} > \{[ET_4 / (1 - p) - ET_2] / ET_3\} - p / (1 - p).$$

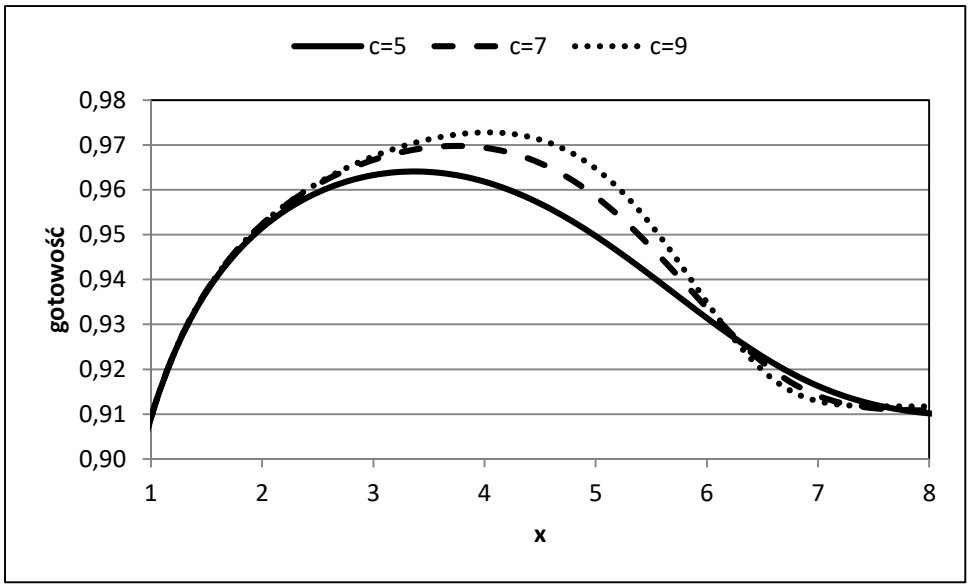
Uwzględniając, że $\beta > 0$ i $\gamma = 0$, można teraz sformułować warunek dostateczny istnienia maksimum współczynnika gotowości.

Wniosek 4. Jeśli Tϵ IFR, $\lambda(t)$ jest różniczkowalna, $\lambda(\infty)\, \alpha\, ET + \beta - \alpha < 0$, $p_{23} > \{[ET_4 / (1 - p) - ET_2] / ET_3\} - p / (1 - p)$, to współczynnik gotowości osiąga dokładnie jedną wartość maksymalną.
Dowód.
Pochodna g’(x) funkcji ma postać: $g'(x) = \{\alpha[f(x)\, ET(x) - R(x)\, F(x)] + \beta\, R(x)\} / M^2(x)$, gdzie M(x) jest mianownikiem funkcji kryterialnej g(x).
Jeśli funkcja intensywności uszkodzeń $\lambda(t)$ rośnie, to funkcja H(x) rośnie. Znak pochodnej jest taki sam jak znak funkcji $h(x) = \alpha[\lambda(x)\, ET(x) - F(x)] + \beta$. Wiadomo, że $H(0^+) = 0$, stąd $h(0^+) = \beta > 0$.
Z tego, że $p_{23} > \{[ET_4 / (1 - p) - ET_2] / ET_3\} - p / (1 - p)$, wynika, że $\alpha < 0$ i funkcja h(x) maleje od wartości $h(0^+) = \beta > 0$ do wartości $h(\infty)$. Jeśli $h(\infty) = \lambda(\infty)\alpha\, ET + \beta - \alpha < 0$, to oznacza, ze pochodna g’(x) dokładnie jeden raz zmienia znak z „+” na „–”. Stąd wnioskuje się, że współczynnik gotowości g(x) osiąga dokładnie jedno maksimum. □

Jeśli $\lambda(\infty) = \infty$, to do istnienia maksimum współczynnika gotowości wystarczają warunki Tϵ IFR, $p_{23} > \{[ET_4 / (1 - p) - ET_2] / ET_3\} - p / (1 - p)$.

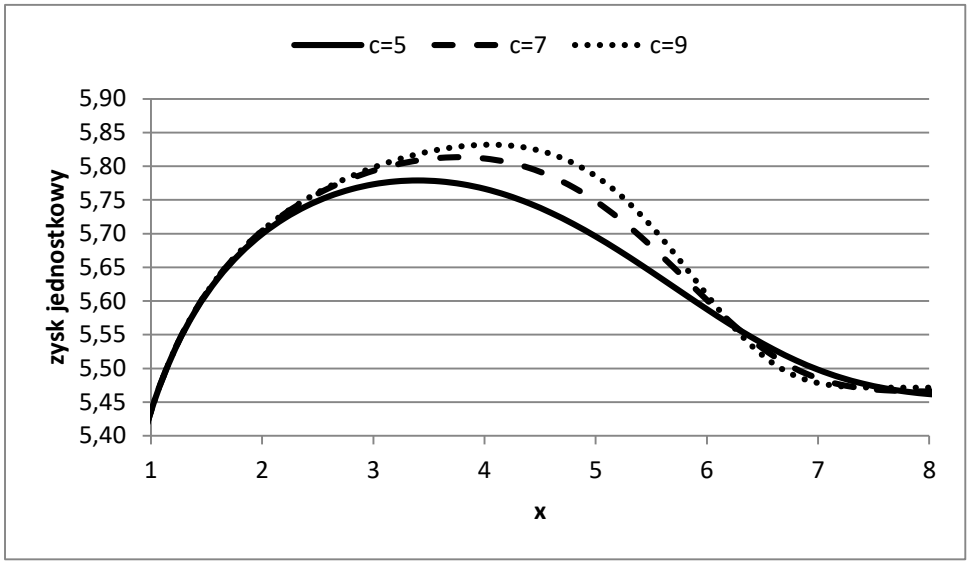
4. Przykłady numeryczne

Przykład 1. W tym przykładzie analizuje się przykład wyznaczania wartości funkcji g(x), w przypadku, gdy g(x) jest współczynnikiem gotowości. W obliczeniach przyjęto następujące dane: wartości średnie czasów przebywania obiektu technicznego w stanach $ET_2 = 0,2$, $ET_3 = 0,5$, $ET_4 = 0,1$, przyjęto dla czasu do uszkodzenia T rozkład Weibulla z parametrem skali b = 6. Przyjęto niezerowe elementy macierzy P jako $p_{12} = 0,2$, $p_{13} = 0,6$, $p_{14} = 0,2$, $p_{21} = 0,2$, $p_{23} = 0,7$. Przyjęto wartości parametru c formy (kształtu) rozkładu Weibulla $c \in \{5, 6, 7\}$. W każdym z trzech analizowanych przypadków istnieje optymalna wartość czasu wymiany.



Rys. 2. Wykresy zmian wartości współczynnika gotowości w zależności od czasu wymiany prewencyjnej x, dla $c \in \{5, 6, 7\}$
Fig. 2. Charts of changes in the value of availability coefficient depending on the time of preventive replacement x, for $c \in \{5, 6, 7\}$

Przykład 2. W tym przykładzie analizuje się przypadek wyznaczania wartości funkcji g(x), w przypadku, gdy g(x) jest zyskiem na jednostkę czasu. W obliczeniach przyjęto wartości średnie czasów przebywania w stanach, macierz prawdopodobieństw P i parametru skali rozkładu Weibulla takie same jak w przykładzie 1. Wartości parametru c formy (kształtu) rozkładu Weibulla $c \in \{5, 6, 7\}$. Do obliczeń jako zyski jednostkowe przyjęto $z_1 = 6$, $z_2 = -0,1$, $z_3 = -0,8$, $z_4 = -0,2$.



Rys. 3. Wykresy zmian wartości zysku na jednostkę czasu w zależności od czasu wymiany prewencyjnej x, dla c ϵ {5, 6, 7}
Fig. 3. Charts of changes in the value of profit per unit time coefficient depending on the time of preventive replacement x, for c ϵ {5, 6, 7)

Dla wszystkich wartości parametru c formy rozkładu Weibulla funkcja kryterialna osiąga wartość maksymalną. Analiza zależności punktu x_{max}, w którym funkcja kryterialna g(x) osiąga wartość maksymalną pokazuje, że wraz ze wzrostem wartości parametru c rośnie wartość x_{max} i maksymalna wartość funkcji kryterialnej.

6. Wnioski

Systemy utrzymania realizujące dwa rodzaje napraw: naprawy minimalne i naprawy dokładne posiadają szeroką literaturę. Jednak zastosowanie procesów semi-Markowa jest niewielkie. W tej pracy pokazano, że zastosowanie do wyznaczania optymalnych strategii działań prewencyjnych w systemach z minimalną naprawą procesów semi-Markowa pozwala na sformułowanie ciekawych wniosków. Dla analizowanych w pracy funkcji kryterialnych (gotowość i zysk na jednostkę czasu) sformułowano warunki dostateczne istnienia maksimum tych funkcji kryterialnych. Funkcje kryterialne rozważa się w niekończonym horyzoncie czasowym. Sformułowanie silniejszych warunków wymaga ustalenia relacji między średnimi czasami przebywania obiektu technicznego i zyskami jednostkowymi w stanach minimalnej naprawy i wymiany prewencyjnej.

Literatura

1. Block H W, Borges W, Savits T H. Age-dependent minimal repair. Journal of Applied Probability, 1985; 22: 370-386.

2. Brown M, Prochan F. Imperfect Repair. Journal of Applied Probability, 1983; 20: 851-859.

3. Chan P K W, Downs T. Two criteria for preventive maintenance. IEEE Transaction on Reliability, 1978; 27: 272-273.

4. Chang C C, Sheu S H, Chen Y L, Zhang Z G. A multi-criteria optimal replacement policy for a system subject to shocks. Computer and Industrial Engineering, 2011; 61: 1035-1043.

5. Chang C C, Sheu S H, Chen Y L. A bivariate optimal replacement policy for system with age-dependent minimal repair and cumulative repair-cost limit. Communications in statistics – Theory and methods, 2013; 42(22): 4108-4126.

6. Chin-Chih, Chang. Optimum preventive maintenance policies for system subject to random working time, replacement, and minimal repair. Computers & Industrial Engineering, 2014; 67: 185-194.

7. Grabski F. Semi-markowskie modele niezawodności i eksploatacji [Semi-Markov models of reliability and maintenance]. Warszawa: IBS PAN, 2002.

8. Grabski F. Semi-Markov Processes: Applications in System Reliability and Maintenance. Elsevier, Amsterdam, 2014.

9. Knopik L. Some results on ageing class. Control and Cybernetics, 2005; 34(4): 1175-1180.

10. Knopik L. Characterization of a class of lifetime distributions. Control and Cybernetics, 2006; 35(2): 1175-1180.

11. Knopik L, Migawa K. Multi-state model of maintenance policy. Maintenance and Reliability, 2018; 20(1): 125-130.

12. Knopik L, Migawa K. Optimal age-replacement policy for non-repairable technical objects with warranty. Maintenance and Reliability, 2017; 19(2): 172-178.

13. Pham H, Wang H. Imperfect maintenance. European Journal of Operational Research, 1996; 94: 425-438.

14. Pham H, Wang H. Probability and Optimal Maintenance. Springer Verlag, London, 2006.

15. Tadj L, Ouali M S, Yacount S, Ait-Kadi D. Replacement Models with minimal Repair, Chapter: A Survey of Replacement Models with Minimal Repair, Springer Verlag, London, 2011.

16. Xia, Xi, Zhou. Modeling and optimizing maintenance schedule for energy systems subject to degradation. Computer and Industrial Engineering, 2012; 63(3): 607-614.

17. Xu, Chen, Yang. Optimal replacement policy for safety-related multi-component multi-state systems. Reliability Engineering and System Safety, 2012; 99: 87-95.

dr inż. Tomasz Rymarczyk

Centrum Badawczo-Rozwojowe Nowoczesnych Technologii, Netrix S.A.
ul. Związkowa 26, 20-148 Lublin
Wyższa Szkoła Ekonomii i Innowacji w Lublinie
ul. Projektowa 4, 20-209 Lublin
E-mail: tomasz@rymarczyk.com

dr inż. Grzegorz Kłosowski

Politechnika Lubelska
Katedra Organizacji Przedsiębiorstwa na Wydziale Zarządzania
ul. Nadbystrzycka 38, 20-618 Lublin, Poland
E-mail: g.klosowski@pollub.pl

Nowatorskie metody neuronowej rekonstrukcji obrazów tomograficznych w eksploatacji zbiornikowych reaktorów przemysłowych

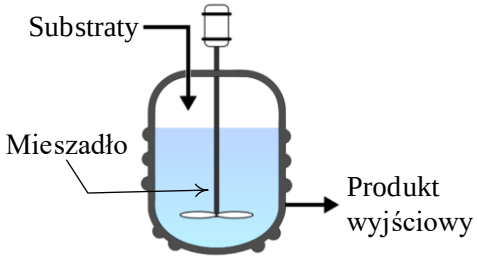
Słowa kluczowe: *tomografia elektryczna, procesy przemysłowe, sterowanie procesami, sieci neuronowe, uczenie maszynowe*

Streszczenie: W artykule przedstawiono nowatorską koncepcję usprawnienia monitoringu i optymalizacji procesów przemysłowych. Opracowana metoda bazuje na systemie osobno wytrenowanych wielu sieci neuronowych, w którym każda sieć generuje pojedynczy punkt obrazu wyjściowego. Dzięki zastosowaniu metody elastic net zaimplementowany algorytm redukuje z wejściowego wektora pomiarowego zmienne skorelowane i nieistotne, czyniąc go bardziej odpornym na zjawisko zaszumienia danych. Przewagą opisywanego rozwiązania nad znanymi metodami nieinwazyjnymi jest uzyskanie wyższej rozdzielczości obrazów dynamicznie pojawiających się wewnątrz reaktora artefaktów (kryształów lub pęcherzy gazowych), co zasadniczo przyczynia się do wczesnego wykrycia zagrożeń i problemów związanych z eksploatacją systemów przemysłowych, a tym samym zwiększa efektywność sterowania procesami chemicznymi.

1. Wprowadzenie

Wiele produktów inżynierii chemicznej powstaje jako rezultat procesów realizowanych przy użyciu ciągów technologicznych, w których kluczową rolę pełnią chemiczne reaktory przemysłowe. Reaktor chemiczny jest naczyniem przystosowanym do przeprowadzania reakcji zachodzących wewnątrz niego. Zadaniem zbiornikowych reaktorów przemysłowych jest zapewnienie optymalnych parametrów ekonomicznych procesów chemicznych [18]. Można to osiągnąć poprzez odpowiednią konstrukcję reaktora oraz poprzez umiejętne nakładanie się trzech rodzajów podprocesów zachodzących wewnątrz reaktora, a mianowicie na przenoszeniu masy, pędu i ciepła. W ten sposób sterowanie procesem można oprzeć na dynamicznym doborze parametrów, takich jak: intensywność mieszania, temperatura, ciśnienie, proporcje substratów i innych. Przedstawione badania obejmowały reaktory, w których reakcje zachodzą

pomiędzy ciałem stałym i cieczą oraz gazem i cieczą. Schemat reaktora zbiornikowego przedstawiono na Rys. 1.



Rys. 1. Schemat reaktora zbiornikowego z systemem mieszania

Pierwszy rodzaj reakcji dotyczy krystalizacji ciał stałych w otoczeniu cieczy. Obejmuje on przemysłowe procesy syntezy i oczyszczania substancji stałych oraz zmiany właściwości cząstek. Reaktory, w których następuje krystalizacja znajdują zastosowanie w wielu gałęziach gospodarki, w tym: w branży chemicznej, spożywczej [9], metalurgicznej oraz utylizacji odpadów [27]. Jednym z zadań układów monitorujących tego rodzaju procesy jest dostarczanie precyzyjnej informacji o ilości, wielkości i położeniu kryształów powstających w cieczy w czasie rzeczywistym.

Drugi rodzaj reakcji dotyczy fazy gazowej i fazy ciekłej. Procesy tego rodzaju są wykorzystywane między innymi przy produkcji biogazu. Fizykochemiczne reaktory fermentacyjne są kluczowym elementem instalacji biogazowych. Wewnątrz tych reaktorów dokonuje się fermentacja metanowa odpadów organicznych. Prawidłowa eksploatacja systemów technicznych jest jednym z kluczowych warunków uzyskania odpowiedniego poziomu niezawodności procesów przemysłowych [10].

Istnieją dwa zasadnicze powody aby monitorować stany procesów dynamicznych. Pierwszym z nich jest wykrywanie zbliżających się awarii [11], do których zaliczyć można uszkodzenie infrastruktury technicznej, ponadnormatywne odchylenie krytycznych parametrów procesu lub przerwanie jego ciągłości. Efektywny system monitoringu ma za zadanie umożliwić identyfikację problemu na tyle wcześnie, aby możliwe było podjęcie skutecznych działań korygujących.

Drugim powodem stosowania monitoringu stanów procesów przemysłowych jest konieczność sprawowania kontroli nad przebiegiem procesu przemysłowego [29]. Jest to niezbędne z punktu widzenia zapewnienia odpowiedniego poziomu jakości. Aby skutecznie sterować procesami wielofazowymi, w których uczestniczą substancje mogące dynamicznie zmieniać stany skupienia, należy zastosować efektywne metody monitoringu. To trudne zadanie, biorąc pod uwagę agresywne warunki w jakich zachodzą reakcje wewnątrz reaktora. Stosując czujniki inwazyjne, problemem jest brak możliwości bezpośredniego zbadania dowolnego fragmentu wnętrza reaktora, punktowość prowadzonych pomiarów, konieczność stosowania wielu różnych systemów monitoringu jednocześnie oraz duża niepewność w określaniu stanu dynamicznego procesu na podstawie niepełnych danych (metoda pośrednia). Wśród nieinwazyjnych metod wykorzystywanych w monitorowaniu procesów przemysłowych można wyróżnić: elektryczną tomografię pojemnościową [2, 4, 5, 13, 14, 16, 20, 26] elektryczną tomografię impedancyjną [3, 8, 28], tomografię magnetoakustyczną [30], ultradźwiękową i radiową [21], tomografię promieniami X [1] i wiele innych. Ostatnio, coraz więcej prac badawczych z zakresu eksploatacji systemów przemysłowych uwzględnia zastosowanie różnorodnych metod obliczeniowych, takich jak: inteligentne metody predykcyjne [25], logika rozmyta [6], uczenie maszynowe [22], modelowanie numeryczne [15], głębokie uczenie [8, 19] oraz programowanie binarne [12].

Obecnie stosowane techniki nieinwazyjnego monitorowania procesów przemysłowych nie w pełni zaspokajają bieżące potrzeby eksploatacyjne. Otrzymywane odwzorowania obrazów badanych zjawisk i procesów bywają nieostre, niejednoznaczne, trudne do zinterpretowania, obarczone niedokładnościami zarówno w zakresie ilości wykrytych w reaktorze artefaktów (kryształów lub pęcherzy gazu), jak również ich rozmiarów i położenia. W rezultacie, aby otrzymać precyzyjne informacje odnośnie stanu monitorowanego procesu stosuje się systemy redundantne, co znacząco podnosi koszty eksploatacji.

Wyżej wymienione utrudnienia i niedoskonałości stosowanych metod monitoringu zbiornikowych reaktorów chemicznych są powodem, z którego wynika potrzeba ich doskonalenia. Zastosowanie ulepszonej metody monitorowania spowoduje wzrost niezawodności procesów zachodzących wewnątrz reaktorów oraz obniży koszty eksploatacji systemów przemysłowych.

Celem niniejszego opracowania jest zaprezentowanie udoskonalonej metody monitoringu i optymalizacji procesów chemicznych zachodzących w zbiornikowych reaktorach heterogenicznych, w których reakcje zachodzą pomiędzy ciałem stałym i cieczą oraz gazem i cieczą. Zastosowana metoda dotyczy tomografii elektrycznej [22], a innowacją jest oryginalny sposób równoległego wykorzystania systemu hybrydowych modułów stanowiących kombinację metody elastic net i sztucznych sieci neuronowych [7, 23]. Przewagą opisywanej koncepcji nad innymi, znanymi metodami nieinwazyjnymi jest zwiększona odporność na zakłócenia powstające w trakcie pomiarów, wyższa dokładność rekonstrukcji, nieograniczona rozdzielczość obrazowania, niski koszt oraz duża szybkość działania. W dalszej części niniejszego opracowania przedstawiono opis systemu neuronowego umożliwiającego efektywny monitoring reakcji chemicznych z wykorzystaniem tomografii elektrycznej.

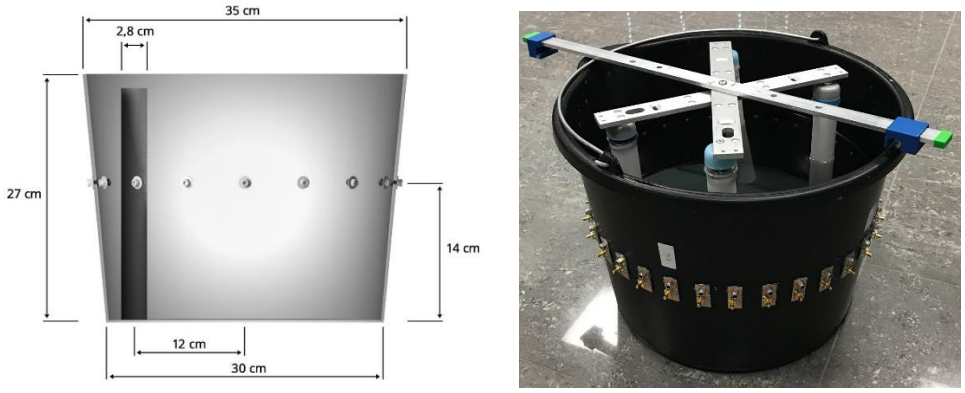
2. Modele, metody, algorytmy

Elektryczna tomografia impedancyjna (EIT – electrical impedance tomography), podobnie jak elektryczna tomografia pojemnościowa (ECT – electrical capacitance tomography) należą do nieinwazyjnych metod diagnostyki obiektów technicznych [15, 22]. Zastosowana w opisywanych badaniach metoda EIT przetwarza dane generowane przez układ 16 elektrod umieszczonych na powierzchni reaktora. Na Rys. 2 i 3 przedstawiono stanowisko badawcze, fizyczny model z układem elektrod oraz tomograf hybrydowy przystosowany do pomiarów EIT.

(a) (b)

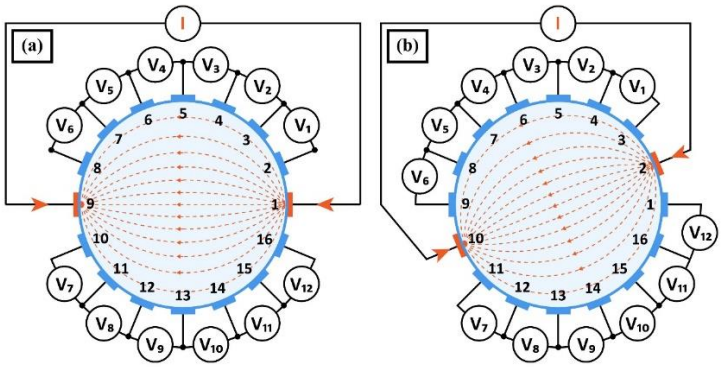
Rys. 2. Stanowisko badawcze:
a) tomograf hybrydowy, b) reaktor z podłączonymi elektrodami

(a) (b)



Rys. 3. Model fizyczny z układem elektrod EIT:
a) rysunek schematyczny, b) reaktor z zanurzonymi w cieczy artefaktami

Na Rys. 4 przedstawiono sposób pomiaru napięć generowanych za pomocą układu 16 elektrod. Z uwagi na nieznaną wartość spadków napięć między badanym obiektem a elektrodami, do których podłączone jest źródło prądu elektrycznego (I), elektrody te nie są uwzględniane podczas pomiarów. Dla każdego kąta projekcji można uzyskać 12 niezależnych pomiarów napięcia (V) pomiędzy poszczególnymi parami sąsiadujących ze sobą elektrod. Liczba ta wynika z następującego wyliczenia: n-4=12, gdzie n=16 jest całkowitą liczbą elektrod w układzie, natomiast 4 jest liczbą elektrod wyłączonych z pomiarów w ramach kąta projekcji. W ten sposób, całkowita liczba pomiarów wynosi $(n-4)(n/2) = 12 \times 8 = 96$.

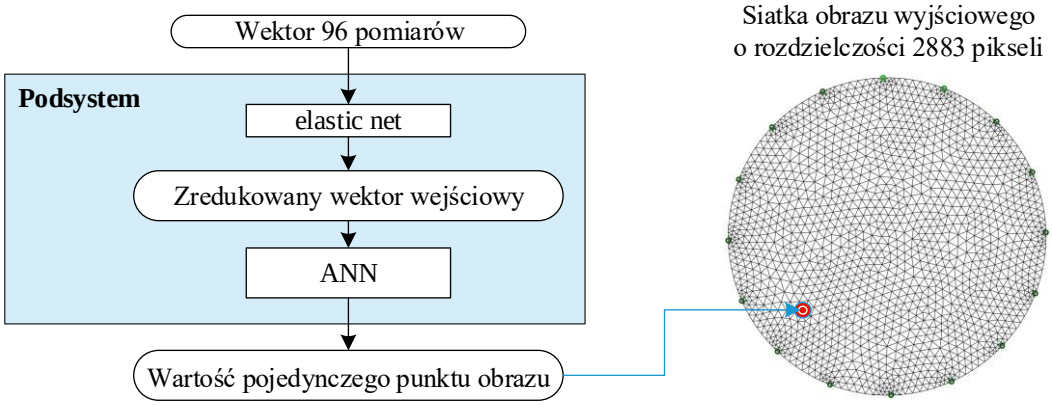


Rys. 4. Sposób pomiaru napięć w układzie 16 elektrod:
a) pierwszy cykl pomiarowy, b) kolejny cykl pomiarowy

2.1. Hybrydowy algorytm neuronowy

Dynamiczny postęp w dziedzinie tanich rozwiązań technologicznych i dostęp do zaawansowanych metod obliczeniowych sprawia, że koszty wykorzystania mocy obliczeniowej i nośników pamięci spadają [16]. Rośnie popularność technik obliczeniowych wykorzystujących obliczenia równoległe i wymagających przetwarzania dużych zbiorów danych [3]. Dzięki temu, coraz rzadziej mają miejsce sytuacje, w których badacze i projektanci systemów tomograficznych są zmuszeni skracać czas obliczeń kosztem jakości wyników [17]. Prezentowany system tomograficzny przetwarza wektor 96 wartości napięć na obraz o rozdzielczości 2883 pikseli. Ponieważ zaprojektowany algorytm wykorzystuje 2883 równolegle działające sieci neuronowe, moc obliczeniowa jest istotnym czynnikiem determinującym szybkość działania układu pomiarowego. Na Rys. 5 przedstawiono model

jednego z 2883 hybrydowych podsystemów do generowania wartości rzeczywistej pojedynczego punktu rekonstruowanego obrazu.



Rys. 5. Model jednego z 2883 podsystemów generujący pojedynczy punkt obrazu

Algorytm trenowania hybrydowego systemu neuronowego ma następujący przebieg:

1. ustal warunki początkowe:
   a. liczba pikseli na siatce obrazu wyjściowego: m=2883;
   b. liczność wektora pomiarów $\boldsymbol{X_n}$ dla każdej z rekonstrukcji: n=96;
   c. liczba przypadków uczących potrzebnych do wytrenowania sieci neuronowej: N=99900;
   d. struktura sieci neuronowej: $\tilde{n}_\upsilon$-10-1, gdzie $\tilde{n}_\upsilon$ oznacza zredukowaną metodą elastic net liczbę pomiarów na wejściu dla sieci neuronowej generującej υ-ty piksel, 10 neuronów w warstwie ukrytej oraz 1 neuron w warstwie wyjściowej. W obu warstwach zastosowano logistyczne funkcje transferowe;
2. wykorzystując metodę elastic net wygeneruj macierz indywidualnie zredukowanych wektorów pomiarowych dla każdego piksela obrazu wyjściowego osobno: $\boldsymbol{X_n} \rightarrow elastic\ net \rightarrow \boldsymbol{X_{\tilde{n},\upsilon}}$, gdzie $n = 96$, $n > \tilde{n} > 0$. Można do tego celu użyć podzbioru o liczności znacznie mniejszej niż wynosi N, np. 3000 losowo wybranych przypadków ze zbioru uczącego;
3. dla υ=1 do m trenuj sieć neuronową ANN_υ z wykorzystaniem zbioru uczącego o liczności N;
4. zapisz wytrenowane 2883 sieci neuronowe do zmiennej strukturalnej.

Powyższy algorytm został zaimplementowany za pomocą języka Matlab, natomiast siatkę obrazu wyjściowego zamodelowano przy użyciu pakietu Eidors. W Tabeli 1 przedstawiono rezultaty procesu uczenia jednej z 2883 sieci neuronowych, generujących liczbę rzeczywistą determinującą kolor przykładowego piksela na siatce obrazu tomograficznego. Do poniższego testu wybrano punkt nr 600. Dane wejściowe zawarto w macierzy $\boldsymbol{X_n}$ o wymiarach 96x99900 (96 pomiarów, 99900 przypadków). Zbiorem wyjściowym był wektor $\boldsymbol{Y_\upsilon}$ o wymiarach 1x99900 (1 punkt obrazu υ=600, 99900 przypadków). Przed wytrenowaniem sieci, zredukowano wektor wejściowy z $n = 96$ do $\tilde{n} = 30$ stosując metodę elastic net. Struktura sieci neuronowej dla punktu nr 600 była więc następująca: ANN_{600}=30-10-1.

Wygenerowany symulacyjnie zbiór 99900 danych został podzielony na 3 części: uczącą, walidacyjną i testową w proporcjach 70/15/15, co znajduje swoje odzwierciedlenie w pierwszych dwóch kolumnach Tabeli 1. Kolumny 3 i 4 zawierają informacje o błędzie MSE oraz regresji R dla wszystkich 3 zbiorów, jakie zostały osiągnięte w procesie uczenia sieci. Niska wartość MSE i R zbliżona do 1 świadczą o dobrej jakości wytrenowanej sieci.

Tabela 1. Rezultaty procesu uczenia wraz z podziałem danych

Podział zbioru danych	**Liczba przypadków w danym zbiorze**	**Błąd średniokwadratowy (MSE)**	**Regresja (R)**
1	*2*	*3*	*4*
Zbiór uczący (70%)	69930	$7.65053 \cdot 10^{-3}$	0.813877
Zbiór walidacyjny (15%)	14985	$7.27605 \cdot 10^{-3}$	0.806954
Zbiór testowy (15%)	14985	$8.27152 \cdot 10^{-3}$	0.822387

Formuła (1) przedstawia sposób wyliczenia MSE:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i' - y^*_i\right)^2 \tag{1}$$

gdzie: n – rozdzielczość obrazu; y_i' – wartość referencyjna i-tego piksela; y^*_i – wartość rekonstrukcyjna i-tego piksela.

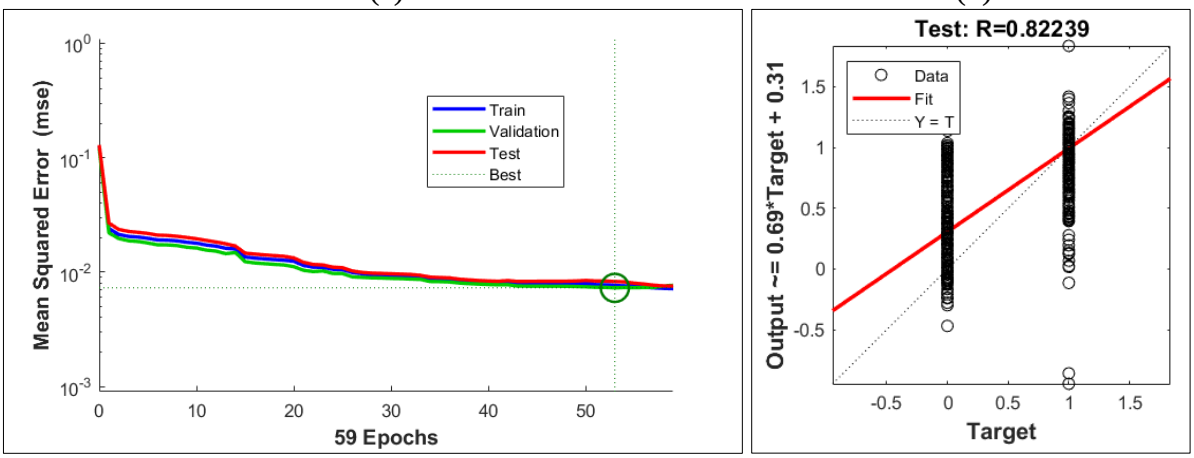
Sposób kalkulacji współczynnika regresji R przedstawia formuła (2):

$$R(y', y^*) = \frac{cov(y', y^*)}{\sigma_{y'}\sigma_{y^*}} \tag{2}$$

gdzie: $\sigma_{y'}$ – odchylenie standardowe wartości referencyjnych, σ_{y^*} – odchylenie standardowe wartości zrekonstruowanych.

Na Rys. 6a przedstawiono wykres przebiegu uczenia wybranej ANN. Jednorodny, hiperboliczny kształt linii wykresu świadczy o braku przeuczenia sieci. Brak fluktuacji oraz zgodność linii dla wszystkich 3 zbiorów danych (uczącego, walidacyjnego i testowego) świadczy o braku przeuczenia oraz o zdolności sieci do generalizacji. Wykres przedstawia także moment, w którym nastąpiło zatrzymanie uczenia sieci. Miało to miejsce po 53 epoce, kiedy MSE zbioru walidacyjnego osiągnęło wartość 0.0072761. Tę samą wartość można również odczytać z kolumny 3 Tabeli 1. Zbiór walidacyjny został wykorzystany do ustalenia warunku zatrzymania uczenia ANN. Warunek ten jest spełniony, jeśli przez 6 kolejnych epok MSE zbioru walidacyjnego nie zmniejsza się.

(a) (b)



Rys. 6. Rezultaty procesu uczenia dla wybranej sieci: a) wykresy błędów MSE dla zbioru uczącego, walidacyjnego i testowego, b) statystyka regresyjna dla zbioru testowego

Informacje na temat wytrenowanej sieci uzupełnia Rys. 6b, na którym przedstawiono wykres statystyki regresyjnej dla zbioru testowego. Zbiór testowy daje najbardziej wiarygodne rezultaty odnośnie jakości otrzymanej sieci, bowiem dane w nim zawarte nie miały żadnego wpływu na proces uczenia. Jak widać R=0.82239 jest zgodna z informacją z Tabeli 1 zawartą w dolnym rzędzie kolumny 4. Charakterystyczny sposób rozmieszczenia danych na Rys. 6b wynika z faktu, że wzorcowy obraz wyjściowy odzwierciedla jedynie 2 wartości: 1 – dla koloru tła oraz 0 – kolor artefaktu (kryształu lub pęcherza gazu).

2.2. Poprawa jakości danych pomiarowych metodą elastic net

W przypadku rekonstrukcji obrazów tomograficznych obiektów rzeczywistych o stosunkowo niskiej przewodności, dane z elektrod są zazwyczaj zaszumione. Jest to wynikiem niedoskonałości izolacji elektrod, wpływom szybkozmiennych prądów o niskim natężeniu generowanych przez multipleksery, wpływom pól elektromagnetycznych oraz wielu innych czynników. Przykładem obiektów technicznych, z których dane tomograficzne wykazują wysoki poziom szumu są także przemysłowe reaktory chemiczne [18]. Zakłócenia sygnałów elektrycznych są jedną z głównych barier utrudniających opracowanie metod tomograficznych dla takich obiektów [23].

Aby uodpornić dane wejściowe na zakłócenia i zniekształcenia, zastosowano regularyzację elastic net [22]. W metodzie tej zakładamy pewien system liniowy, który może być opisany za pomocą równania stanu (3),

$$Y = X\beta + \varepsilon \tag{3}$$

gdzie $Y \in R^n$ jest macierzą zmiennych wyjściowych (rekonstrukcji), $X \in R^{n\times(k+1)}$ reprezentuje macierz zmiennych wejściowych, współczynnik $\beta \in R^{k+1}$ oznacza wektor o nieznanych parametrach, a $\varepsilon \in R^n$ odzwierciedla sekwencję zakłóceń. Jeżeli zadanie liniowe (4) posiada rozwiązanie, w którym prosta regresji przecina oś y, wtedy pierwsza kolumna macierzy X w równaniu liniowym (3) jest jednostkowym wektorem kolumnowym.

W przypadku gdy predyktory wprowadzane do modelu regresji są ze sobą silnie skorelowane, możliwym sposobem określenia regresji liniowej jest rozwiązanie zadania (4).

$$\min_{(\beta_0,\beta')\in R^{k+1}} \frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \beta_0 - x_i\beta'\right)^2 + \lambda P_\alpha\left(\beta'\right) \tag{4}$$

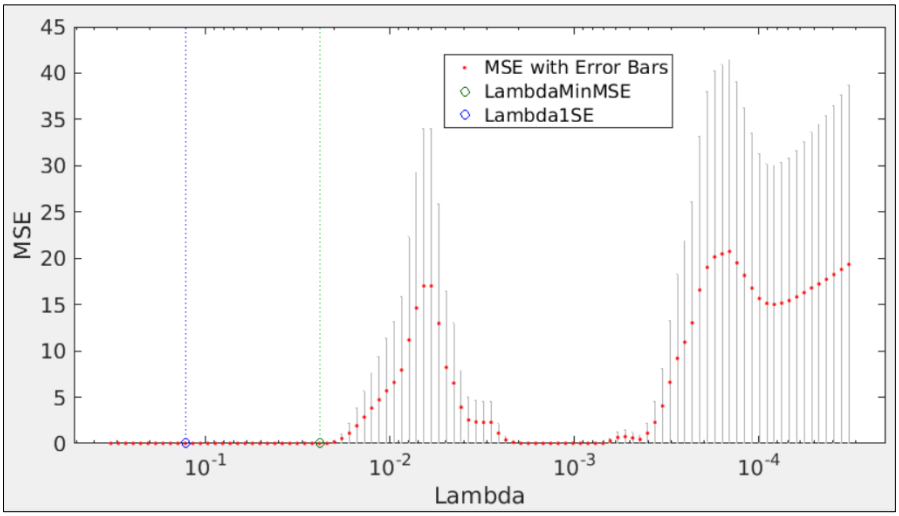
gdzie: $x_i = (x_{i1}, \ldots, x_{ik})$, $\beta' = (\beta_1, \ldots, \beta_k)$ dla $1 \le i \le n$

P_α – funkcja kary elastic net określona równaniem (3).

$$P_\alpha\left(\beta'\right) = (1-\alpha)\frac{1}{2}\left\|\beta'\right\|_{L_2} + \alpha\left\|\beta'\right\|_{L_1} = \sum_{j=1}^{k}\left(\frac{1-\alpha}{2}\beta_j^2 + \alpha\left|\beta_j\right|\right) \tag{5}$$

Można zauważyć, że kara w tym przypadku jest liniową kombinacją norm L_1 i L_2 z nieznanych parametrów β'. Wprowadzenie funkcji kar zależnych od parametrów do funkcji celu redukuje prognozy nieznanych parametrów. Parametr λ w zadaniu (5) reprezentuje współczynnik kary. Wprowadzenie parametru $0 \le \alpha \le 1$ sprawia, że zadanie (4) zachowuje kompromis między Lasso (ang. least absolute shrinkage and selection operator) i regresją grzbietową. Jeżeli $\alpha = 0$ mamy do czynienia z czystą regresją grzbietową zwaną regularyzacją Tichonowa. Regresja grzbietowa jest popularną metodą regularyzacji modeli liniowych [22].

Jeżeli $\alpha = 1$, w zadaniu (4) występuje czysta metoda Lasso. Lasso to metoda statystyczna, która umożliwia dobór zmiennych niezależnych i regularyzację modeli liniowych. W przypadku regresji grzbietowej współczynnik kar obliczany jest w normie L_1 a dla Lasso w L_2. Lasso jest także obojętna na korelację predyktorów.



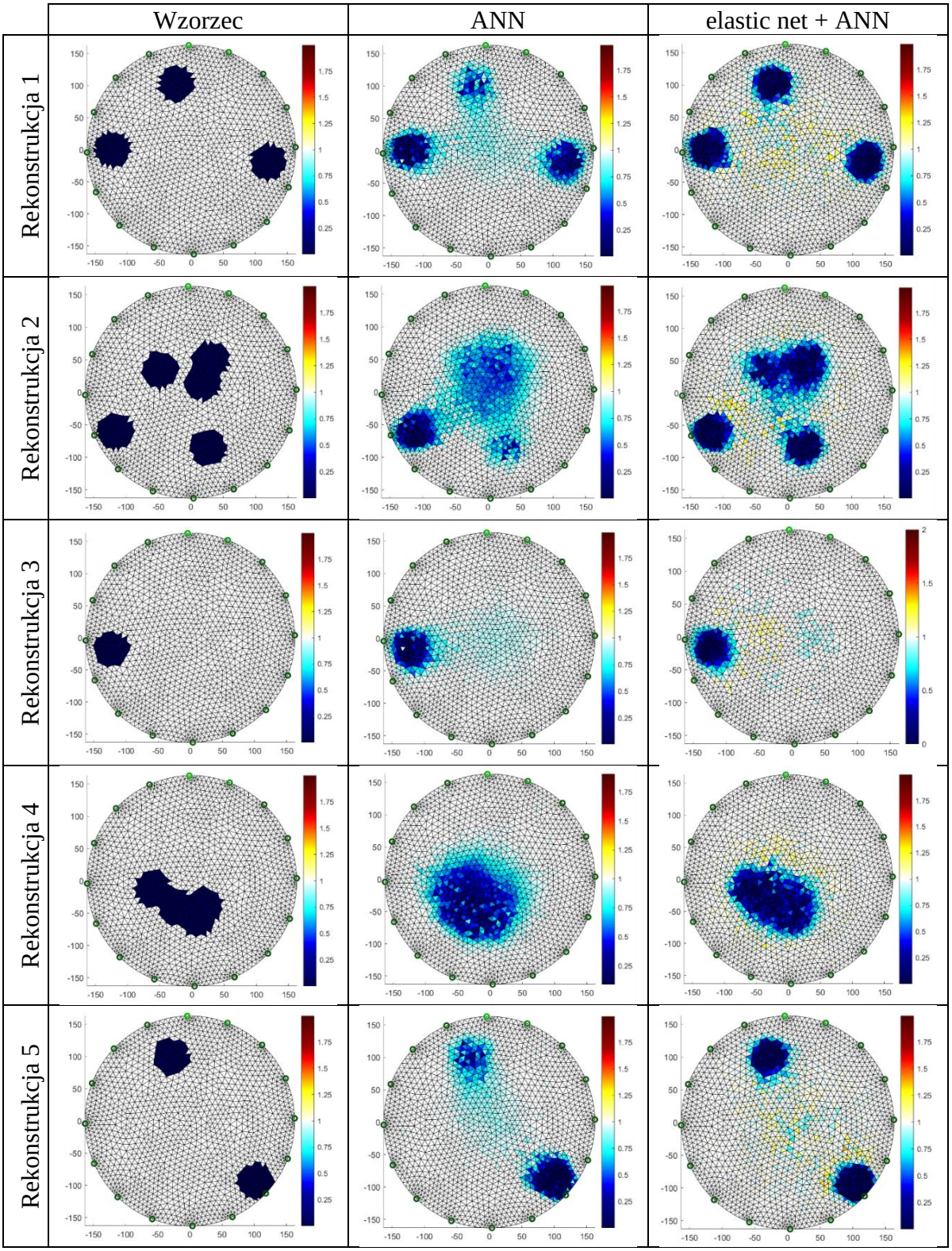
Rys. 7. Dobór parametru Lambda w metodzie elastic net z α=0.25 w oparciu o minimalny błąd średniokwadratowy (MSE)

Rysunek 7 przedstawia proces doboru parametru λ (lambda) w metodzie elastic net z przyjętym $\alpha = 0.25$ dla jednego z punktów obrazu wyjściowego. Dwie wartości Lambdy oznaczone są zieloną i niebieską linią przerywaną. Linia zielona wskazuje wartość Lambda z minimalnym błędem średniokwadratowym (MSE) uzyskaną przy wykorzystaniu sprawdzianu krzyżowego (kroswalidacji). Wartość ta została oznaczona zmienną LambdaMinMSE. Niebieska linia przerywana wskazuje największą wartość Lambda, która mieści się w granicach odchylenia standardowego minimalnego MSE (Lambda1SE). Taka wartość Lambda sprawia, że nawet najgorszy z testowanych modeli (z największym MSE) ma wciąż relatywnie niską wartość błędu. W opisywanym przypadku przyjęto $\lambda = 0.129$ dla MSE=0.0154. Wektor pomiarowy liczy 96 wartości spadków napięć. W prezentowanym przypadku, dzięki zastosowaniu elastic net zredukowano wektor wejściowy z 96 do 30 pomiarów.

3. Rezultaty

Na Rys. 8 przedstawiono porównanie obrazów tomograficznych uzyskanych dwoma metodami: ANN oraz metodą hybrydową, łączącą elastic net z ANN. W pierwszej kolumnie umieszczono obrazy wzorcowe dla pięciu testowanych przypadków. Stanowisko badawcze zostało wyskalowane w taki sposób, że wartość na wyjściu, odpowiadająca konduktancji cieczy referencyjnej (wody) wynosi 1. Z kolei wartość punktów obrazu odpowiadająca artefaktom wynosi 0. Następnie, wykorzystując model fizyczny dokonano pomiarów rzeczywistych metodami ANN oraz elastic net + ANN (algorytm hybrydowy).

Porównując obrazy uzyskane dwoma metodami, można zauważyć, że w podejściu hybrydowym, pomimo zredukowania wektora pomiarów wejściowych o prawie 70% dla każdej sieci neuronowej, wizualna jakość rekonstrukcji nie ustępuje czystej ANN. Należy także zauważyć, że algorytm najgorzej rekonstruuje artefakty zlokalizowane w pobliżu środka reaktora. Ma to najprawdopodobniej związek z odległością punktów siatki obrazu od elektrod. Lepsze rekonstrukcje uzyskiwane są dla artefaktów położonych bliżej elektrod.



Rys. 8. Porównanie obrazów rekonstrukcyjnych dla metod ANN i „elastic net + ANN”

Wizualna ocena rekonstrukcji jest mało precyzyjna i niewystarczająca do obiektywnego porównania dwóch analizowanych metod. W tym celu wprowadzono 3 ilościowe mierniki jakości obrazowania: MSE (Mean Squared Error), RIE (Relative Image Error), ICC (Image Correlation Coefficient).

Względny błąd obrazowania RIE wyliczany jest zgodnie z formułą (6):

$$RIE = \frac{\left\| Y^* - Y' \right\|}{\left\| Y' \right\|} \tag{6}$$

gdzie: Y' – rozkład prawdopodobieństwa punktów obrazu wzorcowego; Y^* – wartości punków zrekonstruowanych.

Formuła (7) przedstawia sposób wyliczenia współczynnika korelacji obrazu ICC:

$$ICC = \frac{\sum_{i=1}^{l} \left(y_i^* - \bar{Y}^* \right) \left(y_i' - \bar{Y}' \right)}{\sqrt{\sum_{i=1}^{l} \left(y_i^* - \bar{Y}^* \right)^2 \sum_{i=1}^{l} \left(y_i' - \bar{Y}' \right)^2}} \tag{7}$$

gdzie: $\bar{Y}'$ – średni rozkład prawdopodobieństwa punktów obrazu wzorcowego; $\bar{Y}^*$ – średnia wartości punków zrekonstruowanych.

Im mniejsze są wartości współczynników MSE i RIE, tym lepsza jakość rekonstrukcji. W przypadku ICC jest odwrotnie - im bliżej 1, tym lepsza jest korelacja obrazu wyjściowego z obrazem referencyjnym, co przekłada się na wierniejszą rekonstrukcję.

Tabela 2 zawiera zestawienie wszystkich trzech wskaźników jakości rekonstrukcji z podziałem na metody i testowane przypadki. W górnej części tabeli znajdują się precyzyjnie wyliczone wartości MSE, RIE i ICC. W prawej kolumnie zamieszczono wartości uśrednione, dające możliwość dokonania uogólnionej oceny wszystkich pięciu przypadków jednym wskaźnikiem. W celu ułatwienia wyboru metody lepszej, w dolnej części tabeli zamieszczono zerojedynkową klasyfikację wszystkich wskaźników. Trzy pytania zostały tak sformułowane, że w przypadku odpowiedzi twierdzącej („prawda") lepszą metodą jest „elastic net + ANN". W przypadku odpowiedzi negatywnej, wskaźnik preferuje metodę czystej ANN.

Tabela 2. Porównanie wskaźników rekonstrukcji

Wskaźnik jakości rekonstrukcji		Numer rekonstrukcji					Średnia
		1	*2*	*3*	*4*	*5*	
ANN	MSE	0.0232	0.0573	0.0061	0.0337	0.0198	0.0280
	RIE	0.1587	0.2547	0.0795	0.1913	0.1451	0.1658
	ICC	0.8319	0.6715	0.8694	0.7301	0.7982	0.7802
elastic net + ANN	MSE	0.0158	0.0388	0.0047	0.0238	0.0115	0.0189
	RIE	0.1312	0.2096	0.0695	0.1610	0.1107	0.1364
	ICC	0.8804	0.7892	0.8987	0.8177	0.8907	0.8553
Czy MSE algorytmu hybrydowego jest mniejszy od MSE czystej ANN?		PRAWDA	PRAWDA	PRAWDA	PRAWDA	PRAWDA	PRAWDA
Czy RIE algorytmu hybrydowego jest mniejszy od RIE czystej ANN?		PRAWDA	PRAWDA	PRAWDA	PRAWDA	PRAWDA	PRAWDA
Czy ICC algorytmu hybrydowego jest większy od ICC czystej ANN?		PRAWDA	PRAWDA	PRAWDA	PRAWDA	PRAWDA	PRAWDA

Jak wynika z Tabeli 2, we wszystkich 18 przypadkach, łącznie z odpowiedziami obejmującymi wartości średnie, lepszą okazała się metoda hybrydowa „elastic net + ANN".

4. Dyskusja i konkluzje

W pracy przedstawiono wyniki badań nad opracowaniem ulepszonego algorytmu rekonstrukcji obrazów z zakresu tomografii procesowej. W szczególności, skoncentrowano się na obrazowaniu wnętrza modelu zbiornikowych reaktorów przemysłowych, w których zachodzą procesy krystalizacji i fermentacji beztlenowej w biogazowniach. Aby dokonać obiektywnej oceny jakości działania algorytmu hybrydowego (elastic net + ANN), przeprowadzono badania porównawcze wykorzystując fizyczny model laboratoryjny. Przygotowano jednakowe przypadki układów artefaktów, a następnie dokonano pomiarów spadków napięć odczytanych z układu elektrod. Z uwagi na to, że podczas odczytu danych wiele stanów prądowo-napięciowych jest nieustalonych, dane zawierały pewien poziom zaszumienia. Po przefiltrowaniu danych przy użyciu wartości referencyjnych tła (reaktora bez artefaktów) uzyskano rezultaty, które zostały zobrazowane i przeliczone na wskaźniki jakości: MSE, RIE i ICC.

Najprawdopodobniej, gdyby dane pomiarowe użyte do obrazowania były wygenerowane symulacyjnie i nie zawierały zakłóceń, usunięcie 70% predyktorów (np. redukcja z 96 do 30) spowodowałoby pogorszenie jakości obrazów. Przeprowadzone eksperymenty wykazały, że inaczej stało się w przypadku danych obarczonych pewnym poziomem zaszumienia. Choć uzyskane rekonstrukcje, zarówno dla czystej ANN jak i dla metody hybrydowej, wciąż nie są idealne, to jednak są wystarczająco dokładne aby określić ilość, kształt i lokalizację artefaktów. Czas uzyskiwania rekonstrukcji we wszystkich badanych przypadkach na maszynie z procesorem Intel i7 nie przekroczył 1 sekundy. Oznacza to, że opracowany algorytm może być stosowany także w procesach o dużej dynamice reakcji, a nawet w układach przepływowych [24].

Znamienne jest, że w Tabeli 2, we wszystkich testowanych przypadkach i dla wszystkich trzech wskaźników jakości, lepsze rezultaty otrzymano przy wykorzystaniu algorytmu „elastic net + ANN", a nie czystej ANN. Okazuje się, że zastosowanie elastic net spowodowało, że wraz z usunięciem 70% zmiennych niezależnych, usunięto także wiele danych zaszumionych, generujących różnego rodzaju zaburzenia obrazu wyjściowego. Fakt ten jest dowodem na to, że nowoopracowany algorytm hybrydowy „oczyszcza" dane i uodparnia układ tomograficzny na różnego rodzaju zakłócenia i zaszumienie danych.

Rezultaty badań udowodniły, że zastosowanie opisywanego rozwiązania umożliwia uzyskanie wyższej rozdzielczości obrazów kryształów lub pęcherzy gazowych pojawiających się wewnątrz reaktora, co zasadniczo przyczynia się do wczesnego wykrycia zagrożeń i ułatwia rozwiązywanie problemów eksploatacyjnych systemów przemysłowych.

Podziękowanie: autorzy składają podziękowania władzom i pracownikom Wydziału Matematyki, Fizyki i Informatyki UMCS w Lublinie za udostępnienie zasobów superkomputerowych.

Literatura

1. Babout L, Grudzień K, Wiącek J, Niedostatkiewicz M, Karpiński B, Szkodo M. Selection of Material for X-Ray Tomography Analysis and DEM Simulations: Comparison between Granular Materials of Biological and Non-Biological Origins. Granular Matter 2018; 20 (3): 38.

2. Banasiak R, Wajman R, Sankowski D, Soleimani M. Three-Dimensional Nonlinear Inversion of Electrical Capacitance Tomography Data Using a Complete Sensor Model. Progress In Electromagnetics Research (PIER) 2010; 100: 219-234..

3. Dusek J, Hladky D, Mikulka J. Electrical Impedance Tomography Methods and Algorithms Processed with a GPU. Progress In Electromagnetics Research Symposium - Spring (PIERS) 2017; 1710–14.

4. Garbaa H, Jackowska-Strumiłło L, Grudzień K, Romanowski A. Application of Electrical Capacitance Tomography and Artificial Neural Networks to Rapid Estimation of Cylindrical Shape Parameters of Industrial Flow Structure. Archives of Electrical Engineering 2016; 65 (4): 657–69.

5. Grudzien, Krzysztof, Zbigniew Chaniecki, Andrzej Romanowski, Dominik Sankowski, Jacek Nowakowski, and Maciej Niedostatkiewicz. Application of Twin-Plane ECT Sensor for Identification of the Internal Imperfections inside Concrete Beams. IEEE International Instrumentation and Measurement Technology Conference Proceedings 2016; May, 1–6.

6. Kłosowski G, Gola A, Świć A. Application of Fuzzy Logic Controller for Machine Load Balancing in Discrete Manufacturing System. In International Conference on Intelligent Data Engineering and Automated Learning 2015; 256–63.

7. Kłosowski G, Rymarczyk T, Gola A. Increasing the Reliability of Flood Embankments with Neural Imaging Method. Applied Sciences 2018; 8 (9): 1457.

8. Kłosowski G, Rymarczyk T. Using neural networks and deep learning algorithms in electrical impedance tomography. Informatyka Automatyka Pomiary w Gospodarce i Ochronie Środowiska 2017; 7 (3): 99–102.

9. Korzeniewska E, Gałązka-Czarnecka I, Czarnecki A, Piekarska A, Krawczyk A. Influence of PEF on Antocyjans in Wine. Przegląd Elektrotechniczny 2018; 1 (1): 59–62.

10. Korzeniewska E., Walczak M., Rymaszewski J., Elements of Elastic Electronics Created on Textile Substrate, Proceedings of the 24th International Conference Mixed Design of Integrated Circuits and Systems - MIXDES 2017; 2017, 447-45.

11. Kosicka E, Kozłowski E, Mazurkiewicz D. Intelligent Systems of Forecasting the Failure of Machinery Park and Supporting Fulfilment of Orders of Spare Parts. In: Burduk A., Mazurkiewicz D. (eds) Intelligent Systems in Production Engineering and Maintenance – ISPEM 2017. ISPEM 2017. Advances in Intelligent Systems and Computing, vol 637. Springer, Cham, 2018.

12. Kozłowski E., Mazurkiewicz D., Kowalska B., Kowalski, D. Binary Linear Programming as a Decision-Making Aid for Water Intake Operators. In: Burduk A., Mazurkiewicz D. (eds) Intelligent Systems in Production Engineering and Maintenance – ISPEM 2017. ISPEM 2017. Advances in Intelligent Systems and Computing, vol 637. Springer, Cham, 2018.

13. Kryszyn J, Smolik W T, Radzik B, Olszewski T, Szabatin R. Switchless Charge-Discharge Circuit for Electrical Capacitance Tomography. Measurement Science and Technology 2014; 25 (11): 115009.

14. Kryszyn J, Waldemar S. Toolbox for 3d Modelling and Image Reconstruction in Electrical Capacitance Tomography. Informatics Control Measurement in Economy and Environment Protection 2017; 7 (1).

15. Lopato P, Tomasz C, Sikora R, Gratkowski S, Ziolkowski M. Full Wave Numerical Modelling of Terahertz Systems for Nondestructive Evaluation of Dielectric Structures. COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering 2013; 32 (3): 736–49.

16. Majchrowicz M, Kapusta P, Jackowska-Strumiłło L, Sankowski D. Acceleration of image reconstruction process in the electrical capacitance tomography 3d in heterogeneous, multi-GPU system. Informatics Control Measurement in Economy and Environment Protection 2017; 7 (1): 37–41.

17. Mikulka J. Accelerated Reconstruction of T2 Maps in Magnetic Resonance Imaging. Measurement Science Review 2015; 4: 210–18.

18. Park S, Na J, Kim M, Lee J M. Multi-Objective Bayesian Optimization of Chemical Reactor Design Using Computational Fluid Dynamics. Computers & Chemical Engineering 2018; 119 : 25–37.

19. Psuj G. Multi-Sensor Data Integration Using Deep Learning for Characterization of Defects in Steel Elements. Sensors 18 (2): 292.

20. Romanowski, Andrzej. 2018. Big Data-Driven Contextual Processing Methods for Electrical Capacitance Tomography. IEEE Transactions on Industrial Informatics 2018; 1–1.

21. Rymarczyk T, Adamkiewicz P, Polakowski K, Sikora J. Effective Ultrasound and Radio Tomography Imaging Algorithm for Two-Dimensional Problems. Przegląd Elektrotechniczny 2018; 94 (6): 62–69.

22. Rymarczyk T, Kłosowski G, Kozłowski E. A Non-Destructive System Based on Electrical Tomography and Machine Learning to Analyze the Moisture of Buildings. Sensors 2018; Vol. 18, Page 2285 18 (7): 2285.

23. Rymarczyk T, Kłosowski G. Application of Neural Reconstruction of Tomographic Images in the Problem of Reliability of Flood Protection Facilities. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (3): 425–34.

24. Rymarczyk T, Sikora J. Applying Industrial Tomography to Control and Optimization Flow Systems. Open Physics 2018; 16 (1): 332–45.

25. Sobaszek Ł, Gola A, Świć A. Predictive Scheduling as a Part of Intelligent Job Scheduling System: in, 358–67. Springer, Cham 2018;

26. Soleimani M, Mitchell C N, Banasiak R, Wajman R, Adler A. Four-dimensional electrical capacitance tomography imaging using experimental data. Progress In Electromagnetics Research 2009; 90: 171–86.

27. Tian G, Yang B, Dong M, Zhu R, Yin F, Zhao X, Wang Y, Xiao W, Wang Q, Zhang W. The Effect of Temperature on the Microbial Communities of Peak Biogas Production in Batch Biogas Reactors. Renewable Energy 2018; 123: 15–25.

28. Voutilainen A, Lehikoinen A, Vauhkonen M, Kaipio J P. Three-Dimensional Nonstationary Electrical Impedance Tomography with a Single Electrode Layer. Measurement Science and Technology 2010; 21 (3): 035107.

29. Wang Mi. Industrial Tomography: Systems and Applications. Edited by Elsevier Ltd. Woodhead Publishing 2015.

30. Ziolkowski M, Gratkowski S, Zywica A R. Analytical and Numerical Models of the Magnetoacoustic Tomography with Magnetic Induction. COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering 2018; 37 (2): 538–48.

Prof. dr hab. inż. Andrzej Gołaś
Dr inż. Wojciech Ciesielka
Dr inż. Krystian Szopa

AGH Akademia Górniczo-Hutnicza
Wydział Inżynierii Mechanicznej i Robotyki
Katedra Systemów Energetycznych i Urządzeń Ochrony Środowiska
Al. Mickiewicza 30, 30-059 Kraków, Polska
E-mail: ghgolas@cyf-kr.edu.pl, ghciesie@cyf-kr.edu.pl, kszopa@agh.edu.pl

Dr hab. inż. Paweł Zydroń, prof. n. AGH
Dr inż. Wojciech Bąchorek
Dr inż. Mariusz Benesz
Dr inż. Aleksander Kot
Dr inż. Szczepan Moskwa

AGH Akademia Górniczo-Hutnicza
Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej
Katedra Elektrotechniki i Elektroenergetyki
Al. Mickiewicza 30, 30-059 Kraków, Polska
E-mail: pzydron@agh.edu.pl, wojbach@agh.edu.pl, mben@agh.edu.pl, akot@agh.edu.pl, szczepan@agh.edu.pl

Analiza możliwości poprawy niezawodności napowietrznej linii 15 kV narażonej na katastrofalne oblodzenie w warunkach polskich

Słowa kluczowe: elektroenergetyczne sieci dystrybucyjne, linie napowietrzne, niezawodność, oblodzenie, szadź

Streszczenie: Praca jest efektem synergicznej współpracy dwóch zespołów akademickich: elektroenergetycznego i mechanicznego oraz operatora systemu dystrybucyjnego. Analizie poddano rzeczywistą, napowietrzną linię średniego napięcia 15 kV narażoną na katastrofalne obciążenia lodem i szadzią. Zbadano możliwość zastosowania trzech rozwiązań mogących poprawić niezawodność badanego obiektu w takich warunkach. Rozważono: skrócenie długości przęseł linii, podgrzewanie magistrali zwiększonym prądem roboczym oraz przebudowę linii do linii kablowej. W celu realizacji pracy wykonano badania modelowo-symulacyjne MES z uwzględnieniem najnowszych wytycznych normatywnych, zrealizowano eksperyment dociążenia linii wraz z pomiarem temperatury przewodów oraz przeprowadzono wielowariantowe obliczenia niezawodnościowe prowadzące do wyznaczenia wskaźników SAIDI i SAIFI. W wyniku szczegółowych analiz sprecyzowano wnioski końcowe pozwalające na zwiększenie niezawodności linii elektroenergetycznych narażonych na katastrofalne oblodzenie, które powinny być rozważone i stosowane przez wszystkich operatorów systemów dystrybucyjnych w Polsce.

1. Wprowadzenie

Dostarczanie energii elektrycznej jest bardzo istotnym elementem funkcjonowania współczesnego społeczeństwa. Pewność i ciągłość dostaw energii oraz minimalizacja przerw w zasilaniu jest obecnie jednym z podstawowych priorytetów operatorów systemów dystrybucyjnych.

Wobec obserwowanej ekstremalizacji klimatu problem ekspozycji napowietrznej infrastruktury sieciowej na oddziaływanie trudnych warunków środowiskowych (oblodzenie, wiatry) nabiera szczególnego znaczenia.

Najpoważniejszym problemem w projektowaniu i budowie linii napowietrznych jest precyzyjne określenie dodatkowych obciążeń linii wywołanych lodem i/lub szadzią osadzającą się na przewodach i konstrukcjach wsporczych. Zagadnieniem tym zajmowano się między innymi podczas standaryzacji tych obciążeń w IEC [15].

W ostatnich latach miało miejsce w Europie i na świecie wiele awarii energetycznych wywołanych oddziaływaniem na elementy systemów elektroenergetycznych niekorzystnych czynników atmosferycznych. Większość krajów świata zlokalizowanych na półkuli północnej, takich jak Chiny, Niemcy [2], [11], północno-wschodnia część USA i Kanada [5], Czechy [17], Północny Kaukaz w Rosji [10], Japonia [1], Finlandia [14], Norwegia [7], Rumunia [9], Węgry [13], Wielka Brytania [20], Islandia [4], [8], czy część Indii boryka się z problemem związanym z tworzeniem się lodu na liniach napowietrznych. Również w Polsce wystąpiły katastrofalne awarie spowodowane ekstremalnymi warunkami pogodowymi, które opisano w [2] i [11].

Wagę tego problemu dostrzeżono na forum międzynarodowym, między innymi w materiałach CIGRE poświęcono temu zagadnieniu bardzo dużo uwagi [18], [19]. Jest to także temat podstawowy odbywających się cyklicznie spotkań w ramach International Workshop on Atmospheric Icing of Structures.

Po doświadczeniach jak opisane powyżej narodowe komitety wielu państw zdecydowały o konieczności zmian standardów obowiązujących przy obliczeniach projektowych linii. Także w Polsce zostały zmienione strefy obciążenia oblodzeniem [16].

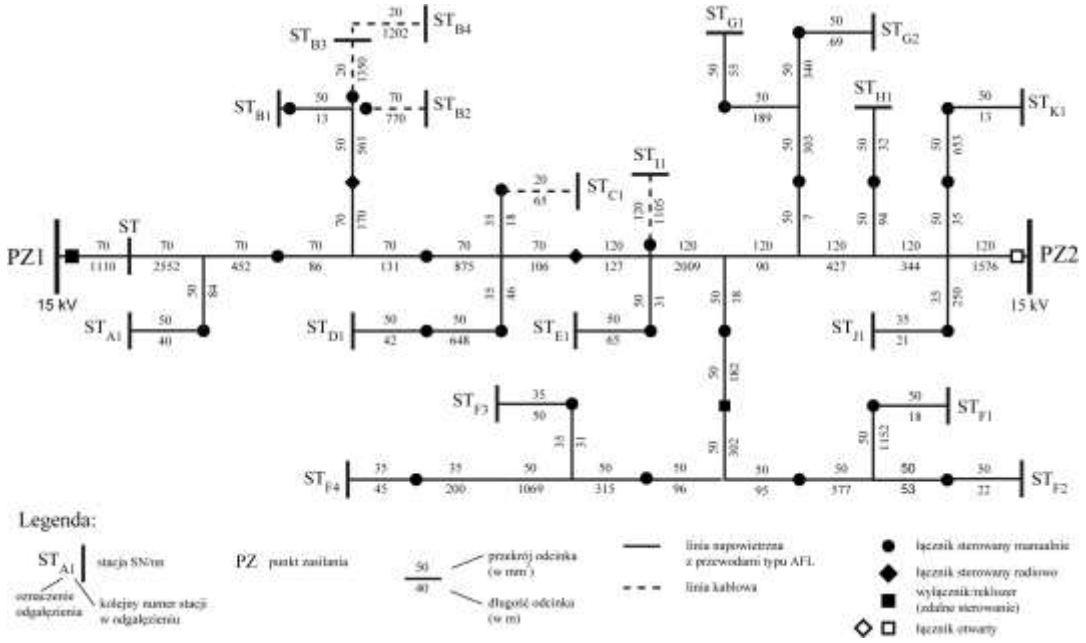
Doniesienia w literaturze światowej dokumentują, szereg metod i technik mających na celu zapobieganie oblodzeniu i usuwanie lodu i/lub szadzi z napowietrznych linii elektroenergetycznych [6] jednak ich skuteczność jest znikoma w warunkach oddziaływania katastrofalnego oblodzenia. Wspomniane rozwiązania dotyczą głównie sieci przesyłowych.

Prezentowana praca analizuje trzy rozwiązania mogące poprawić niezawodność obiektu narażonego na katastrofalne obciążenia oblodzeniem. Rozważono: zmniejszenie rozpiętości przęseł (dogęszczenie konstrukcji wsporczych), podgrzewanie magistrali dodatkowym prądem roboczym oraz przebudowę linii do linii kablowej. W celu realizacji pracy wykorzystano badania modelowo-symulacyjne MES uwzględniając najnowsze wytyczne normatywne, zrealizowano pomiary termowizyjne w warunkach dodatkowego obciążenia obiektu oraz przeprowadzono wielowariantowe obliczenia niezawodnościowe prowadzące do wyznaczenia wskaźników SAIDI i SAIFI. Sprecyzowano wnioski końcowe mające na celu poprawę niezawodności badanego obiektu.

2. Charakterystyka obiektu

Przedmiotem analizy w niniejszej pracy jest rzeczywista napowietrzna linia rozdzielcza 15 kV. Obiekt jest zlokalizowany w południowej części Polski. Schemat linii

przedstawiono na rysunku 1. Jest to ciąg napowietrzny zasilający 11 odgałęzień, z których trzy zawierają odcinki kablowe. Odgałęzienia oznaczono kolejnymi literami alfabetu. Magistrala zasilana jest z rozdzielni PZ1, a punkt stałego podziału w normalnym układzie pracy znajduje się w rozdzielni PZ2.



Rys. 1. Schemat linii 15 kV będącej przedmiotem analizy

Wybrane dane charakteryzujące badaną linię SN zebrano w tabeli 1.

Tabela 1. Podstawowe dane analizowanego ciągu sieciowego 15 kV

	Rodzaj budowy/typ	**Długość/Liczba**	**Razem**
Magistrala	linia napowietrzna AFL	9,9 km	9,9 km
	linia kablowa	0 km	
Odgałęzienia	linia napowietrzna AFL	7,8 km	12,3 km
	linia kablowa	4,5 km	
Konstrukcje wsporcze (słupy)	żelbetowe	12 szt.	174 szt.
	stalowe	4 szt.	
	strunobetonowe	41 szt.	
	wirowane	114 szt.	
	drewniane	3 szt.	
Łączniki	sterowane manualnie	25 szt.	28 szt.
	sterowane radiem	2 szt.	
	reklozery	1 szt.	
Liczba zasilanych stacji SN/nn	19 szt.		
Liczba zasilanych odbiorców	945 szt.		
Roczny zakres obciążeń (min – max)	6 A – 16 A		

Dla analizowanej linii napowietrznej zimą 2010 roku wystąpiły katastrofalne warunki oblodzeniowe, które doprowadziły do rozległej destrukcji infrastruktury sieciowej.

3. Badania modelowo-symulacyjne MES wybranych elementów infrastruktury linii dla katastrofalnych warunków atmosferycznych

Najistotniejszym elementem badań modelowo-symulacyjnych MES było uwzględnienie rzeczywistych oddziaływań katastrofalnych jakie wystąpiły na znacznym obszarze, na którym zlokalizowana jest analizowana linia. W styczniu 2010 roku miała miejsce anomalia pogodowa, która przyczyniła się do powstania na liniach napowietrznych oblodzenia o rozmiarze daleko przekraczającym założenia normatywne (nawet 18 kg/m i 15 cm średnicy). W tym punkcie przedstawiono wpływ takiego katastrofalnego oblodzenia na siły reakcji w punktach zawieszenia przewodów oraz siły w samym przewodzie. Ponieważ powstawaniu tak dużej szadzi sprzyja stosunkowo niewielki wiatr, dlatego skoncentrowano się głównie na przypadkach z samym oblodzeniem.

Celem prowadzonych badań symulacyjnych jest określenie wpływu obciążeń atmosferycznych na konstrukcje wsporcze oraz przewody robocze. Pokazano jak asymetria oblodzenia oraz długości przęseł wpływają na zmianę sił w układzie. Nie analizowano wpływu tych oddziaływań na same poprzeczniki konstrukcji. Konstrukcje wsporcze zostały obciążone statycznym układem sił równoważnym działaniu sił pochodzących od przewodów na poprzecznik.

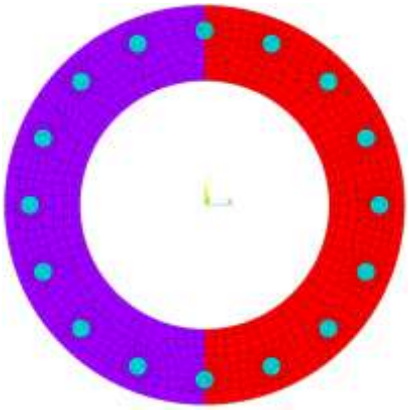
W pierwszej kolejności zbudowano model cyfrowy słupa oraz zweryfikowano poprawność przyjętych założeń.

Obiektem badań jest wybrany odcinek analizowanej linii 15 kV. Jako konstrukcję wsporczą zastosowano żerdź wirowaną typu E 12/10, co oznacza, że jej długość jest równa 12 m (rys. 2), a nominalna siła wierzchołkowa jaką przenosi wynosi 10 kN. Konstrukcja jest wykonana z betonu klasy C 40/50, a druty stalowe, z których wykonano cięgna poprowadzone od podstawy do wierzchołka są zgodne z normami dla zbrojonych konstrukcji betonowych.

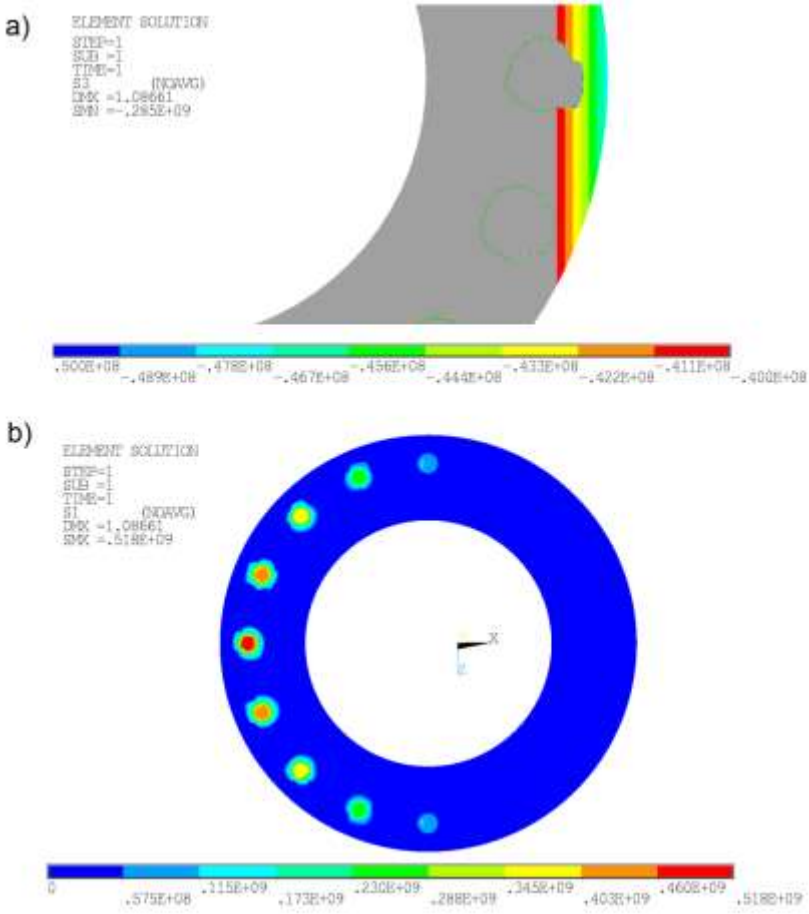
Przekrój słupa ma kształt pierścienia, którego średnica zmniejsza się proporcjonalnie do wysokości konstrukcji (rys. 2). Średnica zewnętrzna słupa przy jego podstawie wynosi 398 mm, a w części wierzchołkowej 218 mm. Na obwodzie słupa rozłożono równomiernie 16 cięgien stalowych o średnicy zastępczej 16 mm. Rozmiar siatki przekroju dobrano 3 mm dla cięgien oraz 10 mm w obszarze betonu (rys. 3).

Dokonano weryfikacji modelu elementów skończonych i przeprowadzono obliczenia dla podanej przez producenta siły niszczącej P_n = 18 kN. Maksymalne naprężenia ściskające w betonie wynoszą 47,5 MPa (rys. 4a), co w zasadzie pokrywa się z granicznym naprężeniem ściskającym dla betonu C 40/50 wynoszącym 48 MPa. Tak więc należy przyjąć, że model został wykonany poprawnie.

Rys. 2. Geometria słupa E 12/10



Rys. 3. Kształt przekroju żerdzi typu E 12/10 wraz z nałożoną siatką



Rys. 4. Naprężenia w żerdzi wirowanej E 12/10 dla siły wierzchołkowej
a) ściskające w betonie, b) rozciągające w cięgnach

Obliczenia dla konstrukcji wsporczej typu E12/10 przeprowadzono ze względu na obciążenia wyznaczone jak dla słupa przelotowego oraz mocnego. Na przewody fazowe wybrano AFL-6-70, ponieważ gołe przewody są bardziej podatne na występowanie dużych oblodzeń, a dodatkowo przewód tego typu nadal jest stosowany w liniach SN. Obliczenia przeprowadzono dla zastępczej rozpiętości przęsła 80 m. Wartość ta jest w przybliżeniu średnią odległością między słupami stosowaną w strefach o dużym charakterystycznym

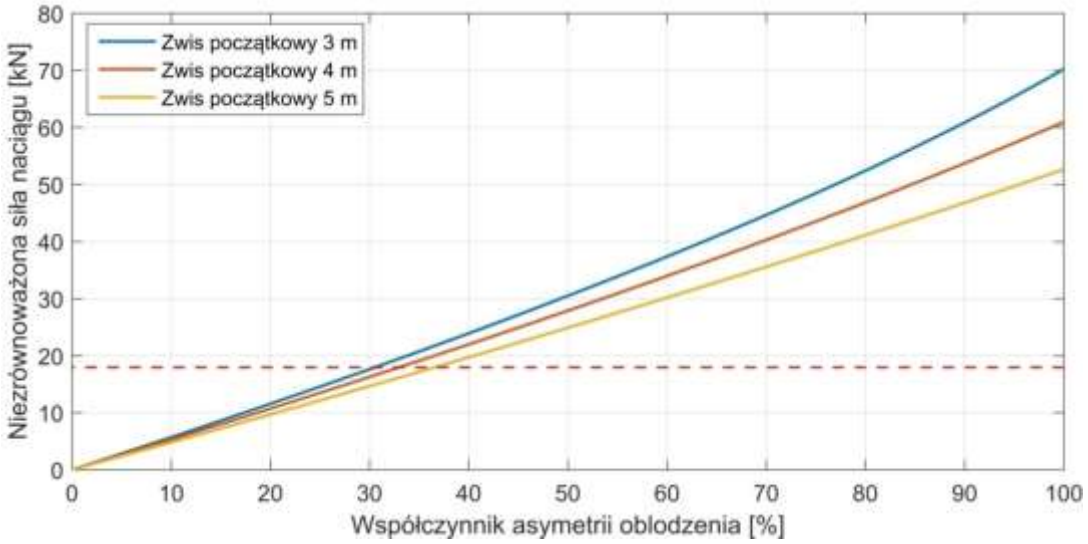
oblodzeniu, z drugiej strony jest na tyle odpowiednia, aby pokazać jak duże siły potrafią wystąpić w stosunkowo krótkim przęśle.

W tabeli 2 przedstawiono wyniki dla opisanego przypadku. Tak wielka masa oblodzenia powoduje powstawanie bardzo dużych sił pionowych, naciągu oraz osiowych w przewodzie. Ponownie przy zrównoważonym oblodzeniu w sąsiednich przęsłach, naprężenia ściskające są niewielkie, natomiast problemy zaczynają się pojawiać dla przypadków asymetrycznego oblodzenia. Jeżeli słup pracuje jako konstrukcja przelotowa to dla przypadku 2c, czyli niezrównoważonego obciążenia wzdłużnego, naprężenia ściskające przekraczają dopuszczalne wartości i dochodzi do zmiażdżenia betonu oraz zniszczenia konstrukcji wsporczej. Jeżeli konstrukcja wsporcza ma pełnić rolę słupa mocnego to dodatkowo musi spełnić warunki z grupy 5. Już dla przypadku zerwania pojedynczego przewodu (5a) pojawia się na tyle duża siła poprzeczna, że naprężenia ściskające w betonie są na granicy wytrzymałości konstrukcji. Natomiast gdyby doszło do sytuacji, w której słup ma za zadanie przeniesienie obciążenia jednostronnego naciągu, to przy tak katastrofalnym oblodzeniu nastąpiłoby zniszczenie konstrukcji wsporczej. Opisane w tym akapicie sytuacje zakładają, że przewody robocze mogą ulec zerwaniu (5a, b, c), albo wytrzymać obciążenie oblodzeniem 18 kg/m (2a, b). Obliczeniowa siła zrywająca dla AFL-6-70 to 22,75 kN. W tabeli nie uwzględniono zwisów ze względu na możliwość wejścia w zakres odkształceń plastycznych.

Ponieważ podany przykład przedstawia wyniki dla konkretnej długości przęsła, zwisu oraz przypadków oblodzenia, przeprowadzono dodatkowe obliczenia celem wyznaczenia zależności pomiędzy wielkością sił poprzecznych oddziałujących na konstrukcję wsporczą typu E 12/10 a rozpiętością przęsła.

Tabela 2. Oddziaływania na konstrukcję wsporczą E 12/10 siłami pochodzącymi od przewodów dla oblodzenia katastrofalnego o masie 18 kg/m

Przypadek	**V**	**H**	**W**	**N**	σ_C	σ_S
	[N]	[N]	[N]	[N]	[MPa]	[MPa]
0	106,0	706,3	0	714,2	0,19	0
2a	4996,4	24133,0	0	24645,0	0,51	0
$\psi_I = 0,3$	1573,1	9258,7	0	9391,4	-	-
$\psi_I = 0,5$	2551,2	13979,2	0	14209,1	-	-
$\psi_I = 0,7$	3529,3	18264,1	0	18602,3	-	-
2b	Kombinacja obciążeń $\psi_I = 0,5$ i $\psi_I = 1$				7,65	48,9
2c	Kombinacja obciążeń $\psi_I = 0,3$ i $\psi_I = 0,7$				71,10	730,0
5a	Zerwanie przewodu dla oblodzenia $\psi_I = 0,7$				48,21	492,2
5b	75% jednostronnego naciągu przy oblodzeniu $\psi_I = 1$				142,25	1471,0
5c	Całkowity jednostronny naciąg przy oblodzeniu $\psi_I=1$				189,20	1962,8

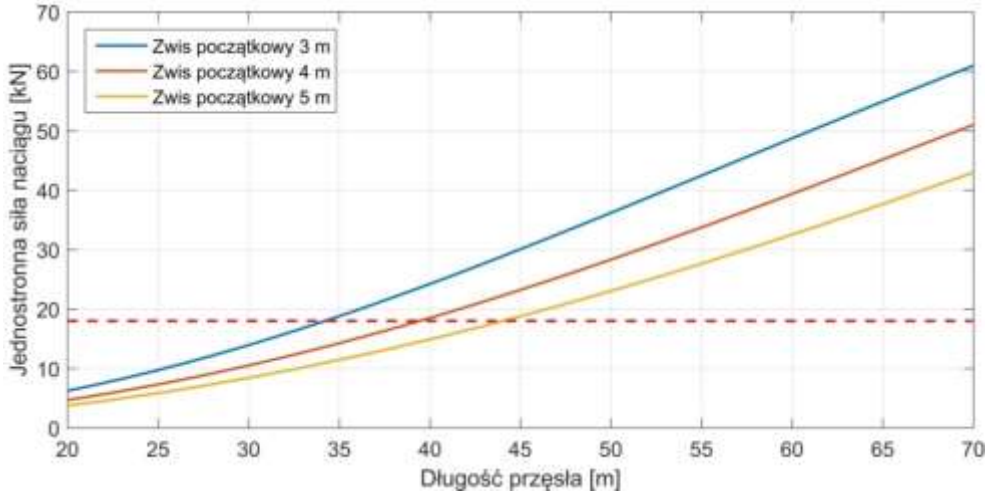


Rys. 5. Zależność niezrównoważonej siły naciągu od współczynnika asymetrii oblodzenia

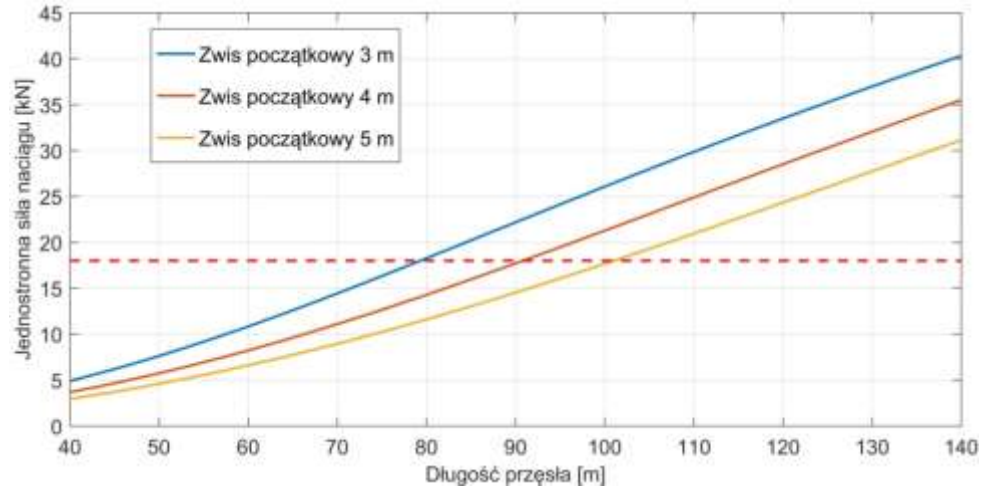
W pierwszej kolejności sprawdzono jednak wpływ asymetrycznego oblodzenia na niezrównoważenie siły naciągu. Przypadek ten przedstawia sytuację, w której wszystkie przewody jednego przęsła są obciążone maksymalnym oblodzeniem katastrofalnym 18 kg/m, natomiast oblodzenie w przęśle sąsiednim zmienia się od wartości maksymalnej ψ_I = 1 do 0. Współczynnik asymetrii jest tutaj rozumiany jako (1-ψ_I)·100%, więc dla maksymalnego oblodzenia obu przęseł współczynnik ten wynosi 0%, natomiast gdy jedno przęsło jest nieoblodzone to asymetria obciążenia oblodzeniem wynosi 100%. Z zależności przedstawionych na rysunku 5 wynika, że dla przęseł od długości 80 m, już przy współczynniku asymetrii 30% – 37% (w zależności od zwisu montażowego) niezrównoważona wartość siły naciągu (od trzech przewodów) osiąga krytyczną wartość siły niszczącej dla słupa E 12/10 równej 18 kN. Potwierdza to wcześniejsze wyniki, gdzie problemem zazwyczaj było nie tyle duże oblodzenie, co jego nierównomierne rozłożenie w sąsiednich przęsłach.

Kolejno sprawdzono jaka może być maksymalna długość przęsła dla obciążenia 18 kg/m, przy założeniu że słup mocny ma za zadanie przenieść naciąg jednostronny (przypadek 5c). Z wykresów przedstawionych na rysunku 6 wynika, że już dla długości przęsła 35 m i zwisu początkowego 3 m, suma składowych naciągu pochodzących od trzech przewodów osiąga wartość siły niszczącej wybraną konstrukcję wsporczą. Dla porównania, przyjmując obciążenie maksymalne według wytycznych normatywnych, długość przęsła dla słupa E 12/10 mogłaby wynosić 80 m (rys. 7).

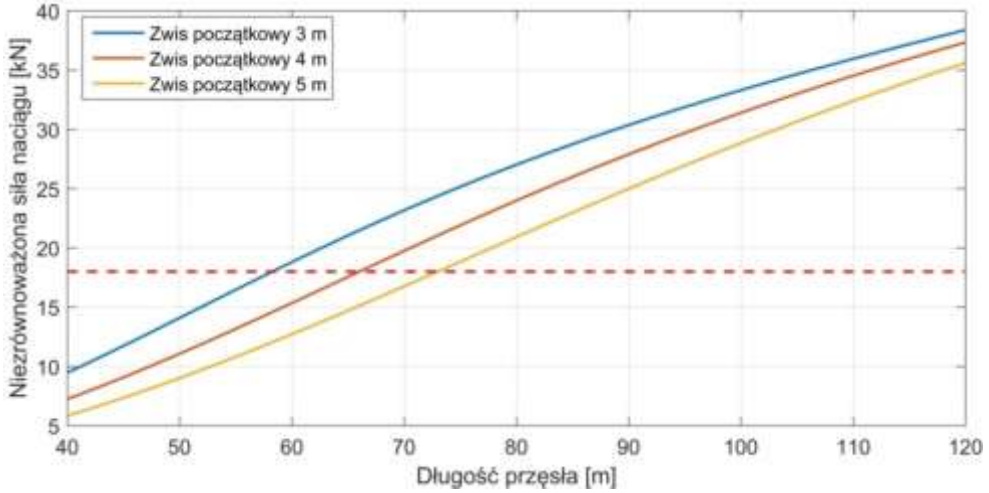
Jeżeli słup pełni rolę słupa przelotowego to dla rozpatrywanych warunków, najbardziej niebezpieczny będzie przypadek 2c. Niezrównoważony wzdłużnie przypadek oblodzenia katastrofalnego będzie zagrożeniem dla przęsła o długości zastępczej wynoszącej 58 m przy zwisie początkowym 3 m (rys. 8).



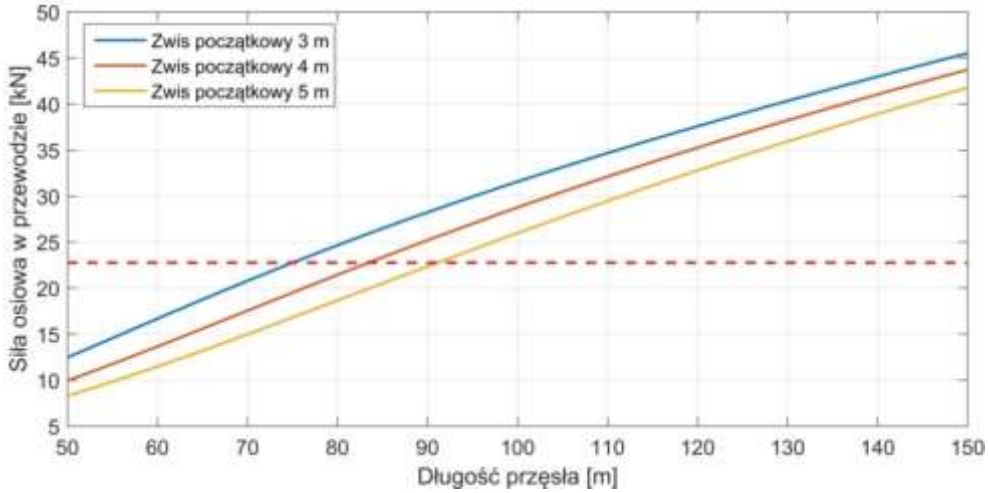
Rys. 6. Zależność jednostronnej siły naciągu w funkcji długości przęsła dla słupa mocnego i oblodzenia katastrofalnego 18 kg/m (przypadek 5c)



Rys. 7. Zależność jednostronnej siły naciągu w funkcji długości przęsła dla słupa mocnego i normatywnego oblodzenia maksymalnego (przypadek 5c)



Rys. 8. Zależność niezrównoważonej siły naciągu w funkcji długości przęsła dla oblodzenia katastrofalnego 18kg/m niezrównoważonego wzdłużnie (przypadek 2c)



Rys. 9. Zależność maksymalnej siły osiowej w pojedynczym przewodzie w funkcji długości przęsła dla oblodzenia katastrofalnego 18kg/m

Oczywiście rozpatrując tak duże oblodzenie należy rozpatrzyć maksymalne siły w przewodach ze względu na możliwość ich zerwania. Dla przewodu AFL-6-70 obliczeniowa siła zrywająca przewód wynosi 22,75 kN. Jeżeli przewód jest obciążony lodem o masie 18 kg/m to dla zwisu początkowego 3 m, może dojść do jego zerwania przy długości przęsła 75 m (rys. 9).

Z analiz wynika, że przy pojawieniu się tak nietypowych warunków atmosferycznych jakie zaistniały w 2010 roku, powstaje oblodzenie o masie, która wykracza daleko poza wytyczne normatywne, nawet dla trzeciej strefy obciążenia oblodzeniem. Słupy typu E, które charakteryzują się dużą wytrzymałością w porównaniu do konstrukcji BSW mogą nie przenieść tak dużych obciążeń i ulec złamaniu. Oczywiście najgorszym przypadkiem dla odcinka linii energetycznej jest sytuacja, gdy przez nadmierne obciążenie dochodzi do zerwania przewodów w jednym z przęseł, a słup jest obciążony jednostronnym naciągiem i poddany dużym siłom gnącym. Są to przypadki z grupy 5 układów oddziaływań. Chociaż dotyczą one słupów mocnych, to może dojść do sytuacji, w której ze względu na duże oblodzenie przewód nie będzie mógł się swobodnie przesunąć w podporze bądź wyczepić z uchwytu zabezpieczającego, tym samym przenosząc siły naciągu na słup przelotowy, doprowadzając do jego zniszczenia.

4. Badania terenowe z wykorzystaniem termowizji

W kolejnym etapie prac przeprowadzono pomiary temperatury przewodów analizowanej linii 15 kV w warunkach normalnego oraz zwiększonego jej obciążenia. Pomiarów dokonano na przewodach fazowych AFL6 – 70. Eksperyment przeprowadzono w dniu 23.11.2017 r. dla układu normalnego oraz specjalnie przygotowanego układu pracy sieci powodującego dociążenie badanej linii.

Na rysunku 10 zaprezentowano zbiorcze zestawienie wyników badań: wartości natężenia prądu, temperatury otoczenia i temperatury przewodu zarejestrowanych w trakcie eksperymentu. Ze względów bezpieczeństwa - zapewnienia ciągłości zasilania oraz zagwarantowania odpowiedniej jakości zasilania - próbę przeprowadzono w porze dziennej.

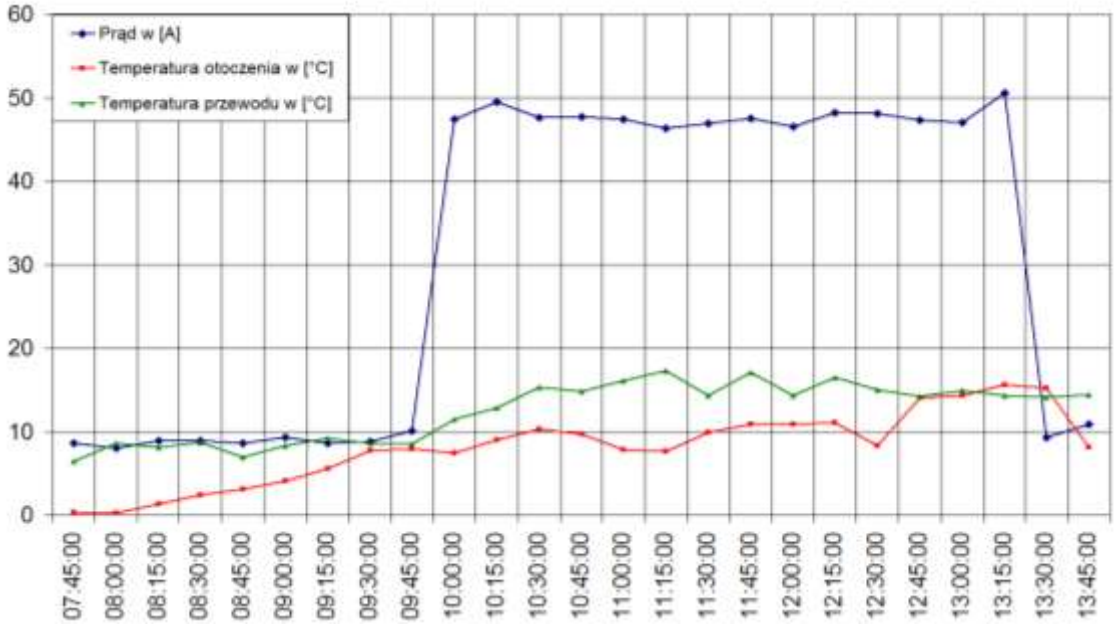
Badania wykonano dla warunków bezwietrznych z niewielkim zachmurzeniem nieba, występującymi chmurami piętra średniego, które pojawiło po godzinie 9.30. Z tego powodu obserwowano podczas eksperymentu zmianę temperatury w zależności od stopnia nasłonecznienia. W początkowej fazie eksperymentów od godziny 7:45 do godziny 9:45 zaobserwowano wzrost temperatury otoczenia z wartości 0,3°C do wartości 7,9°C. W tym okresie wartość natężenia prądu zmieniała się od 8 A do 10 A, a temperatura przewodu od 6,3°C do 9,2°C.

O godzinie 9:52 dociążono linię SN (zmiana konfiguracji sieci), a normalny układ pracy przywrócono o godzinie 13:25. O godzinie 10.00 zaobserwowano prąd o wartości 47 A. Podczas dociążenia wartość natężenia prądu zawierała się w przedziale od 46 A do 51 A. W tym czasie obserwowana temperatura przewodu zmieniała się w przedziale od 11,3°C do 17,3°C. W obserwowanym okresie zmieniała się również wartość temperatury otoczenia w przedziale od 7,4°C do 15,6°C.

Na szczególną uwagę zasługuje okres pomiarowy od 9:45 do 10:45, w którym zaobserwowano największy wzrost temperatury przewodu z wartości 8,5°C do wartości 14,8°C przy wzroście natężenia prądu z wartości 10 A do 48 A, w tym okresie temperatura otoczenia zwiększyła się o 2,3°C.

Stwierdzono, że podczas zmiany obciążenia linii przy wzroście prądu o 36 A do 40 A zaobserwowano wzrost temperatury przewodu w przedziale od 4,3 do 6,1°C.

Oprócz wzrostu temperatury przewodu zaobserwowano również znaczne pogorszenie się warunków napięciowych u odbiorców. Na podstawie analiz całorocznych przebiegów prądów i napięć w przedmiotowej sieci, dokonanych eksperymentów termowizyjnych oraz analiz spadków napięć stwierdzono, że realizacja dodatkowego obciążenia dla warunków zimowych nie spowoduje wzrostu temperatury przewodów do wartości dodatnich (gwarantujących zapobieganie utworzenia się powłoki lodowo-szadziowej). Wykluczona jest również sytuacja wykorzystania podgrzewania magistrali w celu usuwania powłoki lodowej i/lub szadziowej.



Rys. 10. Wyniki pomiarów wykonanych 2311.2017 r dla badanej linii 15 kV w trakcie eksperymentalnego jej dociążenia

5. Strukturalny model niezawodnościowy

Wykonano obliczenia niezawodnościowe prowadzące do wyznaczenia wskaźników SAIDI i SAIFI dla analizowanej linii SN. Zastosowano własny, dedykowany program komputerowy. Wielowariantowe obliczenia z jego pomocą umożliwiły ilościową, niezawodnościową ocenę różnych scenariuszy modernizacji obiektu.

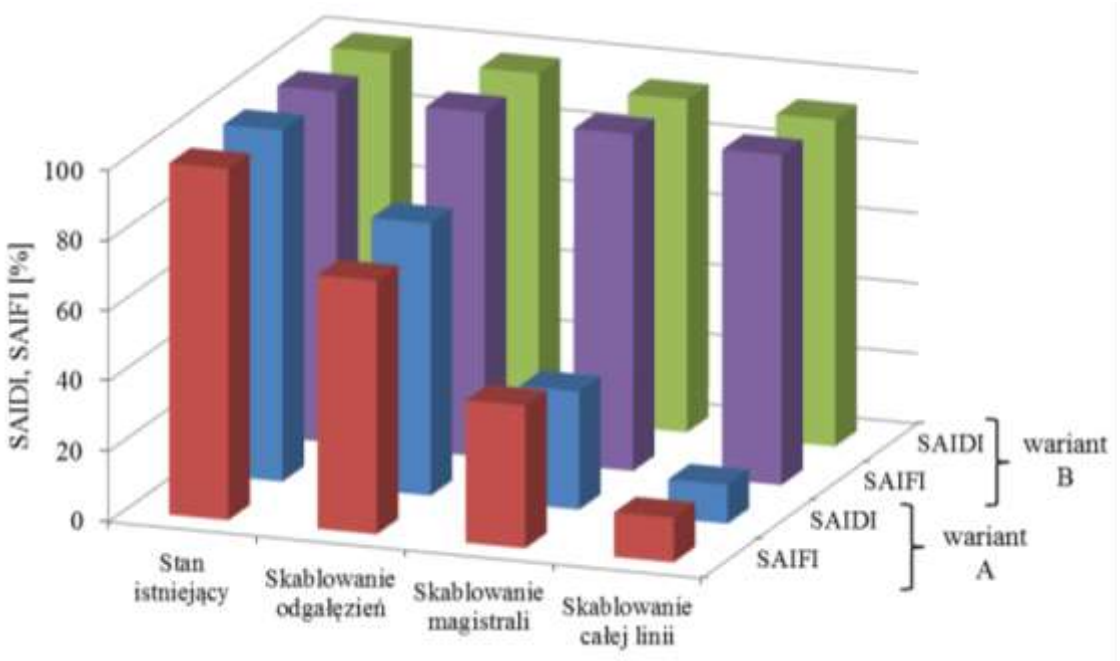
Przeprowadzenie symulacji wymaga danych dotyczących struktury linii wraz ze szczegółowymi danymi odcinkowymi, jak również parametrów niezawodnościowych elementów składowych.

W zakresie parametrów niezawodnościowych elementów przyjęto jednakowe wartości średnich czasów trwania przerw w zasilaniu. Dla częstości uszkodzeń zrealizowano dwa warianty obliczeń A i B. W wariancie A wykorzystano relację parametrów wg publikacji [3] oraz [21] natomiast w wariancie B wskaźniki zaczerpnięto z [12]. Parametry niezawodnościowe zestawiono w tabeli 3.

Tabela 3. Przyjęte parametry zawodnościowe sieci rozdzielczej SN

Element	**Jednostka** ***j***	**Częstość uszkodzeń** ***d*** **[1/(*j*·rok)] wariant A**	**Częstość uszkodzeń** ***d*** **[1/(*j*·rok)] wariant B**	**Średni czas trwania uszkodzenia** t_a **[h]**
Linia napowietrzna SN (AFL)	100 km	8,14	11,02	5
Linia kablowa SN		0,814	10,16	5

Wyniki obliczeń wskaźników niezawodności SAIDI i SAIFI przedstawiono na rysunku 11. Wartości wskaźników dla stanu istniejącego stanowią poziom referencyjny dla wszystkich symulowanych przypadków.



Rys. 11. Spodziewany stopień redukcji wskaźników niezawodności dla różnych wariantów przebudowy analizowanej linii 15 kV

Dla parametrów w wariancie B sukcesywne kablowanie badanej linii w minimalnym stopniu wpływa na poprawę niezawodności. Dla parametrów w wariancie A obserwujemy istotną poprawę niezawodności obiektu na skutek jego przebudowy i modernizacji. Ten zestaw parametrów wydaje się bardziej poprawny w kontekście realiów eksploatacyjnych sieci dystrybucyjnych.

6. Podsumowanie i wnioski

W pracy podjęto niezwykle istotny problem eksploatacji linii średniego napięcia narażonych na katastrofalne oblodzenia. W Polsce linie takie zlokalizowane są na obszarach działania czterech operatorów systemów dystrybucyjnych, a ich długość można szacować na ponad 20 tys. km.

Celem pracy było wskazanie środków poprawy niezawodności analizowanego obiektu, przeciwdziałających skutkom katastrofalnego oblodzenia. Wykonano badania modelowo-symulacyjne MES, eksperymentalne termowizyjne pomiary temperatury przewodów oraz przeprowadzono wielowariantowe obliczenia wskaźników SAIDI i SAIFI.

Badania symulacyjno-modelowe MES prowadzą do konkluzji, że analizowany obiekt jest w stanie wytrzymać obciążenia katastrofalne pod warunkiem zredukowania rozpiętości przęseł do długości 35 metrów.

Przeprowadzony eksperyment, polegający na planowym dociążeniu linii, oraz wykonane pomiary termowizyjne wskazują, że skuteczne wykorzystanie prewencyjnego podgrzewania magistrali dodatkowym prądem nie jest w tym przypadku możliwe. Obciążenie linii w podstawowym układzie pracy jest znikome, a rekonfiguracyjne dociążenie ciągu jest istotnie ograniczone warunkami napięciowymi.

Zasadnicze wyeliminowanie oddziaływań klimatycznych (zarówno zimowych jak i letnich) może być osiągnięte przez przebudowę analizowanego obiektu na linię kablową. Poza efektem pełnego ekranowania klimatycznego wpłynie to niewątpliwie na niezawodność eksploatacyjną obiektu, a stopień tego wpływu będzie istotnie uwarunkowany parametrami niezawodnościowymi elementów sieci.

Wobec znaczącej przewagi sieci napowietrznych w krajowej sieci rozdzielczej niezbędne jest opracowanie strategii postępowania z napowietrzną infrastrukturą dystrybucyjną eksponowaną na oddziaływania ekstremalnych zjawisk atmosferycznych. Taka strategia powinna brać pod uwagę szereg czynników takich jak: stan technicznych, stopień zamortyzowania majątku, uwarunkowania lokalizacyjne i funkcje poszczególnych elementów w sieci dystrybucyjnej.

Reasumując, należy stwierdzić, że wytyczne zaprezentowane w niniejszej pracy pozwalają na zwiększenie niezawodności linii elektroenergetycznych narażonych na katastrofalne oblodzenie i powinny być rozważone oraz wdrożone przez wszystkich operatorów systemów dystrybucyjnych w Polsce.

Literatura:

1. Admirat P, Sakamoto Y. Calibration of a wet snow model on real cases in Japan and France. 4th International Workshop on Atmospheric Icing of Structures (IWAIS), Paris, France 1988; 7-13.

2. Ciesielka W, Czajka I, Filipek R, Gołaś A, Hamiga W, Romik D, Suder-Dębska K, Szopa K, Wołoszyn J. Smart Grid in energetic facilities: modelling, monitoring and diagnostics. Monography of the Department of Power Systems and Environmental Protection. Faculty of Mechanical Engineering and Robotics AGH, Krakow 2017.

3. Damchi Y, Sadeh J. Effect of combined transmission line (overhead line/cable) on power system reliability indices. 4th International Power Engineering and Optimization Conference (PEOCO), Shah Alam, Malaysia 2010; 59–63.

4. Elíasson A J, Thorsteins E, Ólafsson H. Study of wet snow events on the South Coast of Iceland. 9th International Workshop on Atmospheric Icing of Structures (IWAIS), Chester, United Kingdom 2000.

5. Farzaneh M, Savadjiev K. Icing Events Occurrence in Québec: Statistical analysis of field data. International Journal of Offshore and Polar Engineering 2001; 11(1): 9–15.

6. Farzaneh M. Atmosferic Icing of Power Networks. Springer Science+Business Media B.V., 2008.

7. Fikke S M, Johansen O S. Earlier Norwegian iceload research. A review of investigations and results. 2nd International Workshop on Atmospheric Icing of Structures (IWAIS), Trondheim, Norway 1984; 11-18.

8. Fikke S M et al. COST Action 727 Atmospheric icing on structures. Measurements and data collection on icing. State of the art. Veröffentlichung MeteoSchweiz 2007; 75.

9. Goia M L. Damages caused by icing and wind to the Romanian OEL. 9th International Workshop on Atmospheric Icing of Structures (IWAIS), Chester, United Kingdom 2000.

10. Golikova T N, Toporkava G D, Nikitina L G. Ascertaining ice-load maps of the USSR territory. Trans Improving the reliability of high voltage lines. Energoatomizdat, Moscow 1989; 107–122.

11. Gołaś A, Ciesielka W, Czajka I, Czechowski M, Filipek R, Suder-Dębska K, Szopa K, Śliwiński M, Wołoszyn J, Żywiec W. Mechanical engineering in Smart Grid technology. Monography of the Department of Power Systems and Environmental Protection. Faculty of Mechanical Engineering and Robotics AGH, Krakow 2015.

12. Kornatka M. Analysis of the exploitation failure rate in Polish MV networks. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20(3): 413–419.

13. Krómer I. Hungarian icing activity survey. 6th International Workshop on Atmospheric Icing of Structures (IWAIS), Budapest, Hungary 1993; ix-x.

14. Lehtonen P, Ahti K, Makkonen L. The growth and disappearance of ice loads on a tall mast. 3rd International Workshop on Atmospheric Icing of Structures (IWAIS), Vancouver, Canada 1986; 363–368.

15. Overhead Lines – Meteorological Data for Assessing Climatic Loads, 1997; International Electrotechnical Commission Technical Report 61774, First edition: 1997–2008.

16. PN-EN 50341-2-22:2016-04 Elektroenergetyczne linie napowietrzne prądu przemiennego powyżej 1 kV -- Część 2-22: Krajowe Warunki Normatywne (NNA) dla Polski.

17. Popolansky F. Economical aspects of ice failures caused in power transmission on the territory of former Czechoslovakia. 9th International Workshop on Atmospheric Icing of Structures (IWAIS), Chester, United Kingdom 2000.

18. Technical Brochure CIGRE - Guidelines for field measurement of ice loadings on power line conductors, 2001; CIGRE TB No 179.

19. Technical Brochure CIGRE - Big storm events. What we have learned?, 2008; CIGRE TB No 344.

20. Wareing B J, Chetwood P. Ice load data from Deadwater Fell. 9th International Workshop on Atmospheric Icing of Structures (IWAIS), Chester, United Kingdom 2000.

21. Zhu D, Broadwater R P, Tam K, Seguin R, Asgeirsson H. Impact of DG placement on reliability and efficiency with time-varying loads, IEEE Transactions on Power Systems 2006; 21(1): 419–427.

Grzegorz Mieczkowski

Faculty of Mechanical Engineering
Bialystok University of Technology
ul. Wiejska 45C, Poland
E-mail: g.mieczkowski@pb.edu.pl

Kryterium pękania struktury bi-materiałowej z ostrym karbem usytuowanym na interfejsie

Słowa kluczowe: *inicjacja procesu pękania, bi-materiały, osobliwe pola naprężeń, współczynniki intensywności naprężeń*

Streszczenie: W pracy przedstawiono wyniki badań dotyczących pękania struktury bi-materiałowej z karbem usytuowanym na interfejsie. Do prognozowania inicjacji procesu pękania zastosowano kryterium oparte na punktowej teorii krytycznych dystansów. Analizowano elementy wykonane ze stopu aluminium i polimerów (PC, PMMA), które poddane były trójpunktowemu zginaniu. Wartości obciążeń krytycznych wynikających z wykorzystanej hipotezy porównano z wartościami uzyskanymi z eksperymentu. Walidacja wybranego kryterium wymagała określania jakościowego i ilościowego opisu osobliwych pól naprężeń, występujących w okolicy wierzchołkowej karbu strukturalnego. W związku z tym, uzyskano takie rozwiązania i omówiono metodykę ich otrzymywania.

1. Wstęp

Prognozowanie trwałości konstrukcji mechanicznych jest złożonym procesem wymagającym uwzględnienia wielu czynników. Jednym z nich jest złożoność strukturalna konstrukcji. Konstrukcja jest obiektem fizycznym zbudowanym z wielu elementów, często wykonanych z materiałów o różnych właściwościach mechano–fizycznych. Dodatkowo, konstrukcje poddane są zwykle działaniu złożonych obciążeń zewnętrznych, często o charakterze zmiennym, przy różnych czynnikach środowiskowych. Powoduje to różne mechanizmy zniszczenia - zużycie (korozyjne, kawitacyjne, tarciowe), utratę nośności, pękanie – wykluczające często dalszą pracę urządzenia.

Miejscem inicjacji pęknięć są najczęściej pustki i inne wady materiałowe oraz karby konstrukcyjne, których obecność w konstrukcji wynika najczęściej z konieczności połączenia ze sobą poszczególnych komponentów, zapewnienia konstrukcji wymaganych cech funkcjonalnych, czy żądanej struktury materiałowej.

Karby można podzielić na wiele różnych kategorii w zależności od ich kształtu, usytuowania w konstrukcji, niejednorodności materiałowej czy technologii wytwarzania konstrukcji. Wszystkie one mają wspólną cechę – powodują lokalny wzrost naprężenia w obciążonej konstrukcji, przez co wpływają na jej trwałość i wytrzymałość.

Na podstawie obserwacji i badań eksperymentalnych, naukowcy od dawna próbowali określić pewne krytyczne warunki, przy których następuje zniszczenie materiału. Można tu wymienić np. idee Galileusza, Tresci, Beltramiego, Coulomba, Mohra, Misesa czy Hubera. Przyjęli oni pewne hipotezy, określające wytrzymałość materiałów, stanowiące do dziś podstawy obliczeń inżynierskich. Zakładając jednorodność i ciągłość ośrodka, z którego został wykonany element konstrukcyjny, sformułowali graniczne wartości funkcji, po której przekroczeniu następowało zniszczenie materiału. Hipotezy te nie uwzględniały znaczącego wpływu silnych gradientów pól naprężeń i odkształceń na wytrzymałość. Rozwój teoretyczny

mechaniki ciała stałego, szczególnie teorii sprężystości, pozwolił na uzyskanie wielu rozwiązań analitycznych opisujących lokalne pola naprężeń w otoczeniu zarówno łagodnego koncentratora (np. rozwiązania Kirscha [8]), jak i koncentratorów ostrych, generujących osobliwe pola naprężeń (rozwiązania Sneddona [31], Williamsa [38]). Znajomość nowych rozwiązań matematycznych umożliwiła sformułowanie kolejnych kryteriów wytrzymałościowych uwzględniających wpływ obecności koncentratorów naprężeń w jednorodnych materiałach konstrukcyjnych (np. Griffith [6], Sih [295], McClintock [16] i inni).

W ostatnich latach zauważalny jest silny rozwój materiałów kompozytowych o z góry zaprojektowanych właściwościach mechanicznych. Są to z reguły materiały anizotropowe lub kompozyty o złożonej strukturze periodycznej, często zawierające pęknięcia, wtrącenia lub inne wady wewnętrzne wywołujące efekty miejscowego spiętrzenia naprężeń.

Typowym koncentratorem naprężeń występującym w kompozytach warstwowych jest pęknięcie [30] lub ostry karb usytuowany w płaszczyźnie łączenia poszczególnych warstw kompozytu [3, 5, 26]. Tego typu koncentratory występują także często w elementach konstrukcyjnych powstałych przez połączenie dwóch różnych materiałów za pomocą klejenia (taki element można traktować jako swoisty materiał kompozytowy). Pojawia się zatem potrzeba określenia wytrzymałości i odporności na pękanie kompozytów (z uwzględnieniem właściwości mechanicznych warstwy łączącej), w których karby strukturalne generują duże gradienty naprężeń. Rozwiązaniem tego problemu jest odpowiednio sformułowane kryterium wytrzymałościowe. Kryterium powinno zawierać precyzyjnie określone równanie wraz ze zdefiniowanymi stałymi materiałowymi, na podstawie, którego można przewidzieć moment zainicjowania procesu pękania. Prognozowaniem trwałości elementów z karbami strukturalnymi zajmowało się stosunkowo niewielu naukowców.

W pracy [12] analizowano wytrzymałość elementów dwufazowych z karbem strukturalnym. Komponenty materiałowe połączone były adhezyjnie. Autorzy wykonali badania trójpunktowego zginania i wyznaczyli wartości sił inicjujących pękanie. Możliwość zastosowania kryterium Leguillona dla tego typu elementów (elementy z karbem łączone adhezyjnie) została pozytywnie zweryfikowana w pracy [36]. Kryterium, które jest często stosowane dla materiałów jednorodnych i nie wymaga wyznaczania tak dużej ilości stałych materiałowych, jest kryterium MClintocka. Zatem głównym celem prezentowanej pracy jest eksperymentalna weryfikacja możliwości stosowania tego kryterium (z zastosowaniem właściwej modyfikacji) dla struktur bi-materiałowych, w których ostre karby strukturalne generują osobliwe pola naprężeń. Ideę kryterium opisano poniżej.

Nomenklatura

a - Wysokość karbu

b - Gradient ‘naprężeń kombinowanych’

E - Moduł Younga

f_{ik}^{I}, f_{ik}^{II} -Współczynniki wpływu dla naprężeń

F - Obciążenie, przy którym obliczano uogólnione współczynniki intensywności naprężeń

F_k - Prognozowane obciążenie niszczące

g - Grubość próbki

h - Wysokość próbki

H_o, H_1, H_2 - Współczynniki wpływu dla równania charakterystycznego

i - Indeks materiału (=1,2)

K_E - Ekwiwalentny współczynnik intensywności naprężeń

K_{Ec} - Wartość krytyczna ekwiwalentnego współczynnika intensywności naprężeń

K_I, K_{II} -Uogólnione współczynniki intensywności naprężeń

K_{Ic} - Odporność na pękanie
L - Odległość między podporami w próbie trójpunktowego zginania
Lc -Długość próbki
n - Indeks węzłów
r, φ -Współrzędne w biegunowym układzie odniesienia
u_r, u_φ - Przemieszczenia w biegunowym układzie odniesienia
α - Kąt pomiędzy interfejsem, a krawędzią karbu w materiału 1
β - Kąt wierzchołkowy karbu
γ - Kąt pomiędzy interfejsem, a krawędzią karbu w materiału 2
Γ - Proporcja modułów odkształcenia postaciowego
δ - Część urojona wartości własnej macierzy warunków brzegowych λ
λ - Wartość własna macierzy warunków brzegowych
λ_r - Część rzeczywista wartości własnej macierzy warunków brzegowych λ
μ - Moduł odkształcenia postaciowego
ν - Współczynnik Poissona
$\sigma_\varphi, \sigma_r, \tau_{r\varphi}$ -Naprężenia w biegunowym układzie odniesienia
φ_0 - Kąt propagacji pęknięcia
ψ -Proporcja naprężeń stycznych i normalnych

2. Kryterium pękania

W kryterium zaproponowanym w pracy [16] przyjmuje się, że pękanie nastąpi w przypadku, gdy odkształcenie normalne ε_φ w pewnej małej odległości od wierzchołka szczeliny ρ_c osiągnie wartość krytyczną, co można zapisać w następujący sposób:

$$\varepsilon_\varphi(\rho_c) = \varepsilon_c \tag{1}$$

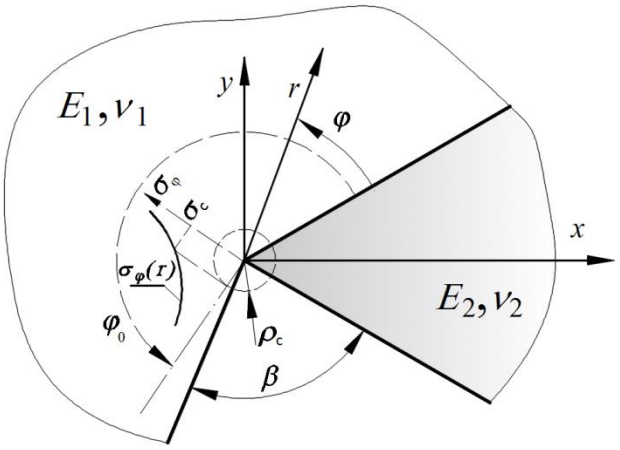
Szersze zastosowanie znalazła jednak postać naprężeniowa tego kryterium, w której odkształcenie zostało zastąpione przez odpowiednią składową naprężeń normalnych.

W podejściu zaproponowanym w pracy [23] przyjęto, że propagacja pęknięcia nastąpi wtedy, gdy naprężenia obwodowe σ_φ w pewnej skończonej odległości $r=\rho_c$ osiągną wartość krytyczną σ_c (2). Stosowanie tak sformułowanego warunku, dla elementów z karbami występującymi w materiałach jednorodnych, zostało pozytywnie zweryfikowane w wielu pracach np. [27].

$$\max_\varphi \sigma_\varphi(\rho_c) = \sigma_c \tag{2}$$

Kąt propagacji pęknięcia φ_0 wyznacza się maksymalizując σ_φ względem kąta φ. W przypadku zagadnienia bi-materiału ze strukturalnym karbem, początkowo pęknięcie propaguje wzdłuż interfejsu. Zatem można z góry założyć, że $\varphi_0 = 0$ (rys.1) i warunek (2) można zapisać w następujący sposób:

$$\sigma_\varphi(\rho_c, 0) = \sigma_c \tag{3}$$



Rys. 1 Graficzna interpretacja kryterium McClintocka.

Parametr ρ_c, traktowany jest jako stała materiałowa i może być wyznaczony doświadczalnie. Na przykład, wykorzystując w warunku (3) zależność na naprężenia obwodowe przy wierzchołku szczeliny oraz kryterium Griffitha – Irwina otrzymuje się:

$$\frac{K_c}{\sqrt{2\pi\rho_c}} = \sigma_c \tag{4}$$

skąd można obliczyć krytyczny dystans:

$$\rho_c = \frac{1}{2\pi}\left(\frac{K_c}{\sigma_c}\right)^2, \tag{5}$$

gdzie K_c - odporność na pękanie, σ_c -: wytrzymałość na rozciąganie.

Zaletą proponowanej metody, opartej na tak zwanej teorii krytycznych dystansów (TDC), jest to, że w większości przypadków do prognozowania procesu pękania nie jest wymagana znajomość analitycznego opisu pól naprężeń- prognozowanie opiera się na rozwiązaniach numerycznych.

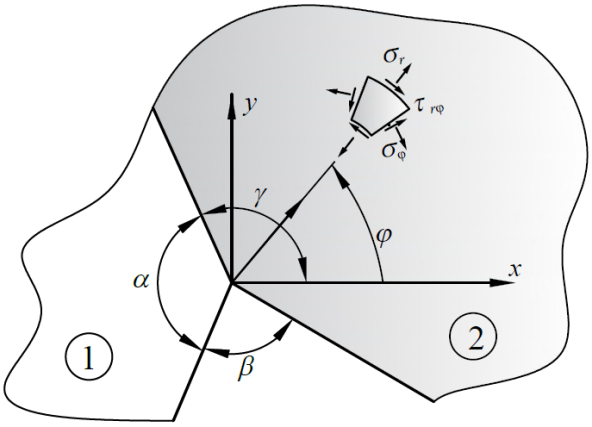
W przypadku, gdy proces pękania odbywa się w płaszczyźnie, w której występuje złożony stan naprężeń, zastosowanie rozwiązań numerycznych może spowodować błędną prognozę [13]. Najczęściej w takiej sytuacji, na podstawie analitycznego opisu lokalnych pól naprężeń, formułowane jest globalne kryterium pękania (przy użyciu lokalnego parametru) oparte na współczynniku intensywności naprężeń [27, 13, 1] lub współczynniku uwalniania energii [9].

W prezentowanej pracy, na podstawie kryterium McClintocka, zaproponowano dwie koncepcje przewidywania inicjacji procesu pękania. Pierwsza z nich oparta jest na ekwiwalentnym współczynniku intensywności naprężeń, a druga na uzależnieniu (w warunku (2)) naprężeń krytycznych od proporcji naprężeń stycznych i normalnych występujących w płaszczyźnie pękania. Szczegółowy opis koncepcji przedstawiono w rozdziale 6.

Jak łatwo zauważyć, zastosowanie kryterium McClintocka wymaga znajomości dystrybucji pól naprężeń występujących w bliskim sąsiedztwie punktu osobliwego. W związku z tym w kolejnym rozdziale pracy podane zostaną postacie funkcji opisujących takie pola naprężeń oraz metodologia ich wyznaczania.

3. Analityczne zależności opisujące pola naprężeń w okolicy wierzchołka karbu strukturalnego.

Rozwiązanie zagadnienia bi-materiału z karbem strukturalnym usytuowanym na interfejsie (rys.2) otrzymano wykorzystując podejście zastosowane przez autorów pracy [21] dla przypadku ostrego naroża w materiale jednorodnym.



Rys. 2 Bi-materiał z karbem strukturalnym usytuowanym na interfejsie

W wykorzystanej metodzie, której dokładny opis można znaleźć w pracy [19], stosując funkcję naprężeń Airy'ego otrzymuję się ogólne rozwiązania asymptotyczne opisujące poszczególne komponenty pól naprężeń i przemieszczeń. Dla zagadnienia karbu usytuowanego na interfejsie bi-materiału rozwiązanie ogólne przyjmuje następująca postać [18]:

$$
\begin{aligned}
u_{ri} &= r^{\lambda}\left(A_i\cos((1+\lambda)\varphi)+B_i\sin((1+\lambda)\varphi)+C_i\cos((1-\lambda)\varphi)+D_i\sin((1-\lambda)\varphi)\right)\\
u_{\varphi i} &= r^{\lambda}\left(-A_i\sin((1+\lambda)\varphi)+B_i\cos((1+\lambda)\varphi)-C_i\frac{\kappa+\lambda}{\kappa-\lambda}\sin((1-\lambda)\varphi)+D_i\frac{\kappa+\lambda}{\kappa-\lambda}\cos((1-\lambda)\varphi)\right)\\
\sigma_{ri} &= r^{\lambda-1}\mu\left(A_i 2\lambda\cos((1+\lambda)\varphi)+B_i 2\lambda\sin((1+\lambda)\varphi)+C_i(3-\lambda)\frac{2\lambda}{\kappa-\lambda}\cos((1-\lambda)\varphi)+D_i(3-\lambda)\frac{2\lambda}{\kappa-\lambda}\sin((1-\lambda)\varphi)\right)\\
\sigma_{\varphi i} &= r^{\lambda-1}\mu\left(-A_i 2\lambda\cos((1+\lambda)\varphi)-B_i 2\lambda\sin((1+\lambda)\varphi)+C_i(1+\lambda)\frac{2\lambda}{\kappa-\lambda}\cos((1-\lambda)\varphi)+D_i(1+\lambda)\frac{2\lambda}{\kappa-\lambda}\sin((1-\lambda)\varphi)\right)\\
\tau_{r\varphi i} &= r^{\lambda-1}\mu\left(-A_i 2\lambda\sin((1+\lambda)\varphi)+B_i 2\lambda\cos((1+\lambda)\varphi)+C_i(1-\lambda)\frac{2\lambda}{\kappa-\lambda}\sin((1-\lambda)\varphi)-D_i(1-\lambda)\frac{2\lambda}{\kappa-\lambda}\cos((1-\lambda)\varphi)\right).
\end{aligned}
\tag{6}
$$

gdzie: $\mu_i=\dfrac{E_i}{2(1+\nu_i)}$ -moduł odkształcenia postaciowego, $\kappa_i=(3-\nu_i)/(1+\nu_i)$ - płaski stan naprężenia, $\kappa_i=(3-4\nu_i)$ - płaski stan odkształcenia, ν_i -współczynnik Poissona, i=1,2.

Rozwiązanie szczególne otrzymuje się poprzez wyznaczenie wykładnika potęgowego λ oraz stałych A_i, B_i, C_i, D_i. Stałe wyznaczono na podstawie następujących warunków brzegowych [18]:

1. dla lewej powierzchni bocznej karbu, $\varphi = \alpha$;

$$\sigma_{\varphi 1}=\tau_{r\varphi 1}=0$$

2. dla prawej powierzchni bocznej karbu $\varphi = -\gamma$;

$$\sigma_{\varphi 2}=\tau_{r\varphi 2}=0$$

3. wzdłuż interfejsu, $\varphi = 0$;

$$u_{r1}=u_{r2};u_{\varphi 1}=u_{\varphi 2};\ \sigma_{\varphi 1}=\sigma_{\varphi 2};\tau_{r\varphi 1}=\tau_{r\varphi 2}.$$

Ponadto z warunku zerowania się wyznacznika macierzy warunków brzegowych wyznaczono równanie charakterystyczne (7), którego kolejne pierwiastki wyznaczają wartość wykładnika potęgowego λ (wartości własne macierzy warunków brzegowych) w uzyskanych rozwiązaniach asymptotycznych (6)[18].

$$H_0 + \Gamma H_1 + \Gamma^2 H_2 = 0 \tag{7}$$

gdzie:

$$H_o = \left(1-2\lambda^2+2\lambda^2\cos[2\alpha]\right)\left(1-\lambda^2+\lambda^2\cos[2\gamma]-\cos[2\gamma\lambda]\right)$$

$$-\kappa_1\begin{pmatrix}\cos\left[2(-\alpha+\gamma)\lambda\right]+\cos\left[2(\alpha+\gamma)\lambda\right]+\cos[2\alpha\lambda]\left(-2+4\lambda^2\sin[\gamma]^2\right)\\+2\left(\lambda^2\sin[\gamma]^2-\sin[\gamma\lambda]^2\right)\kappa_1\end{pmatrix},$$

$$H_1 = 5\lambda^2+\cos[2\alpha\lambda]+\cos[2\gamma\lambda]-2\cos\left[(-\alpha+\gamma)\lambda\right]^2-$$

$$\lambda^2\left(3\cos[2\alpha]-\cos\left[2(\alpha-\gamma)\right]+3\cos[2\gamma]+4\cos[2\gamma\lambda]\sin[\alpha]^2+4\left(\cos[2\alpha\lambda]+4\lambda^2\sin[\alpha]^2\right)\sin[\gamma]^2\right)+$$

$$\kappa_2\left(\cos\left[2(\alpha+\gamma)\lambda\right]-\cos[2\alpha\lambda]+2\sin[\gamma\lambda]^2-\lambda^2\left(1+\left(\cos[2\alpha]+4\cos[2\gamma\lambda]\sin[\alpha]^2-2\sin[\alpha]\sin[\alpha-2\gamma]\right)\right)\right)+$$

$$\kappa_1\begin{pmatrix}\cos\left[2(\alpha+\gamma)\lambda\right]+4\lambda^2\cos[\alpha]\sin[\alpha-\gamma]\sin[\gamma]+\cos[2\alpha\lambda]\left(-1+4\lambda^2\sin[\gamma]^2\right)+2\sin[\gamma\lambda]^2\\+\left(\cos[2\alpha\lambda]+\cos[2\gamma\lambda]-2\cos\left[(-\alpha+\gamma)\lambda\right]^2+4\lambda^2\cos[\alpha-\gamma]\sin[\alpha]\sin[\gamma]\right)\kappa_2\end{pmatrix},$$

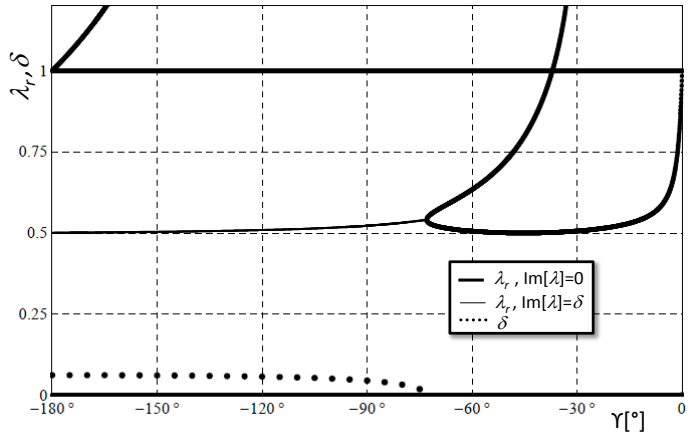
$$H_2 = \left(1-2\lambda^2+2\lambda^2\cos[2\gamma]\right)\left(1-\lambda^2+\lambda^2\cos[2\alpha]-\cos[2\alpha\lambda]\right)-$$

$$\kappa_2\begin{pmatrix}\cos\left[2(-\alpha+\gamma)\lambda\right]+\cos\left[2(\alpha+\gamma)\lambda\right]+\cos[2\gamma\lambda]\left(-2+4\lambda^2\sin[\alpha]^2\right)\\+2\left(\lambda^2\sin[\alpha]^2-\sin[\alpha\lambda]^2\right)\kappa_2\end{pmatrix},$$

$$\Gamma = \frac{\mu_1}{\mu_2}.$$

Z uzyskanego równania charakterystycznego (7) można wywnioskować, że wykładnik potęgowy zależy od stałych materiałowych i kąta wierzchołkowego karbu. Pierwiastki równania (7) nie mogą być wyznaczone analitycznie. Wyznaczono je numerycznie. W tym celu wykorzystano program obliczeniowy napisany w środowisku Mathematica 9.0.

Na rysunku 3 przedstawiono graficznie wartości własne równania (7), gdzie linią grubą zaznaczono rzeczywiste wykładniki λ (λ= λ_r, Im[λ]=0), linią cienką rzeczywiste wartości zespolonych wykładników λ (λ_r=Re[λ], Im[λ]=δ), natomiast linią punktową części urojone zespolonych wykładników λ (δ = Im[λ]).



Rys. 3. Rozwiązanie równania charakterystycznego (7) dla Γ = 0.033, α=180˚, ν_1=0.37, ν_2=0.35, (płaski stan odkształcenia)

Z otrzymanego rozwiązania wynika, że w zależności od stałych materiałowych i geometrii karbu, istnieje jeden lub więcej członów osobliwy rozwiązania asymptotycznego o wykładniku rzeczywistym lub zespolonym. Co więcej, warto zauważyć, że dla problemu

karbu znajdującego się na interfejsie, równania charakterystyczne nie mogą być uzyskane niezależnie dla I i II sposobu obciążania [3, 5]

Otrzymano również wzory analityczne opisujące składniki pola naprężeń w okolicy wierzchołkowej. Ponieważ naprężenia mogą być opisane zespolonym wykładnikiem λ, uogólnione współczynniki intensywności naprężeń zdefiniowano podobnie jak autorzy pracy [32] (dla zagadnienia szczeliny międzyfazowej):

$$\left(\sigma_{\varphi}+\mathrm{i}\,\tau_{r\varphi}\right)_{\varphi=0}=\frac{K_I+iK_{II}}{\sqrt{2\pi}r^{1-\lambda_r}}\left(\frac{r}{2a}\right)^{i\delta}cosh[\pi\delta], \tag{8}$$

gdzie wymiar *a* można traktować jako np. wysokość karbu.

Wykorzystując równanie (6), warunki brzegowe i przyjętą uogólnioną definicję współczynników intensywności naprężeń (8), można uzyskać analityczny opis pól naprężeń występujących w obszarze wierzchołka karbu [18]:

$$\sigma_{ik}=\frac{cosh[\pi\delta]\sqrt{K_I{}^2+K_{II}{}^2}}{\sqrt{2\pi}r^{1-\lambda_r}}\left(\sin\left[arctan\left[\frac{K_{II}}{K_I}\right]+\delta\log\left[\frac{r}{2a}\right]\right]Re\left[f_{ik}^{I}\right]+cos\left[\frac{K_{II}}{K_I}+\delta\,log\left[\frac{r}{2a}\right]\right]Re\left[f_{ik}^{II}\right]\right), \tag{9}$$

gdzie:

$$f_{rr}^{\mathrm{I}}=M^{-1}\begin{pmatrix}(\lambda-1)cos[(1+\lambda)\varphi](\lambda\,sin[2\epsilon]-sin[2\epsilon\lambda])-(1+\lambda)(\lambda-\lambda cos[2\epsilon]+cos[2\epsilon\lambda]-1)sin[(1+\lambda)\varphi]\\+(\lambda-3)(\lambda\,sin[2\epsilon+(\lambda-1)\varphi]+(1+\lambda)sin[\varphi-\lambda\varphi]-sin[2\epsilon\lambda+\varphi-\lambda\varphi])\end{pmatrix},$$

$$f_{rr}^{II}=-M^{-1}\begin{pmatrix}(1-\lambda)(\lambda cos[2\epsilon]+cos[2\epsilon\lambda]-1-\lambda)cos[(1+\lambda)\varphi]+\\(\lambda-3)(-\lambda cos[2\epsilon+(\lambda-1)\varphi]+(\lambda-1)cos[\varphi-\lambda\varphi]+cos[2\epsilon\lambda+\varphi-\lambda\varphi])-\\+(\lambda-1)(\lambda\,sin[2\epsilon]+sin[2\epsilon\lambda])sin[(1+\lambda)\varphi]\end{pmatrix},$$

$$f_{\varphi\varphi}^{I}=M^{-1}(1+\lambda)\begin{pmatrix}(1-\lambda)sin[(1+\lambda)\varphi]+\lambda\,sin[2\epsilon+(\lambda-1)\varphi]+\\(1+\lambda)sin[\varphi-\lambda\varphi]-sin[2\epsilon\lambda+\varphi-\lambda\varphi]-\lambda\,sin[2\epsilon-(1+\lambda)\varphi]+sin[2\epsilon\lambda-(1+\lambda)\varphi]\end{pmatrix},$$

$$f_{\varphi\varphi}^{II}=-M^{-1}\begin{pmatrix}(\lambda-1)(-1-\lambda+\lambda cos[2\epsilon]+cos[2\epsilon\lambda])cos[(1+\lambda)\varphi]+(\lambda-1)(\lambda\,sin[2\epsilon]+sin[2\epsilon\lambda])sin[(1+\lambda)\varphi]\\+(1+\lambda)(\lambda cos[2\epsilon+(\lambda-1)\varphi]-\lambda cos[\varphi-\lambda\varphi]+2sin[\epsilon\lambda]sin[\epsilon\lambda+\varphi-\lambda\varphi])\end{pmatrix},$$

$$f_{r\varphi}^{I}=M^{-1}\begin{pmatrix}(1+\lambda)(1-\lambda+\lambda cos[2\epsilon]-cos[2\epsilon\lambda])cos[(1+\lambda)\varphi]+\\(\lambda-1)(\lambda cos[2\epsilon+(\lambda-1)\varphi]-(1+\lambda)cos[\varphi-\lambda\varphi]+cos[2\epsilon\lambda+\varphi-\lambda\varphi])+\\+(1+\lambda)(\lambda\,sin[2\epsilon]-sin[2\epsilon\lambda])sin[(1+\lambda)\varphi]\end{pmatrix},$$

$$f_{r\varphi}^{II}=M^{-1}(\lambda-1)\begin{pmatrix}(1+\lambda)sin[(1+\lambda)\varphi]-\lambda\,sin[2\epsilon+(\lambda-1)\varphi]\\-(\lambda-1)sin[\varphi-\lambda\varphi]-sin[2\epsilon\lambda+\varphi-\lambda\varphi]+\lambda\,sin[2\epsilon-(1+\lambda)\varphi]+sin[2\epsilon\lambda-(1+\lambda)\varphi]\end{pmatrix},$$

$M=2\left(\lambda^2-\lambda^2 cos[2\epsilon]+cos[2\epsilon\lambda]-1\right),\ \epsilon=\alpha$ -dla materiału 1 i $\epsilon=-\gamma$ dla materiału 2.

Poniżej podano szczególną postać pól naprężeń dla kąta φ=0, czyli wzdłuż linii interfejsu.

$$\sigma_{\varphi 1,2_{\varphi=0}}=\frac{1}{\sqrt{2\pi}}\sqrt{K_I{}^2+K_{II}{}^2}\,r^{\lambda_r-1}cos\left[arctan\left[\frac{K_{II}}{K_I}\right]+\delta\,ln\left[\frac{r}{2a}\right]\right]cosh[\pi\delta],$$

$$\tau_{r\varphi 1,2_{\varphi=0}}=\frac{1}{\sqrt{2\pi}}\sqrt{K_I{}^2+K_{II}{}^2}\,r^{\lambda_r-1}sin\left[arctan\left[\frac{K_{II}}{K_I}\right]+\delta\,ln\left[\frac{r}{2a}\right]\right]cosh[\pi\delta], \tag{10}$$

W sytuacji, gdy wykładnik λ opisany jest liczbą rzeczywistą wzory (10) uproszczą się do następującej postaci:

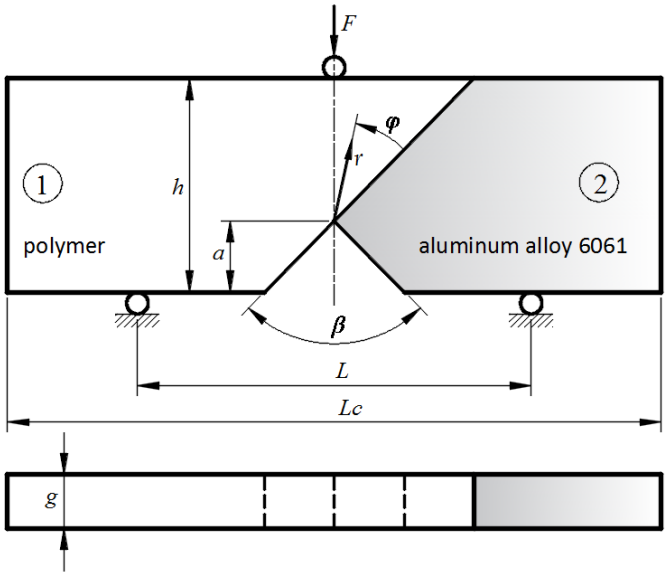
$$\sigma_{\varphi 1,2_{\varphi=0}}=\frac{1}{\sqrt{2\pi}}K_I\ r^{\lambda_r-1},\ \tau_{r\varphi 1,2_{\varphi=0}}=\frac{1}{\sqrt{2\pi}}K_{II}\ r^{\lambda_r-1}\ . \tag{11}$$

Do ilościowego opisu naprężeń niezbędne jest wyznaczenie wartości współczynników K_j. Wyznaczono je na podstawie porównania uzyskanych rozwiązań analitycznym z

naprężeniami otrzymanymi z rozwiązania MES. W związku z tym, że głównym celem prezentowanej pracy była eksperymentalna weryfikacja możliwości stosowania kryterium McClintocka, wykonano modele MES próbek, których wytrzymałość została zbadana doświadczalnie w pracy [11,12].

4. Próbki badawcze i model MES

W symulacjach numercznych modelowano próki, których geometria i właściwości materiałowe były identyczne jak próbek użytych w badaniach eksperymrntalnych opisanych w pracy [11, 12].



Rys. 4 Greometraia i sposób obciążenia próbek z karbem strukturalnym.

Analizowano dwa rodzaje próbek:
- element 1 wykonany z PC (Poliwęglanu) oraz komponent 2 ze stopu aluminium 6061;
- element 1 wykonany z PMMA (Poli (metakrylan metylu)), natomiast komponent 2 ze stopu aluminium 6061.

Komponenty zostały sklejone za pomocą kleju Weld-on® 10. Klej został tak dobrany, aby jego właściwości sztywnościowe były zbliżone do właściwości polimerów. Dzięki temu możliwe było spełnienie założenia, że modelowany jest bi-materiał, a nie kompozyt trójwarstwowy. W takiej sytuacji interfejs klejowy można było potraktować, jako warstwę bez żadnej grubości, ale o różnej wytrzymałości i odporności na kruche pękanie niż polimery lub stop aluminium rozpatrywane oddzielnie.

Aby uzyskać minimalną grubość kleju, próbki zostały połączone pod wysokim ciśnieniem za pomocą specjalnego uchwytu. Pozostawiono je do utwardzenia na okres 24 godzin, aby uzyskać wymaganą wytrzymałość. Grubość warstwy klejącej nie została zmierzona. Takie informacje nie były potrzebne do prognozowania odporności na kruche pękanie, przy założeniu, że analizowana struktura jest bi-materiałem. Warto zauważyć, że stosując wybrany klej w przypadku, gdy łączone byłyby dwa różne stopy metali np. stal i stop aluminium, pominięcie grubości warstwy kleju i traktowanie takiej struktury, jako bi-materiał byłoby niedopuszczalne. W takim przypadku do przewidywania obciążenia krytycznego można wykorzystać podejście oparte o kryterium Traction-Separation [25]. Mianowicie warstwa kleju powinna być modelowana przy użyciu specjalnych elementów skończonych (cohesive elements).

Wykonane analizy miały na celu określenie możliwości zastosowania zaproponowanego kryterium pękania dla struktur bi-materiałowych w sytuacji, gdy w

płaszczyźnie pękania (na interfejsie) występuję złożony stan naprężeń, a karb strukturalny generuje osobliwe pola naprężeń. Przy czym pola naprężeń, w zależności od cech geometryczno-materiałowych struktury, opisane mogą być za pomocą rzeczywistych lub zespolonych wykładników potęgowych λ. W badaniach wykorzystano próbki o różnym kącie wierzchołkowym karbu. Kąty wierzchołkowe β dobrano w ten sposób, żeby uzyskać przypadki, gdy naprężenia opisane są zarówno rzeczywistym jak i zespolonym wykładnikiem potęgowym λ. Co więcej zmienność kątów wierzchołkowych pozwoliła uzyskać różne proporcję naprężeń stycznych i normalnych występujących w płaszczyźnie połączenia. We wszystkich próbkach przyjęto jednakową wysokość karbu (mierzoną od dolnej powierzchni próbki), wymiary gabarytowe oraz położenie punktów podparcia i obciążania w próbie trójpunktowego zginania. Miało to na celu zapewnienie jednakowych warunków brzegowych (mocowania i obciążania) dla wszystkich badanych próbek. Wymiary gabarytowe dobrano arbitralnie uwzględniając możliwości stanowiska badawczego oraz przyrządu wykorzystywanego do klejenia.

Wymiary próbek wynosiły odpowiednio: długość całkowita Lc=254 mm, odległość między podporami L=90 mm, wysokość karbu a=19.1 mm, wysokość próbki h=50.8 mm, grubość g=5.4 mm. Analizowano trzy rodzaje próbek, o różnym kącie wierzchłkowym karbu strukturalnego: $\beta = 30^0$, $\beta = 90^0$ i $\beta = 120^0$. Próbki użyte w badaniach przedstawiono w tabeli 1. Dane materiałowe poszczególnych komponentów podano w tabeli 2.

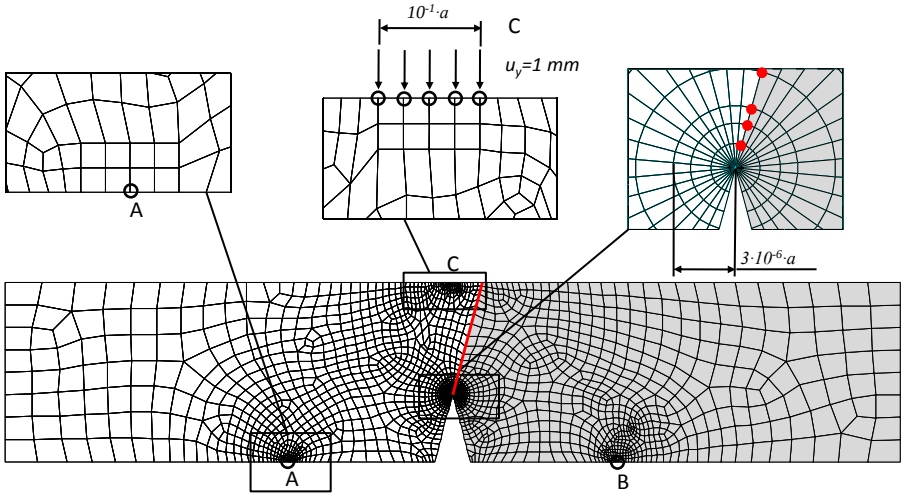
Tab. 1 Próbki badawcze

Lp.	próbki badawcze	kąt wierzchołkowy karbu $\beta[°]$
1	PC/ stop aluminium 6061	30
2	PC/ stop aluminium 6061	90
3	PC/ stop aluminium 60611	120
4	PMMA/ stop aluminium 6061	30
5	PMMA stop aluminium 6061	90
6	PMMA/ stop aluminium 6061	120

Tab. 2 Właściwości mechaniczne poszczególnych komponentów próbek [12]

	moduł Younga E [GPa]	Współczynnik Poissona ν
stop aluminium 6061	70	0.35
PC	2.38	0.37
PMMA	3.79	0.37

Badane próbki (rys.4) zamodelowano za pomocą MES, wykorzystując program ANSYS. Na rysunku 5 pokazano, dla wybranej próbki, podział na elementy skończone oraz warunki brzegowe.



Rys. 5. Podział na elementy skończone oraz warunki mocowania i obciążenia próbki z karbem o kącie wierzchołkowym β= 30^0, kolorem czerwonym zaznaczono węzły, z których naprężenia zostały wykorzystane do obliczenia współczynników intensywności naprężeń

Płaskie próbki opisano siatką czworokątnych, ośmiowęzłowych elementów skończonych, o zwiększonym zagęszczeniu w okolicy wierzchołkowej, z trójkątnymi elementami specjalnymi [35] otaczającymi punkt osobliwy (rys. 5). Całkowita długość bocznych krawędzi ostatnich trzech elementów zależała od wysokości karbu a i założono, że wynosi 3 10^{-6} a dla wszystkich próbek. Ze względu na dużą gęstość siatki elementów skończonych w obszarze wierzchołkowym karbu, przygotowane modele zawierały około 10000 elementów skończonych. Jak już wspomniano wcześniej, próbki można traktować jako strukturę bi-materiałową. Dlatego warstwa adhezyjna nie została uwzględniona w przygotowanych modelach numerycznych. Jeśli chodzi o warunki połączenia poszczególnych komponentów, to węzły leżące na interfejsie były wspólne dla obydwu materiałów (nie ma możliwości poślizgu między komponentami na interfejsie).

Ze względu na fakt, że ciężko jest określić rzeczywiste warunki tarciowo-kontaktowe, między podporami a materiałem próbki, jakie występują podczas próby trójpunktowego zginania należało przyjąć pewne uproszczenia. Zatem przetestowano dwa sposoby zamocowania:

I) podpory przesuwne (odebrana możliwość przemieszczenia pionowego w węzłach usytuowanych w punktach podparcia A i B);

II) podpory nieprzesuwne (odebrana możliwość przemieszczenia pionowego i poziomego w węzłach usytuowanych w punktach podparcia A i B).

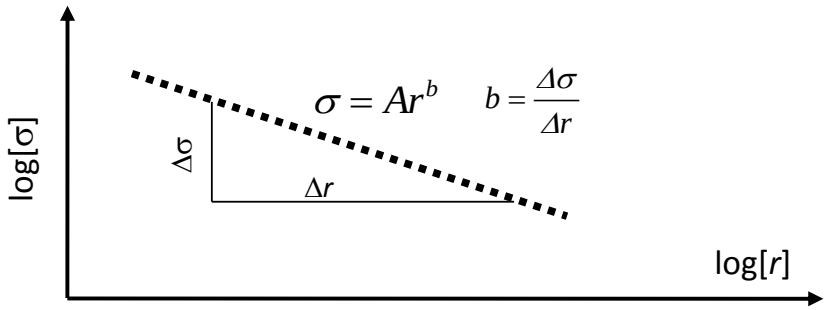
Co się tyczy warunków obciążenia, to próbki obciążano stałym przemieszczeniem pionowym u_y = 1mm, aplikowanym w wybranych węzłach (punkt C). Siła obciążająca F została określona na podstawie naprężeń w węzłach, w których aplikowano przemieszczenie u_y.
Obliczenia numeryczne wykonano dla dwóch przypadków stanu naprężeń: płaskiego stanu naprężenia i płaskiego stanu odkształcenia.

Jak już wspomniano wcześniej, w modelach *MES* klej, jako oddzielna warstwa materiału, nie został uwzględniony. Mimo to możliwe było uwzględnienie właściwości wytrzymałościowych interfejsu w zastosowanym kryterium pękania. Na podstawie symulacji numerycznych wyznaczono współczynniki intensywności naprężeń. Wykorzystano je do obliczenia wartości przewidywanej funkcji zniszczenia (opisanej dalszej części pracy), którą następnie porównano z wartościami krytycznymi. Właściwości wytrzymałościowe interfejsu w badanych próbkach zostały uwzględnione w ten sposób, że wartości krytyczne funkcji zniszczenia i parametru ρ_c zostały określone w oparciu o wytrzymałość na rozciąganie i

odporność na pękanie warstwy adhezyjnej (Tab. 3). Oba parametry zostały określone eksperymentalnie dla bi-materiałów wykonanych z PMMA i stopu aluminium, a także PC i stopu aluminium.

5. Opis metody wykorzystanej do wyznaczania uogólnionych współczynników intensywności naprężeń

Do wyznaczenia wartości współczynników K_j wykorzystano metodę ekstrapolacji. Metoda ta, w przeciwieństwie np. do metod energetycznych [37] czy też metod opartych na stosowaniu specjalnych elementów skończonych [4], jest mniej skomplikowana. Wadą tej metody jest konieczność stosowania dużego zagęszczenia siatki podziału na elementy skończone w okolicy wierzchołkowej koncentratora naprężeń. Dodatkowo na dokładność wyników ma wpływ wybór obszaru, w którym porównywane jest rozwiązanie numeryczne z analitycznym. Niedogodność tą można wyeliminować poprzez użycie w opisie analitycznym członów wyższych rzędów [17, 33, 24] lub ustalenie odpowiedniego kryterium wyboru węzłów, dla których wartości naprężeń otrzymane z MES porównuje się ze znanym rozwiązaniem analitycznym. Kryterium takie, określono w pracach [20] (dla zagadnienia szczeliny międzyfazowej) oraz [18] (dla karbu strukturalnego). Jak powszechnie wiadomo, jeżeli wykres naprężeń typu - $\sigma = Ar^b$ -w układzie podwójnie logarytmicznym jest liniowy (rys. 6), to gradient naprężeń wynosi *b*.



Rys. 6. Graficzna interpretacja osobliwych pól naprężeń o teoretycznym gradiencie *b*

Zatem, wyznaczając współczynniki intensywności naprężeń, porównuje się rozwiązania numeryczne i analityczne tylko dla węzłów o gradiencie równym *b*.

Dla zagadnienia karbu strukturalnego, składowe naprężeń zawsze uzależnione są jednocześnie od współczynników K_I i K_{II}, w związku z tym, aby skorzystać z powyższego kryterium, konieczne jest określenie tzw. 'naprężeń kombinowanych' [18]:

$$\sigma_{1(r,0)} = Sech[\pi\delta]\left(\sigma\cos\left[\delta ln\left[\frac{r}{2a}\right]\right] + \tau\sin\left[\delta ln\left[\frac{r}{2a}\right]\right]\right) = \frac{K_I}{\sqrt{2\pi}r^{1-\lambda_r}}, \tag{12}$$

$$\sigma_{2(r,0)} = \left(\tau\cos\left[\delta\log\left[\frac{r}{2a}\right]\right] - \sigma\sin\left[\delta\log\left[\frac{r}{2a}\right]\right]\right)Sech[\pi\delta] = \frac{K_{II}}{\sqrt{2\pi}r^{1-\lambda_r}}, \tag{13}$$

gdzie- σ,τ to odpowiednio naprężenia obwodowe i styczne uzyskane z MES.

Zgodnie z przyjętym kryterium przy wyznaczaniu poszukiwanych współczynników pod uwagę brano tylko pary węzłów gradiencie naprężeń $b = (\lambda_r - 1) \pm 0.01$.

'Naprężenia kombinowane' $\sigma_{i(r,0)}$ w odległości r_n i r_{n+1} od wierzchołka karbu można zapisać w następujący sposób:

$$\sigma_{j(r_n,0)} = \frac{K_{j(r)}}{\sqrt{2\pi}r^{1-\lambda_r}}(1+cr_n), \sigma_{j(r_{n+1},0)} = \frac{K_{j(r)}}{\sqrt{2\pi}r^{1-\lambda_r}}(1+cr_{n+1}). \tag{14}$$

Wykorzystując zależności (12÷14) po prostych przekształceniach matematycznych otrzymuje się wzory (15) umożliwiające wyznaczenie współczynników $K_{j(r)}$ (w pewnej odległości od wierzchołka karbu):

$$K_{I(r)} = \frac{\sqrt{2\pi}\left(r_n r_{n+1}\right)^{1-\lambda_r}}{r_n - r_{n+1}} sech[\pi\delta] \left\{ \begin{array}{l} r_n^{\lambda_r}\left(\sigma_{(r_{n+1})} cos\left[\delta ln\left[\frac{r_{n+1}}{2a}\right]\right] + \tau_{(r_{n+1})} sin\left[\delta ln\left[\frac{r_{n+1}}{2a}\right]\right]\right) \\ -r_{n+1}^{\lambda_r}\left(\sigma_{(r_n)} cos\left[\delta ln\left[\frac{r_n}{2a}\right]\right] + \tau_{(r_n)} sin\left[\delta ln\left[\frac{r_n}{2a}\right]\right]\right) \end{array} \right\}$$

$$K_{II(r)} = \frac{\sqrt{2\pi}\left(r_n r_{n+1}\right)^{1-\lambda_r}}{r_n - r_{n+1}} sech[\pi\delta] \left\{ \begin{array}{l} r_n^{\lambda_r}\left(\tau_{(r_{n+1})} cos\left[\delta ln\left[\frac{r_{n+1}}{2a}\right]\right] - \sigma_{(r_{n+1})} sin\left[\delta ln\left[\frac{r_{n+1}}{2a}\right]\right]\right) \\ -r_{n+1}^{\lambda_r}\left(\tau_{(r_n)} cos\left[\delta ln\left[\frac{r_n}{2a}\right]\right] - \sigma_{(r_n)} sin\left[\delta ln\left[\frac{r_n}{2a}\right]\right]\right) \end{array} \right\} \quad (15)$$

Obliczone współczynnik, dla wyselekcjonowanych węzłów, za pomocą wzoru (15) aproksymuje się linią prostą i w ten sposób wyznacza się uogólnione współczynniki intensywności naprężeń $K_{j(r=0)}$.

Warto zauważyć, że przypadku, gdy wykładnik potęgowy λ jest liczbą rzeczywistą ($\delta = 0$) zależność (15) upraszcza się do postaci przedstawionej w pracy [14].

$$K_I = \frac{\sqrt{2\pi}\left(r_n r_{n+1}\right)^{1-\lambda_r}\left(r_{n+1}^{\lambda_r}\sigma_{(r_n)} - r_n^{\lambda_r}\sigma_{(r_{n+1})}\right)}{r_{n+1} - r_n}. \quad (16)$$

6. Wyniki badań oraz ich dyskusja

Jak już wspomniano do zweryfikowania zmodyfikowanego kryterium McClintocka niezbędna jest znajomość jakościowego i ilościowego opisu pól naprężeń występujących w płaszczyźnie pękania, parametrów krytycznych i danych eksperymentalnych (obciążenie niszczące). Obciążenia niszczące zaczerpnięto z prac [11, 12]. Ponieważ po zainicjowaniu procesu pękania pęknięcie propagowało wzdłuż interfejsu, to w testowanym kryterium użyto parametrów krytycznych charakteryzujących właściwości warstwy adhezyjnej, dla których określono długość strefy zniszczenia ρ_c, zgodnie ze wzorem (5). Metody wyznaczania krytycznych parametrów warstwy adhezyjnej omówiono w pracach [7, 10, 22, 25].

Tab. 3. Właściwości wytrzymałościowe warstwy adhezyjnej [10]

	wytrzymałość na rozciąganie σ_c [MPa]	odporność na pękanie K_{Ic} [MPa m $^{0.5}$]	dystans krytyczny ρ_c [mm]
PC/ stop aluminium 6061	11.35	0.24	0.071
PMMA/ stop aluminium 6061	12.85	0.28	0.075

W celu określenia ilościowego opisu pól mechanicznych, obliczono uogólnione współczynniki intensywności naprężeń K_j. Wyznaczone je numerycznie metodą ekstrapolacji, wykorzystując dane uzyskane z MES i rozwiązania analityczne. Metodę ekstrapolacji, modelowanie MES i rozwiązanie analityczne przedstawiono i omówiono wcześniejszych częściach pracy. Wartości obliczonych współczynników przedstawiono w tabelach 4-5.

Tab. 4 Wartości uogólnionych współczynników intensywności naprężeń K_j oraz obciążania F, przy którym obliczano współczynnik K_j
PC /stop aluminium 6061

β [°]	Warunki podparcia					
	podpory przesuwne			podpory nieprzesuwne		
	K_I [Pa m$^{1-\lambda r}$]	K_{II} [Pa m$^{1-\lambda r}$]	F [N]	K_I [Pa m$^{1-\lambda r}$]	K_{II} [Pa m$^{1-\lambda r}$]	F [N]
30	6621327.4* 5654415.8**	16207.8* 565904.7**	4139.0* 3508.3**	2670172.3* 2337935.8**	334892.9* 576351.2**	5598.9* 4718.3**
90	6953117.8* 7929593.6**	-2675922.6* -1851771.4**	3532.6* 3004.2**	2654647.7* 3089615.1**	-1281221.0* -622169.7**	5071.6* 4276.7**
120	3224964.6* 3206237.2**	-4631416.6* -6451442.2**	3345.1* 2853.9**	1156739.9* 1185145.1**	-1821056.5* -2538667.0**	5001.9* 4225.0**
*-płaski stan odkształcenia, **- płaski stan naprężenia						

Tab. 5 Wartości uogólnionych współczynników intensywności naprężeń K_j oraz obciążania F, przy którym obliczano współczynnik K_j
PMMA/stop aluminium 6061

β [°]	Warunki podparcia					
	podpory przesuwne			podpory nieprzesuwne		
	K_I [Pa m$^{1-\lambda r}$]	K_{II} [Pa m$^{1-\lambda r}$]	F [N]	K_I [Pa m$^{1-\lambda r}$]	K_{II} [Pa m$^{1-\lambda r}$]	F [N]
30	10590286.9* 9092521.1**	88797.8* 449519.4**	6510.0* 5522.1**	4304874.73* 3774299.6**	466048.7* 845462.9**	8818.2* 7435.5**
90	15655266.1* 13838295.9**	-8259321.6* -4731359.1**	5572.9* 4742.3**	6032440.86* 5427305.2**	-1915526.6* -578767.5**	8002.9* 6753.4**
120	4796910.9* 4040648.9**	-6494991.4* -6507052.4**	5279.8* 4507.1**	1798019.9* 1525983.9**	-2641592.9* -2594161.8**	7892.7* 6671.4**
*-płaski stan odkształcenia, **- płaski stan naprężenia						

Wartości wykładników potęgowych λ, otrzymane z równania (7), dla stałych materiałowych podanych w tabeli 2, zestawiono w tabeli 6.

Tab. 6 Wartości wykładników potęgowych λ

β [°]	Typ próbki			
	PC /stop aluminium 6061		PMMA/stop aluminium 6061	
	λ_r	δ	λ_r	δ
30	0.5032* 0.5033**	0.0611* 0.0958**	0.5051* 0.5052**	0.0579* 0.0913**
90	0.5222* 0.5231**	0.0450* 0.0810**	0.5339* 0.5352**	0.0235* 0.0646**
120	0.5058* 0.5324**	0* 0**	0.5003* 0.5071**	0* 0**
*-płaski stan odkształcenia, **- płaski stan naprężenia				

Jak już wspomniano, jeżeli w płaszczyźnie pękania występują zarówno naprężenia styczne, jak i normalne, w celu prognozowania procesu pękania można zastosować ekwiwalentny współczynnik intensywności naprężeń K_E. Dla analizowanego przypadku został on określony (na podstawie analitycznego opisu lokalnych pól naprężeń) za pomocą wzoru (17):

$$K_E = \cosh(\pi\delta)\sqrt{K_I^{\,2} + K_{II}^{\,2}}, \tag{17}$$

Krytyczną wartość współczynnika K_{Ec} można wyznaczyć poprzez rozwiązanie poniższego układu równań (18):

$$\frac{\sigma_\varphi(\rho_c,0)}{\sigma_c}=1,\left(\frac{\partial\sigma_\varphi}{\partial\varphi}\right)_{\varphi=0,r=\rho_c}=0. \tag{18}$$

Po rozwiązaniu układu równań (18) – z wykorzystaniem wzorów (5), (9) i (10) uzyskuje się zależność, pozwalającą wyznaczyć wartość krytycznego współczynnika K_{Ec} (19):

$$K_{Ec}=(2\pi)^{\lambda_r-\frac{1}{2}}\left(\frac{K_{Ic}{}^2}{\sigma_c^2}\right)^{1-\lambda_r}\sigma_c. \tag{19}$$

Warto zauważyć, że dla przypadku rozciąganego elementu ze szczeliną ($\lambda_r=0.5$) lub z karbem, o kącie wierzchołkowym równym π ($\lambda_r=1$), zależność (19) upraszcza się odpowiednio do następujących postaci: $K_{Ec}=K_{Ic};K_{Ec}=\sqrt{2\pi}\sigma_c$, co jest zgodne z danymi literaturowymi.

Zakładając, że proces pękania zostanie zainicjowany, w sytuacji, gdy:

$$K_E=K_{Ec}, \tag{20}$$

przewidywana siła krytyczna może być obliczona na podstawie następującego warunku:

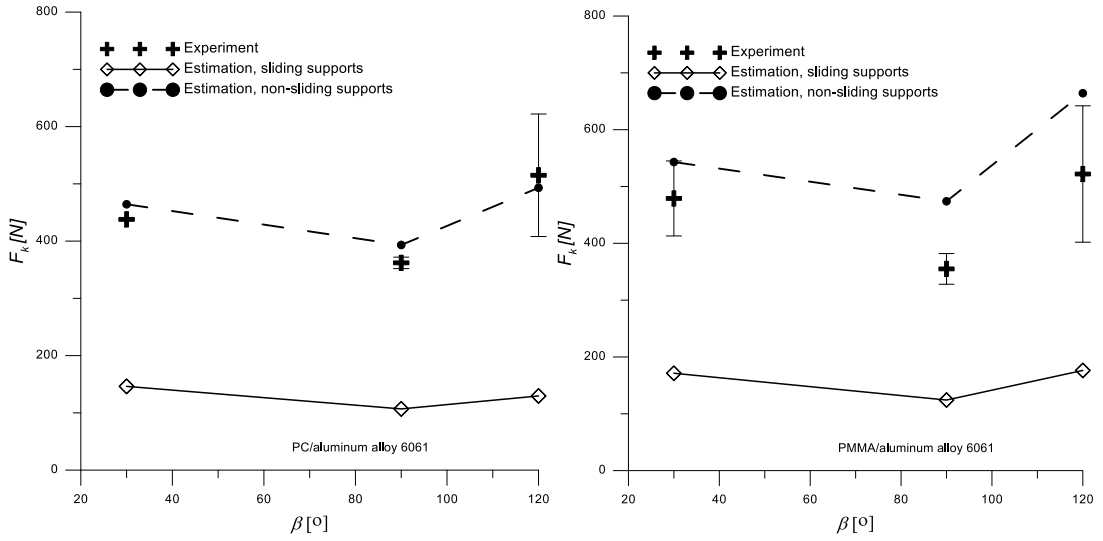
$$F_k=\frac{K_{Ec}F}{K_E}, \tag{21}$$

gdzie F jest obciążeniem, przy którym obliczano ekwiwalentny współczynnik intensywności naprężeń K_E (17).

Jak już wspomniano wcześniej, trudno jest odzwierciedlić rzeczywiste warunki tarciowo- kontaktowe [2, 15] występujące w miejscu podparcia przy numerycznym modelowaniu trójpunktowej próby zginania. Dlatego też na rysunku 7 (dla płaskiego stanu naprężenia) i tabela 7 (dla płaskiego stanu odkształcenia) podano wartości prognozowanych sił F_k (21) określonych dla dwóch wariantów zamocowania próbek w modelach MES. Oczywistym jest fakt, że rzeczywista siła krytyczna przyjmie wartości z zakresu ograniczonego przez siły oszacowane przy stosowaniu podparcia przesuwnego i nieprzesuwnego.

Tab. 7 Wartości sił krytycznych F_k (21) obliczone dla różnych typów próbek, płaski stan odkształcenia

β [°]	PC/ stop aluminium 6061		PMMA/ stop aluminium 6061	
	experiment [12]	estimation (21)	experiment [12]	estimation (21)
30	438±1	151.9* 505.7**	479±66	177.7* 588.6**
90	362±10	139.2* 403.3**	355±27	121.3* 487.2**
120	515±107	150.3* 587.9**	522±120	183.6* 693.6**
*- podpory przesuwne, **- podpory nieprzesuwne				

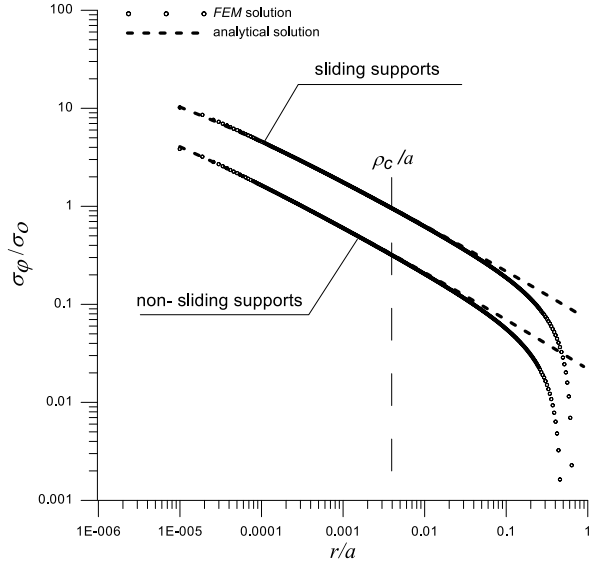


Rys. 7 Porównanie krytycznych wartości obciążeń, uzyskanych z testowanych kryteriów z wartościami otrzymanymi z badań eksperymentalnych [12], a) PC/ stop aluminium 6061, b) PMMA/ stop aluminium 6061, płaski stan naprężenia

Wartości sił krytycznych określone za pomocą wzoru (21) zostały porównane z danymi eksperymentalnymi, co pokazano na rys. 7 i w tabeli 6. Eksperymentalne obciążenie inicjujące proces pękania jest średnią wartością siły niszczącej określonej na podstawie, co najmniej trzech prób wykonanych dla każdego typu próbki.

Na podstawie uzyskanych wyników można stwierdzić, że rozkład prognozowanych sił krytycznych jest zgodny z danymi eksperymentalnymi. Lepszą zbieżność wyników, zarówno dla płaskiego stanu naprężenia, jak i odkształcenia, uzyskano przy zastosowaniu w modelach MES podpór nieprzesuwnych.

Przy określaniu sił krytycznych w opisie analitycznym użyto tylko pierwszego członu osobliwego rozwiązania asymptotycznego. Używanie członów wyższych rzędów nie było konieczne ze względu na fakt, że rozwiązania analityczne (przy użyciu tylko pierwszego członu osobliwego) i numeryczne zgadzały się ze sobą w obszarze większym niż krytyczny dystans ρ_c, co pokazano na rysunku 8.



Rys. 8 Unormowane naprężenia obwodowe w funkcji *r/a* dla próbki PC/ stop aluminium 6061specimen z kątem wierzchołkowym β= 30^0, $\sigma_0 = 3FL/\left(2gh^2\right)$, płaski stan naprężenia

Warto zauważyć, że prognozowanie odporności na kruche pękanie za pomocą koncepcji ekwiwalentnego współczynnika intensywności naprężeń jest dość skomplikowane. Konieczne jest bowiem wyznaczenie wykładników λ i uogólnionych współczynników intensywności naprężeń K_j. W przypadku materiałów jednorodnych nie stanowi to poważnego problemu, ponieważ w literaturze dostępne są przybliżone formuły, które pozwalają na obliczenie uogólnionych współczynników intensywności naprężeń, a wykładnik λ zależy tylko od kąta wierzchołka karbu i można go łatwo określić (wartości wykładników można znaleźć np. w pracy [28]). W przypadku bi-materiału z karbem usytuowanym na interfejsie, zarówno K_j, jak i λ zależą od cech geometryczno- materiałowych struktury i powinny być wyznaczone indywidualnie dla każdego analizowanego przypadku. Ze względu na powyższe niedogodności, w artykule podjęto próbę opracowania procedury prognozowania odporności na kruche pękanie, która może być bardziej praktyczna z inżynierskiego punktu widzenia.

Autorzy wielu prac, np. [30] wskazują, że krytyczna wartość współczynnika uwalniania energii (dla zagadnienia bi-materiału)) zależy od proporcji naprężeń stycznych i normalnych występujących w płaszczyźnie pękania. Dlatego też wartości krytyczne σ_c również muszą być uzależnione od takiego czynnika. W związku z tym warunek (3) można zapisać w następujący sposób (22):

$$\sigma_\varphi(\rho_c, 0) = \sigma_c(\psi) \tag{22}$$

Biorąc pod uwagę wzajemne zależności między poszczególnymi parametrami mechaniki pękania zaproponowano następującą formę funkcji uzależniającej naprężenia krytyczne od proporcji naprężeń stycznych i normalnych ψ:

$$\sigma_c(\psi) = \sigma_c / \sqrt{1+\psi^2}, \tag{23}$$

gdzie: $\psi = \dfrac{\tau_{r\varphi}(\rho_c, 0)}{\sigma_\varphi(\rho_c, 0)}$.

Warto zauważyć, że zaproponowana modyfikacja kryterium pękania (23) jest zbieżna z koncepcją przedstawioną w pracy [34]. Autor pracy [34] sugeruje zastosowanie, w procesie prognozowania inicjacji pękania, dodatkowego współczynnika, uwzgledniającego wpływ ilościowego udział naprężeń stycznych i normalnych. Współczynnik ten nie jest stały i zależy od geometrii elementu i warunków obciążeniowych. Wartość siły krytycznej, wykorzystując kryterium (23), można obliczyć z zależności (24):

$$F_k = \frac{\sigma_c(\psi) F}{\sigma_\varphi(\rho_c, 0)}. \tag{24}$$

Siła krytyczna F_k (24) może być określona na dwa sposoby:
- przy wykorzystaniu zarówno rozwiązań analitycznych (opis jakościowy (9, 10)), jak i numerycznych (opis ilościowy (15));
- przy wykorzystaniu tylko rozwiązania numerycznego (określenie naprężeń $\tau_{r\varphi}(\rho_c, 0)$ i $\sigma_\varphi(\rho_c, 0)$ np. przy użyciu MES).

Oczywiście sposób drugi, z praktycznego punktu widzenia, jest mniej skomplikowany i zalecany do stosowania w obliczeniach inżynierskich.

Jednak w prezentowanej pracy zastosowano podejście pierwsze. Podyktowane to było koniecznością sprawdzenia zbieżności rozwiązań uzyskiwanych ze wzorów (21) i (24). Ze względu na fakt, że siły krytyczne (21) zostały oszacowane przy użyciu rozwiązań analitycznych i numerycznych, to identyczne podejście powinno zostać użyte przy prognozowaniu z wykorzystaniem zależności (24). W związku z tym wykorzystując rozwiązania analityczne i numeryczne wyznaczono siły krytyczne F_k (24) dla wszystkich analizowanych próbek. Uzyskane wyniki były identyczne z wynikami uzyskanymi na

podstawie wzoru (21) i w związku z tym nie było potrzeby zamieszczenia ich w prezentowanej pracy.

7. Podsumowanie i wnioski

W pracy analizowano proces inicjacji pęknięć w strukturze bi-materiałowej z karbem usytuowanym na interfejsie. Zweryfikowano możliwość zastosowania kryterium McClintocka do prognozowania odporności na pękania tego typu elementów konstrukcyjnych. Stosowanie tego kryterium wymaga znajomości ilościowego i jakościowego opisu pól naprężeń wokół wierzchołka koncentratora. Zatem uzyskano opis analityczny i przedstawiono metodykę jego otrzymywania. Ponadto omówiono metodę wyznaczania uogólnionych współczynników intensywności naprężeń, uwzględniającą jakościowy charakter osobliwości pól generowanych przez karb strukturalny, i dla wybranych przypadków wyznaczono ich wartości.

Uzyskane rozwiązania analityczne i numeryczne pozwoliły na sformułowanie postaci kryterium pękania i parametrów kryterialnych. Opracowano dwie postacie kryterium pękania, oparte na:

-ekwiwalentnym współczynniku intensywności naprężeń;

-wprowadzeniu modyfikacji w kryterium McClintocka polegającej na uzależnieniu naprężeń krytycznych od proporcji naprężeń stycznych i normalnych występujących w płaszczyźnie pękania.

Wykonane analizy pokazały, że z obydwu postaci kryterium pękania uzyskuje się identyczne wyniki prognozowania sił krytycznych. Jednak praktycznego punktu widzenia, druga postać kryterium pękania jest korzystniejsza. Wynika to z faktu, że w prognozowaniu pękania można wykorzystać tylko naprężenia wyznaczone np. za pomocą *FEM* bez konieczności wyznaczania współczynników intensywności naprężeń.

Wartości obciążeń krytycznych wynikających z hipotez porównano z wartościami uzyskanymi z badań eksperymentalnych. Ze względu fakt, że przy modelowaniu numerycznym próby trójpunktowego zginania nie można odzwierciedlić rzeczywistych warunków tarciowo-kontaktowych, jakie występują w miejscu podparcia próbek, wyznaczono jedynie przedział, w którym znajdują się prognozowane sił krytyczne. W zdecydowanej większości analizowanych przypadków siły krytyczne wyznaczone eksperymentalnie znajdowały się w przedziale określonym za pomocą testowanego kryterium. Co więcej, trend zmienności prognozowanych sił zgadzał się z danymi eksperymentalnymi. Pozwala to sądzić, że analizowane kryterium może być stosowane do prognozowania inicjacji procesu pękania elementów z karbem usytuowanym na interfejsie. Jednak w celu jednoznacznego stwierdzenia takiego faktu, należałoby wykonać dodatkowe badania eksperymentalne. Takie badania powinny być zaplanowane w ten sposób, żeby można było w modelowaniu numerycznym odzwierciedlić rzeczywiste warunki mocowania i obciążenia próbek badawczych. Wykonanie takich badań i ponowne zweryfikowanie przydatności hipotezy McClintocka będzie celem kolejnej pracy autora.

*Badania zostały zrealizowane w ramach pracy nr **S/WM/2/13** i sfinansowane ze środków na naukę MNiSW.*

References

1. Ayatollahi M.R., Torabi A.R. A criterion for brittle fracture in U-notched components under mixed mode loading. Engineering Fracture Mechanics 2009; 39: 1883–1896.

2. Baranowski P., Damaziak K., Małachowski J. Brake system studies using numerical methods, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2013; 15 (4): 337–342.
3. Bogy D.B., Wang K.C. Stress singularities at interface corners in bonded dissimilar isotropic elastic materials. International Journal of Solids and Structures 1971; 1: 993-1005.
4. Byskov E. Calculation of stress intensity factors using finite element method with cracked elements. International Journal of Fracture Mechanics 1970; 6(2): 59-167.
5. Carpinteri, A., Paggi M. Analytical study of the singularities arising at multi-material interfaces in 2D linear elastic problems. Engineering Fracture Mechanics 2007; 74: 59–74.
6. Griffith A.A. The phenomena of rupture and flow in solids. Philosophical Transactions series A 1920; 221: 163-198.
7. Kinloch A. J. Adhesion and adhesives, Science and Technology. London: Springer, 1987.
8. Kirsch G. Die theorie der elastizität und die bedürfnisse der festigkeitslehre. Verein deutscher Ingenieure Zeitschrift 1898; 29: 797-807.
9. Knesl Z., Klusak J., Nahlik L. Crack initiation criteria for singular stress concentrations, Part I: A Universal assessment of singular stress concentrations, Engineering Mechanics 2007; 14(6): 399–408.
10. Krishnan A., Xu LR. Systematic evaluation of bonding strengths and fracture toughnessess of adhesive joints. The Journal of Adhesion 2011; 87(1): 53–71.
11. Krishnan A, Xu LR. Experimental studies on the interaction among cracks, notches and interfaces of bonded polymers. International Journal of Solids and Structures 2013, 50: 1583–1596.
12. Krishnan A., Roy Xu L. An experimental study on the crack initiation from notches connected to interfaces of bonded bi-materials. Engineering Fracture Mechanics 2013; 111: 65–76.
13. Leguillon D. A criterion for crack nucleation at a notch in homogeneous materials. Comptes Rendus de l'Académie des Sciences - Series IIB – Mechanics 2001; 329(2): 97–102.
14. Li Y., Song M. Method to calculate stress intensity factor of V-notch in bi-materials. Acta Mechanica Solida Sinica 2008; 21(4): 337–346.
15. Łukaszewicz A. Nonlinear numerical model of heat generation in the rotary friction welding, Journal of Friction and Wear 2018; 39 (6): 612-619.
16. McClintock F. A. Ductile fracture instability in shear. Journal of Applied Mechanics 1958; 25: 582-588.
17. Mieczkowski G. Description of stress fields and displacements at the tip of a rigid, flat inclusion located at interface using modified stress intensity factors. Mechanika 2015; 21(2): 91-98.
18. Mieczkowski G. Stress fields and fracture prediction for adhesively bonded bi-material structure with sharp notch located on the interface. Mechanics of Composite Materials 2017; 53(3): 305-320.
19. Mieczkowski G. Stress fields at the tip of a sharp inclusion on the interface. Mechanics of Composite Materials 2016; 52(5):601-610.
20. Naik R. A., Crews J. H. Determination of stress intensity factors for interface cracks under mixed-mode loading. Paper presented at the ASTM National Symposium on Fracture Mechanics, June 30-July 2,1992, Gatlinburg, TN.
21. Parton V.Z., Perlin P.I. Mathematical methods of the theory of elasticity. Moscow: Mir Publishers, 1984.

22. Pirondi A., Nicoletto, G. Fatigue crack growth in bonded DCB specimens. Engineering Fracture Mechanics 2004; 71(4–6): 859–871.
23. Ritchie R. O., Knott J. F., Rice J. R. On the relation between critical tensile stress and fracture toughness in mild steel. Journal of the Mechanics and Physics of Solids 1973; 21: 395-410.
24. Rogowski G., Molski K.L. The T-stress effect on the plastic zone size in a thin ductile material layer sandwiched between two elastic adherents. Engineering Fracture Mechanics 2016; 168 (A): 260-270.
25. Rudawska A., Dębski H. Experimental and numerical analysis of adhesively bonded aluminium alloy sheets joints. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2011; 1: 4–10.
26. Savruk M.P, Shkarayev S., Madenci E. Stress near apex of dissimilar material with bilinear behavior. Journal of Applied Fracture Mechanics 1999; 31: 203-212.
27. Seweryn A., Łukaszewicz A. Verification of fracture criteria of elements with V-shaped notches. Eksploatacja i Niezawodnosc — Maintenance and Reliability 2001; 5: 6–8.
28. Seweryn A., Molski K. Elastic stress singularities and corresponding generalized stress intensity factors for angular corners under various boundary conditions. Engineering Fracture Mechanics 1996; 55: 529-556.
29. Sih G.C. Strain-energy-density factor applied to mixed mode crack problems. International Journal of Fracture 1974; 10: 305-321.
30. Sih G.C., Chen E.P. Cracks in composite materials, Ch.3 (Mechanics of Fracture VI) ed. G. C. Sih. Hague: Martinus Nijhoff Publishers, 1981.
31. Sneddon I.N. The distribution of stress in the neighbourhood of a crack in an elastic solid. Proceedings of the Royal Society of London A 1946;187(1009): 229-260.
32. Sun C T., Jih C. J. On strain energy release rates for interfacial cracks in bi-material media. Engineering Fracture Mechanics 1987; 28: 13-20.
33. Sun C. T., Qian H. Brittle fracture beyond the stress intensity factor. Journal of Mechanics of Materials and Structures 2009; 4(4): 743-753.
34. Taylor, D. The Theory of Critical Distances: A new perspective in fracture mechanics. Oxford: Elsevier, 2007.
35. Tracey D. M. Finite elements for determination of crack tip elastic stress intensity factors. Engineering Fracture Mechanics 1971; 3(3): 255-265.
36. Tran V.-X., Leguillon D., Krishnan A. Interface crack initiation at V-notches along adhesive bonding in weakly bonded polymers subjected to mixed-mode loading. International Journal of Fracture Mechanics 2012; 176: 65–79.
37. Treifi M., Oyadiji S.O. Strain energy approach to compute stress intensity factors for isotropic homogeneous and bi-material V-notches. International Journal of Solids and Structures 2013; 50: 2196–2212.
38. Williams M. L. Stress singularities resulting from various boundary conditions in angular corners of plate in extension. Journal of Applied Mechanics 1952; 9: 526-528.

dr hab. Jacek Wawrzosek
dr Szymon Ignaciuk
dr Andrzej Bochniak

Katedra Zastosowań Matematyki i Informatyki
Uniwersytet Przyrodniczy w Lublinie
Ul. Głęboka 28, 20-612 Lublin, Poland
E-mails: jacek.wawrzosek@up.lublin.pl, szymon.ignaciuk@up.lublin.pl (corresponding author), andrzej.bochniak@up.lublin.pl

Wybrane problemy niejednoznaczności ceny dualnej wody w postoptymalizacyjnej analizie systemu wodociągów

Słowa kluczowe: *sieć dystrybucji wody, cena dualna wody, proces eksploatacji, model gospodarowania wodą, postoptymalizacja*

Streszczenie: W literaturze uważa się, że cena dualna wody jest obiektywną przesłanką do kształtowania rynkowej ceny wody. Jednak autorzy zauważają, że pojedynczy wektor cen dualnych w dystrybucji wody, gdy jest niejednoznaczny, nie powinien stać się podstawą do podejmowania decyzji zarówno normującej cenę wody jak i wpływającej na procedury modernizujące sieć wodociągową. Praca uczula inżynierów gospodarki wodnej by nie powielali powszechnych błędów oprogramowania oraz wskazuje jak, pomimo kompletnego braku literaturowych wskazówek, praktycznie rozwiązywać napotykane problemy techniczne. Liniowa zależność wektorów wierszowych parametrów lewych stron wiążących warunków ograniczających w modelu programowania liniowego dla zużycia wody identyfikowana jest tu jako przyczyna niejednoznaczności wektorów cen dualnych. Ta niejednoznaczność w zagadnieniach dystrybucji wody wymaga kształtowania alternatywnych scenariuszy technicznych pozwalających na wariantowy wybór sposobu modyfikacji systemu poboru wody. Dlatego opisano zasady wyznaczania proporcjonalności jednoczesnych zmian niektórych parametrów prawych stron warunków ograniczających. Na uproszczonym modelu dystrybucji wody sformułowano i wskazano te zasady optymalnego doboru najbardziej produktywnych wektorów dla metody parametrycznego programowania liniowego. Opracowana w pracy metodyka umożliwia m.in. wygenerowanie alternatywnych scenariuszy technicznych oszczędzania różnej ilości wody, skutkującej różnymi oszczędnościami finansowymi.

1. Wstęp

1.1. Kształtowanie ceny wody celem zrównoważenia zapotrzebowania na wodę

Analiza przeprowadzona przez Bank Światowy [64] wskazuje, że niedobór wody w niektórych regionach może obniżyć PKB nawet o 6% i prowadzić do zwiększonej migracji, a w niektórych przypadkach do większego ryzyka konfliktu. Susze i okresy niedoboru wody stały się zjawiskiem powszechniejszym i częstszym w Europie [17]. Doświadczenia suszy w Europie w latach 2011, 2012, 2015 i 2018 były najgorsze od stulecia i dotyczyły nie tylko Europy Południowej i Zachodniej, ale także krajów w Europie Północnej (w tym Wielkiej Brytanii, Francji, Niemiec, Szwecji i Polski). Różnica między zaopatrzeniem w wodę a jej rosnącym popytem określa również kluczowe ograniczenia rozwoju gospodarczego Chin. Szacuje się, że przed 2005 r. z powodu niedoboru wody w produkcji, Chiny traciły rocznie 28 511 milionów $. Brown [7] stwierdził, że niedługo niedobór wody w Chinach stanowić będzie zagrożenie dla globalnych potrzeb zbożowych. Dlatego woda wraz z ziarnem i ropą naftową są określane jako zasoby strategiczne. Stąd sugeruje się, że władze publiczne powinny kształtować

cenę wody dla użytkowników w celu odzwierciedlenia jej prawdziwego niedoboru lub kosztów alternatywnych [15].

1.2. Metody optymalizacji i postoptymalizacji w planowej eksploatacji zasobów wodnych

Metody optymalizacji i postoptymalizacji mogą być wykorzystane jako narzędzie pozwalające na dynamiczną korektę i doskonalenie eksploatacji złożonych systemów. Metody te wymagają dwóch podejść: matematycznego i menadżerskiego [29-30]. Narzędzia i zasady optymalizacji umożliwiły opracowanie modeli normatywnych dla optymalnego zarządzania systemami zasobów wodnych na dużą skalę, uwzględniających wszechobecną niepewność w prognozowaniu procesów naturalnych i skutków gospodarczych [13]. Liczne modele optymalizacji bywają wykorzystane do zapewnienia wysokich parametrów niezawodności sieci, parametrów niezawodności jakości wody, odpowiednich schematów operacyjnych uwzględniających liczne ograniczenia hydrauliczne w postaci: wysokości spiętrzonej wody (*hydraulic head*), wycieków, zmiany zużycia energii pompy i sekwencyjnej dyskretnej pracy pompy przy tym minimalizującej koszty operacyjne. Większość modeli optymalizacji używa technik matematycznych, takich jak: programowanie liniowe (*LP*), programowanie dynamiczne (*DP*) i programowanie nieliniowe (*NLP*) lub ich wariacje [2, 10, 13-14, 19, 22, 25, 53]. Niektóre modele problemów niezawodności lub optymalizacji kosztu i czasu konserwacji sieci wodociągowej lub systemu kanalizacyjnego konstruowane są na gruncie statystyczno-stochastycznym [3, 38]. Romaniuk [44-46] prezentuje liczne eksperymenty numeryczne skupiające się np. na optymalizacji wartości oczekiwanej kosztów i czasu eksploatacji sieci wodociągowej, gdy parametrem decyzyjnym jest bezwarunkowy czas wymiany tzn. gdy lepiej jest wymienić fragment sieci zamiast dokonywać jego kolejnej, przyszłej naprawy.

Modele niezawodności, gotowości i bezpieczeństwa w połączeniu z programowaniem liniowym są użyteczne w identyfikacji oraz predykcji niezawodności, gotowości i bezpieczeństwa złożonych systemów technicznych, a także w optymalizacji i analizie kosztów eksploatacji tych systemów [28, 32-33, 66]. Wskazana tu niepewność obranych parametrów typowego modelu programowania liniowego wymaga, by oprócz procedury optymalizacyjnej, mieć łatwy dostęp do procedur postoptymalizacyjnych [56, 58].

Freire-González i in. [18] dokonują przeglądu literatury na temat istniejących badań dotyczących modeli przepływów międzygałęziowych (metody input-output) do oceny ekonomicznych skutków niedoboru wody podczas suszy oraz metod programowania liniowego wejścia-wyjścia (*IO-LP*) w podejściu do kwestii planowania zasobów wodnych w kontekście suszy i niedoboru wody. Gibbons [22] oraz Liu i in. [36] wskazują, że prawidłowe ustalanie cen dla zasobów wodnych odzwierciedlających ich rzeczywistą wartość jest bardzo ważne dla oszczędzania wody i dla złagodzenia niedoborów zasobów wodnych. Dalej uznają cenę dualną wody jako tą, która winna służyć do ilościowego określenia rzeczywistej wartości zasobów wodnych, odzwierciedlającego również ich niedobór. Czynią tak pomimo, że niekiedy twierdzi się, że praktycznie niemożliwe jest uzyskanie ceny dualnej wody poprzez rozwiązanie liniowego modelu programowania. Jednak Liu i in. [36] łącząc metodę analizy wejścia-wyjścia z metodą *LP* opracowali model z ograniczeniami dotyczącymi popytu końcowego, całkowitej produkcji, salda handlu i dostępności wody. Model ten posłużył do oszacowania ceny dualnej wody. Wyniki te stanowią cenne odniesienie do ustalenia racjonalnych cen wody przemysłowej i produktywnej w rejonach dziewięciu głównych dorzeczy chińskich. Przeglądu analogicznych badań modelowania wartości wody w różnych sektorach gospodarki w Afryce Południowej dokonują Nieuwoudt i Backeberg [42], a na świecie Conradie i Hoag [11]. Badania te przeprowadzono

w wyniku powstania zapotrzebowania na modele do pomiaru gotowości do zapłaty za wodę np. do nawadniania upraw rolnych.

1.3. Problemy programowania liniowego w modelowaniu zasobów wodnych

Typowy system zasobów wodnych składa się ze zbiorników wodnych, hydroelektrowni, nawadnianych gruntów, kanałów sztucznych i nawigacyjnych itp. będących w zasięgu rzeki lub dorzecza. Stąd optymalne planowanie wielozadaniowego systemu zasobów wodnych, czyli zaprojektowanie "najlepszego" systemu, który ma zostać zbudowany i wykorzystany w horyzoncie planowania, jest podporządkowane m. in. ograniczeniom technicznym, ekonomicznym, finansowym, społecznym i politycznym. Ograniczenia te obejmują sezonowe wahania zaopatrzenia w wodę, warunki geograficzne i geologiczne wybranych miejsc, istnienie kapitału, pożyczki, siłę roboczą i usługi lokalne, stopę procentową (i jej tendencje), plany rozwoju regionalnego itp. [25]. Cytowani autorzy wykorzystali model *LP* dla bardzo złożonego systemu zasobów wodnych uwzględniając szereg licznych ograniczeń związanych ze zbiornikiem, z nawadnianiem, z hydroelektrownią, ze sztucznymi ściekami i ograniczenia nawigacyjne. Przy tym badano przypadki rzek w południowej Argentynie, a te typowe problemy opisywało około 300 ograniczeń i 300 zmiennych. W takiej masie ograniczeń nie trudno natrafić na liniowo zależne wektory wierszowe parametrów lewych stron warunków ograniczających, a to już rodzi problemy przy analizowanej tu postoptymalizacji.

McKee i in. [41] opracowali model procesu eksploatacji warstwy wodonośnej przez ponad 900 odwiertów, poczynionych głównie na potrzeby przemysłu, zaopatrzenia komunalnego i nawadniania upraw w Arkansas. Uwzględnili 3 warianty modeli *LP* w celu symulacji zoptymalizowanych wypływów wody powierzchniowej i wody gruntowej przy jednoczesnym zachowaniu natężenia przepływu strumienia oraz licznych ograniczeń hydraulicznych. Również to złożone zagadnienie posiada wskazane powyżej problemy z liniową zależnością.

Techniki optymalizacji wielokryterialnej, np. dla skażonego obiektu wodonośnego, sprowadza się do optymalizacji jednokryterialnej poprzez zastosowanie metody ważonej sumy lub metody ograniczeń [16, 27, 40]. W ostatniej z tych metod optymalizuje się jedną z funkcji celu, wykorzystując inne funkcje celu jako ograniczenia, tj. włączając je do warunków ograniczających modelu, tym samym zwiększając liczbę ograniczeń. Prowadzi to zazwyczaj do analizowanego w obecnej pracy problemu liniowej zależności towarzyszącej nadmiernej liczbie warunków ograniczających.

Abdy Sayyed i in. [1] przeprowadzają optymalizację sieci dystrybucji wody poprzez zminimalizowanie kosztów sieci przy warunkach ograniczających wynikających z wymagań ciśnieniowych we wszystkich węzłach. Z uwagi na komplikującą zagadnienie mnogość warunków ograniczających, już w optymalnym projekcie niektóre warunki ograniczające są tam zastępowane poprzez dodatkową karę w funkcji celu. Kara jest stosowana za niespełnienie ograniczeń ciśnieniowych. W cytowanej pracy zostały zastosowane trzy metody wnioskowania karnego. Oznacza to zaznaczającą się niekiedy w literaturze tendencję do ucieczki od nadmiaru warunków ograniczających. Lecz wymaga to właściwego doboru funkcji kary.

Frizzone i in. [19], celem maksymalizacji dochodu netto dla kilku upraw podlegających ograniczeniom dostępności wody i obszaru uprawy, dokonują linearyzacji nieliniowej funkcji celu. Linearyzacja jest typowym narzędziem stosowanym w celu przeprowadzenia procesu optymalizacji [37].

1.4. Krytyczna ocena niektórych wyników prezentowanych w literaturze

Liczne prace [26, 29-30, 40, 48, 55-58] sygnalizują konieczność ostrożności przy wykorzystywaniu metod *LP*. Wynika to głównie z niejednoznaczności raportów wrażliwości zawierających wektory cen dualnych [55-58]. W tej pracy zauważa się, że niejednoznaczność

wektorów cen dualnych jest konsekwencją liniowej zależności wektorów wierszowych parametrów lewych stron wiążących warunków ograniczających modelu *LP*. Zależność ta musi mieć miejsce, gdy liczba m_0 wiążących warunków ograniczających przewyższa liczbę zmiennych decyzyjnych *n*. Lecz zależność ta jest zwykle niekontrolowana przez analityków. Koltai i Terlaky [29] wskazują, że pojawia się ona niemal zawsze i to już dla niewielkich liczby *m* warunków ograniczających. Gdyby problem niejednoznaczności wektora cen dualnych nie występował w modelu dualnym, to cykliczność procedur postoptymalizacyjnych szybko do niego doprowadzi. Zauważa się, że w cytowanej literaturze modele gospodarowania wodą zawierają wartości *m* oraz *n* bardzo często określane nawet w setkach lub tysiącach. A przecież już dla niewielkich ich wartości winno to rodzić u analityków wielką ostrożność w operowaniu cenami dualnymi. Zauważa się tu brak niezbędnego krytycyzmu u autorów wielu prac wykorzystujących ceny dualne do zagadnień ekonomicznych i technicznych. I tak, pojedynczy niejednoznaczny wektor cen dualnych w dystrybucji wody nie powinien stać się podstawą do podejmowania decyzji zarówno normującej rynkową cenę wody jak i wpływającej na procedury modernizujące sieć wodociągową. W tej pracy sygnalizuje się konieczność rozważania kilku alternatywnych technicznych scenariuszy decyzyjnych opartych o niektóre węzłowe rozwiązania w modelu dualnym do pierwotnego modelu *LP*. Stąd dalej formułuje się metodykę tworzenia tych alternatywnych technicznych scenariuszy decyzyjnych. Autorzy dostrzegają, że dotychczas dające się niekiedy zauważyć w literaturze powątpiewania o użyteczności cen dualnych jest prawdopodobnie powiązane z brakiem wiedzy o ich niejednoznaczności jak również z brakiem praktycznych metod wykorzystania tego faktu. Wyjątkiem mogą być tu prace [24, 55-58]. W rozdziale 2 na prostym przykładzie sieci wodociągowej z nieskończenie wieloma wektorami cen dualnych wskazuje się jak zaproponować zręby alternatywnych scenariuszy modernizacji już uprzednio zoptymalizowanej sieci.

1.5. Interpretacja cen dualnych. Problemy w analizie wrażliwości

Raporty analizy wrażliwości dla *LP* wskazują cechy determinujące wybór optymalnego wariantu decyzyjnego:

1. opisują niektóre proste skutki wywołane odstąpieniem od planu optymalnego (tj. wysokość przyrostu krańcowego czyli wielkość, o którą należy skorygować optymalną wartość funkcji celu oraz zakres obowiązywania korekty w tej wysokości);
2. wskazują jak długo nie należy zmieniać wariantu optymalnej decyzji przy zmianie pojedynczych parametrów liniowej funkcji celu;
3. wskazują, poprzez ceny dualne, jak zmieni się optymalna wartość funkcji celu przy zmianie (niekoniecznie pojedynczych) parametrów prawych stron poszczególnych warunków ograniczających w pewnym zakresie;
4. wskazują, czy opłaca się firmie zwiększyć dostępności określonego zasobu o określoną liczbę jednostek (analiza „więcej za mniej" Arsham [4]).

Ogólnie, cena dualna mierzy zmianę wartości funkcji celu wynikającą ze zwiększenia dostępności określonego zasobu o jednostkę, zazwyczaj przy wyraźnie niedomówionym a domyślnym założeniu, że nie zmieni się ilości pozostałych „deficytowych" zasobów. Każdemu ograniczonemu zasobowi odpowiada wówczas osobna cena dualna. Każdorazowo bada się wtedy wpływ zmiany ilości każdego pojedynczego „deficytowego" czynnika (tj. tego dla którego ograniczenie jest wiążące). Ceny dualne pozostają stałe, dopóki nie zmieni się zbiór wiążących ograniczeń rozwiązania optymalnego. Każda z tych cen mierzy wartość korzyści z rozszerzenia zdolności produkcyjnych lub strat wynikających z ich zmniejszenia. Innymi słowy cena dualna odpowiadająca prawej stronie określonego warunku ograniczającego wskazuje, o ile zmieni się wartość funkcji celu przy rozluźnieniu tego ograniczenia. Jeżeli określony czynnik produkcji, czyli pewien zasób, nie jest w pełni wykorzystany w rozwiązaniu

optymalnym (tzn. nie jest „deficytowy", nie stanowi ograniczenia wiążącego / non-binding constraint), to ma cenę dualną równą zeru. Może być on dodatkowo częściowo spożytkowany po zwiększeniu innych zasobów lecz nie musi go zabraknąć. Jednak, jak w obecnej pracy wykazano, to każdy inny czynnik o cenie dualnej równej zeru, gdy został już w całości zużyty (tzn. stał się „deficytowy", bo stanowi ograniczenie wiążące), może stać się dodatkowo wymagany w podlegającej oszacowaniu proporcji do zwiększanej ilości pewnych zasobów o niezerowej cenie dualnej. Analiza zmian nie tylko pojedynczego parametru prawych stron warunków ograniczających jest przedmiotem licznych studiów. Szeroki przegląd literatury w tym zakresie prezentują Shahin i in. [48]. Dokonana tam klasyfikacja pozwala wyróżnić oprócz zwykłej analizy wrażliwości (ordinary sensitivity analysis) jeszcze 7 innych typów analizy postoptymalizacyjnej: 1) regułę 100% [6], 2) „symetryczną tolerancję" – [59-61], 3) „niesymetryczną tolerancję" będącą rozszerzeniem tolerancji symetrycznej, a wprowadzoną przez Arshama i Oblaka [5], Wondolowskiego [63] i Wendella [62], 4) (*PLP*) parametryczne programowanie liniowe [21, 47], 5) wieloparametryczne programowanie liniowe [54], 6) analizę wrażliwości przy funkcjonalnej zależności parametrów prawych stron warunków ograniczających lub współczynników funkcji celu [23], 7) analizę wrażliwości ze skorelowanymi ww. parametrami [48]. Praca Arshama [4] dobrze porównuje większość z ww. metod poprzez konstrukcje największego regionu wrażliwości dla ogólnego *LP*. Tym samym Arsham [4] wskazuje większość z tych typów jako szczególne przypadki swej analizy. Niemniej nadal pozostaje nierozstrzygnięta kwestia wielu szczególnych przypadków zwłaszcza zdegenerowanych o czym wyraźnie Arsham [4] tam sygnalizuje. Należy tu zauważyć, że zachowanie określonych (najlepiej optymalnych) proporcji przy zwiększaniu zasobów stanowi podstawę (optymalnego) *PLP*. W innym przypadku część zwiększonych zasobów może pozostać niewykorzystana, czyli nieproduktywna. Tym samym dostrzega się potrzebę sformułowania zasad optymalnego doboru najbardziej produktywnych wektorów do *PLP*, co jest przedmiotem obecnej pracy. Innym nierozstrzygniętym w literaturze problemem jest praktyczna użyteczność niejednoznacznych raportów analizy wrażliwości lub choćby wskazywanie faktu pojawiania się ich przez popularne pakiety obliczeniowe [26]. Fakt istnienia nieskończenie wielu rozwiązań modelu dualnego [55-56] dla przeciętnego analityka czyni raport wrażliwości najczęściej nieprzydatnym z uwagi na kłopoty z interpretacją uzyskanych niejednoznacznych raportów. Dotychczas ten ostatni problem został opisany tylko częściowo i to w jednostkowych przypadkach modeli transportowych [57-58]. Prezentowane tam modele transportowe stanowią szczególny przypadek analizy wrażliwości przy funkcyjnej zależności parametrów prawych stron warunków ograniczających [23].

Zatem autorzy tej pracy formułują zasady określenia proporcjonalności równoczesnych zmian parametrów prawych stron warunków ograniczających w przypadku niejednoznacznych raportów analizy wrażliwości. Jednocześnie identyfikują liniową zależność wektorów wierszowych parametrów lewych stron wiążących warunków ograniczających jako przyczynę niejednoznaczności cen dualnych. W tych warunkach wykorzystują różne możliwe do uzyskania powszechnie dostępnym oprogramowaniem raporty zwykłej analizy wrażliwości. Wskazują odmienność interpretacji ceny dualnej odpowiadającej pierwszemu warunkowi ograniczającemu dla każdego z uzyskanych raportów analizy wrażliwości (tabele 2, 3, 6). Odmienność ta polega na tym, że jednostkowa zmiana prawej strony pierwszego warunku ograniczającego dla każdego raportu wymusza równoczesną odpowiednio proporcjonalną zmianę prawej strony innych warunków ograniczających. W rozpatrywanym przykładzie autorzy wskazują, że oszczędność zużycia wody o jedną jednostkę może wymagać równoczesnego spełnienia jednego z dwóch alternatywnych scenariuszy działań technicznych. Przy tym każdy z tych dwóch scenariuszy technicznych pozwala na zaoszczędzenie innej ilości wody i skutkuje innymi oszczędnościami finansowymi. Wybór pomiędzy tymi scenariuszami wymaga uwzględnienia dodatkowych informacji nieprecyzowanych w tym przykładzie. Celem

klarownej prezentacji nowej metodyki przydatnej w procesach eksploatacyjnych wielu złożonych systemów technicznych ograniczono się do przeprowadzenia studium uproszczonego modelu gospodarowania wodą.

2. Studium uproszczonego modelu gospodarowania wodą

Autorzy wskazali już powyżej, że każdy model *LP* dla sieci wodociągów z pojedynczym niejednoznacznym rozwiązaniem zadania dualnego samodzielnie nie powinien być podstawą do kształtowania rynkowych cen wody lub modyfikacji parametrów tej sieci bez skonstruowania kilku alternatywnych scenariuszy technicznego postępowania. Zaś pełna analiza już niewielkiego modelu *LP* prowadzi do wielu pobocznych wątków, które nie są istotne dla prezentacji metodyki tworzenia zrębów alternatywnych scenariuszy technicznego postępowania. Stąd w poniższym uproszczonym przykładzie nadano szczegółową interpretację jedynie dwóm warunkom ograniczającym przy częściowej interpretacji większości pozostałych warunków ograniczających modelu *LP*. Jest to celowy zabieg autorów by wskazać na szeroki zakres przydatności poniższej metodyki przy podkreśleniu znaczenia ograniczenia zużycia wody i kosztów jej pozyskania. Dokonując niewielkich zmian interpretacyjnych poniższy przykład równie dobrze może być odnoszony do osiedlowej sieci wodociągów, która podłączona jest do trzech węzłów poboru wody z większego systemu zaopatrzenia w wodę dużej aglomeracji miejsko-przemysłowej [por. 38]. Taki model zapotrzebowania w wodę uzupełniać można woj. o dynamikę zmian sezonowych i rytmu tygodniowego [por. 35]. Wówczas zmianie ulegają niektóre parametry modelu oraz zmianie ulega interpretacja poszczególnych warunków ograniczających. W każdym z tych przypadków problem szczegółowej interpretacji pojedynczych warunków ograniczających będzie musiał być roztrzygnięty indywidualnie, lecz za każdym razem powrócą te same rozwiązywane tu trudności: 1) klarowne opisanie nieprzeliczalnego zbioru wszystkich niejednoznacznych cen dualnych, 2) nadanie praktycznej interpretacji tej niejednoznaczności woj. poprzez sformułowanie alternatywnych scenariuszy postoptymalizacji dla sieci wodociągowej. Aby pokonywać te trudności analityk systemów logistyki miejskiej winien zapoznać się z przedstawionym tu formalizmem matematycznym opisującym metodykę przejścia od typowych dla *LP* raportów wyników i niejednoznacznych raportów wrażliwości do postulowanych scenariuszy. By było to klarowne niezbędnym jest naszkicowanie modelu dla niewielkiego i niezbyt skomplikowanego przykładu dystrybucji wody, który zbędnie nie eksponuje problemów technicznych lecz skupia się na poprawieniu metodyki inżynierskiego wnioskowania.

Stąd celem uzyskania prostoty i ustalenia uwagi na metodyce poszerzającej postoptymalizację załóżmy, że analizowana aglomeracja miejsko-przemysłowa zasilana jest w wodę przez trzy ujęcia wody znajdujące się na tym samym cieku w miejscach poboru P_i dla $i = 1, 2, 3$ (jak na rys. 1).

Rys.1 Trzy ujęcia wody dla aglomeracji miejsko-przemysłowej znajdujące się na tym samym cieku, przekrój wodowskazowy

Dobowy rozbiór wody w aglomeracji jest wielkością zmienną. Zatem ilość wody pobierana z ujęcia nr i to zmienna decyzyjna $x_i \geq 0$, dla $i = 1, 2, 3$, a okresowo może być ona

wyznaczana w różnych jednostkach woj. [$dm^3 \cdot s^{-1}$], [$dm^3 \cdot h^{-1}$], [$m^3 \cdot d^{-1}$], [$m^3 \cdot miesiąc^{-1}$], [$m^3 \cdot rok^{-1}$]. Nadto załóżmy, że aktualne okresowe zapotrzebowanie aglomeracji na wodę jest zmienne i zawiera się w przedziale ⟨200; 300⟩ jednostek, co można zinterpretować jako ograniczenia w postaci:

WO_1: $x_1 + x_2 + x_3 \geq 200$ WO_2: $x_1 + x_2 + x_3 \leq 300$

Z uwagi na zasadę nienaruszalności przepływu wody w rzece w przekroju wodowskazowym *„S”* lub względy techniczne zależności między poborami wyrażają się w postaci ograniczeń:

WO_3: $9 \cdot x_1 + 11 \cdot x_2 + 7 \cdot x_3 \leq 1800$ WO_4: $5 \cdot x_1 + 6 \cdot x_2 + 4 \cdot x_3 \geq 1000$

WO_5: $2 \cdot x_1 + 4 \cdot x_2 + 3 \cdot x_3 \geq 600$ WO_6: $0 \cdot x_1 + 2 \cdot x_2 + 1 \cdot x_3 \geq 200$

Przy tym przepływ nienaruszalny (*ang. Instream flow, minimum acceptable flow*) określany jest jako ilość wody, która powinna być pozostawiona w danym przekroju cieku ze względów: biologicznych, ekologicznych, społecznych. Konieczność zachowania tego przepływu nie powinna podlegać ocenom ekonomicznym. W związku z tym brane pod uwagę w bilansie wodnym zasoby powinny być pomniejszone o przepływy nienaruszalne. Przy tym kryterium hydrobiologiczne określa minimalny przepływ dla zachowania życia flory i fauny w środowisku wodnym. Kryterium ochrony środowiska określa minimalny przepływ zapewniający zachowanie równowagi stanów wód powierzchniowych i podziemnych w obrębie parków narodowych, rezerwatów przyrody i stref chronionego krajobrazu. Kryterium rybacko-wędkarskie określa minimalny przepływ umożliwiający rozwój ryb. Kryterium sportu i turystyki wodnej określa minimalne stany wód i odpowiadające im przepływy umożliwiające turystykę wodną [9-10, 43]. W poruszanych zagadnieniach hydrologicznych ważną rolę odgrywa także kryterium techniczne odnoszące się do technologicznych możliwości systemu poboru wody przez stosowane specjalistyczne urządzenia. Również niektóre takie ograniczenia można odnieść do niezawodnego funkcjonowania kluczowych dziedzin gospodarki jak energetyka czy transport wodny.

Łączny koszt poboru wody z tych trzech ujęć opisuje minimalizowana funkcja celu $FC: \boldsymbol{c} \cdot \boldsymbol{x} \rightarrow \boldsymbol{min}$. Przy tym wektor współczynników kosztów ma postać

$$\boldsymbol{c} = [c_1 \quad c_2 \quad c_3] = [3 \quad 4 \quad 3], \quad \text{zaś} \quad \boldsymbol{x} = [x_1 \quad x_2 \quad x_3]^T$$

stanowi wektor zmiennych decyzyjnych. Celem optymalizacji zmiennych decyzyjnych wykorzystano dodatek *Solver* w programie *Excel* (por. tabele 1 i 2). Przy warunkach ograniczających *woj.* dla $j = 1, \dots, 6$ wektor optymalnych zmiennych decyzyjnych

$$\boldsymbol{x}^* = [x_1^* \quad x_2^* \quad x_3^*]^T = [66.(6) \quad 66.(6) \quad 66.(6)]^T$$

świadczy, że trzy identyczne wielkości poboru wody z trzech ujęć wyznaczają minimalną wartość funkcji celu na poziomie 666.67 jednostek pieniężnych. Uzyskane rozwiązanie pozostanie niezmienione, gdy jednostkowy koszt poboru wody z ujęcia P_1 pozostanie w zakresie od 2 do 3.5 jednostek pieniężnych, a przy tym koszt pozyskania wody z pozostałych źródeł nie ulegnie zmianie. Zatem przykładowo przy wzroście jednostkowego kosztu poboru wody z tego ujęcia o 0.3 jednostki pieniężne rozwiązanie optymalne $\boldsymbol{x}^*$ nie ulegnie zmianie, lecz koszt łączny wrośnie o 20 jednostek pieniężnych. Podobnie z górnej części tabeli 2 można odczytać dopuszczalne zmiany pozostałych pojedynczych parametrów funkcji kosztów całkowitych, aby rozwiązanie optymalne $\boldsymbol{x}^*$ nie uległo zmianie oraz obliczyć odpowiadającą temu zmianę łącznego kosztu pozyskania wody.

Załóżmy, że krytyczny stan systemu wymusza zmniejszenie poboru wody tj. naruszenie prawej strony WO_1. *Które inne parametry modelu i jak należy zmodyfikować, aby zmniejszenie poboru wody było wykonalne oraz powiązane z optymalnym spadkiem kosztów?*

Tabela 1. Optymalny pobór wody dla aglomeracji miejsko-przemysłowej z trzech ujęć wody – MP. Raport wyników w programie *Excel*

Objective Cell (Min)

Cell	Name	Original Value	Final Value
$E$10	Total Cost	10	666.6666667

Variable Cells

Cell	Name	Original Value	Final Value	Integer
$A$2	x1	1	66.66666667	Contin
$B$2	x2	1	66.66666667	Contin
$C$2	x3	1	66.66666667	Contin

Constraints

Cell	Name	Cell Value	Formula	Status	Slack
$E$4	Left Side Constraint 1	200	$E$4>=$G$4	Binding	0
$E$5	Left Side Constraint 2	200	$E$5<=$G$5	Not Binding	100
$E$6	Left Side Constraint 3	1800	$E$6<=$G$6	Binding	0
$E$7	Left Side Constraint 4	1000	$E$7>=$G$7	Binding	0
$E$8	Left Side Constraint 5	600	$E$8>=$G$8	Binding	0
$E$9	Left Side Constraint 6	200	$E$9>=$G$9	Binding	0

Tabela 2. Raport wrażliwości optymalizacji w modelu pierwotnym dla przykładu w programie Excel.

Variable cells

Cell	Name	Final Value	Reduced Cost	Coefficient Objective	Allowable Increase	Allowable Decrease
$A$2	x1	66.66666667	0	3	0.5	1
$B$2	x2	66.66666667	0	4	0.285714286	1
$C$2	x3	66.66666667	0	3	1E+30	0.25

Constraints

Cell	Name	Final Value	Price dual	Constraint R. H. Side	Allowable Increase	Allowable Decrease
$E$4	Left Side Constraint 1	200	0	200	0	1E+30
$E$5	Left Side Constraint 2	200	0	300	1E+30	100
$E$6	Left Side Constraint 3	1800	-0.6666667	1800	0	1.42109E-14
$E$7	Left Side Constraint 4	1000	1.6666667	1000	7.10543E-15	0
$E$8	Left Side Constraint 5	600	0.3333333	600	100	2.84217E-14
$E$9	Left Side Constraint 6	200	0	200	2.84217E-14	1E+30

Z uwagi na 6 warunków ograniczających przy 3 zmiennych decyzyjnych wnioskujemy, że wektory wierszowe parametrów lewych stron warunków ograniczających są liniowo zależne. Ponadto z uwagi na 5 wiążących warunków ograniczających i 1 niewiążący ujawnionych w tabeli 1, przy 3 zmiennych decyzyjnych wnioskujemy, że rozwiązanie zadania pierwotnego jest jednoznaczne, lecz rozwiązania dualne tworzą pewien podzbiór przestrzeni 2 wymiarowej. Stąd celem przeprowadzenia prawidłowej postoptymalizacji należy dokładnie zbadać *model dualny* (*MD*) do zaprezentowanego tu *modelu pierwotnego* (*MP*) [24, 52]. Dzieje się tak z uwagi na niejednoznaczność wyników optymalizacji modelu dualnego i wynikającą z tego problematyczność we wnioskowaniu. Oznacza to, że nieskończenie wiele wyników optymalizacji modelu dualnego sprawia kłopot w interpretacji zagadnień biznesowych i technicznych [55-56]. Uzyskanie wielu węzłowych rozwiązań modelu dualnego na podstawie *MP* jest pewnym problemem. Jednym ze sposobów uzyskania kolejnych rozwiązań węzłowych jest zmiana kolejności wprowadzania warunków ograniczających [24]. Lecz wymaga to przebadania znacznej liczby permutacji spośród aż 6! = 720. Pomocną okazuje się optymalizacja w modelu dualnym do *MP*. Model *MD* przyjmuje postać [12, 62]:

FC_{MD}: $200 \cdot y_1 + 300 \cdot y_2 + 1800 \cdot y_3 + 1000 \cdot y_4 + 600 \cdot y_5 + 200 \cdot y_6 \rightarrow max$

WO_{MD1}: $1 \cdot y_1 + 1 \cdot y_2 + 9 \cdot y_3 + 5 \cdot y_4 + 2 \cdot y_5 + 0 \cdot y_6 \leq 3$

WO_{MD2}: $1 \cdot y_1 + 1 \cdot y_2 + 11 \cdot y_3 + 6 \cdot y_4 + 4 \cdot y_5 + 2 \cdot y_6 \leq 4$

WO_{MD3}: $1 \cdot y_1 + 1 \cdot y_2 + 7 \cdot y_3 + 4 \cdot y_4 + 3 \cdot y_5 + 1 \cdot y_6 \leq 3$

WB_{MD}: $y_j \leq 0$ dla $j = 2, 3$, $y_j \geq 0$ dla $j = 1, 4, 5, 6$.

Z uwagi na to, że WO_2 oraz WO_3 są nierównościami niestandardowymi w *MP*, zatem odpowiadające im dualne zmienne decyzyjne w *MD* są niedodatnie [49, p. 104]. Ponieważ zaś WO_2 jest niewiążącym warunkiem ograniczającym, to odpowiadająca mu wartość dualnej zmiennej decyzyjnej y_2 jest równa zero. Rzeczywiście w *raporcie wrażliwości* dla *MP* uzyskanego w programie *Excel* (tabela 2) kolumna *cena dualna* zawiera jedno z rozwiązań optymalnych w *MD* i ma ono postać:

$$\boldsymbol{y}_A^* = \left[y_{A_1}^*, y_{A_2}^*, y_{A_3}^*, y_{A_4}^*, y_{A_5}^*, y_{A_6}^*\right]^T = [0 \quad 0 \quad -0.(6) \quad 1.(6) \quad 0.(3) \quad 0]^T = 1/3 \cdot \tilde{\boldsymbol{y}}_A^* .$$

Zaś w *raporcie wrażliwości* dla *MD* uzyskanego w programie *Excel* (tabela 3) kolumna *wartość końcowa* zawiera inne rozwiązanie optymalne w MD i ma ono postać

$$\boldsymbol{y}_B^* = \left[y_{B_1}^*, y_{B_2}^*, y_{B_3}^*, y_{B_4}^*, y_{B_5}^*, y_{B_6}^*\right]^T = [0.(6) \quad 0 \quad 0 \quad 0.(3) \quad 0.(3) \quad 0]^T = 1/3 \cdot \tilde{\boldsymbol{y}}_B^* .$$

Granice dopuszczalnego wzrostu i spadku wartości $\boldsymbol{b}$ prawej strony warunków ograniczających *MP* można odczytać odpowiednio z dolnej części tabeli 2 lub z górnej części tabeli 3.

Tabela 3. Raport wrażliwości w modelu dualnym uzyskany w programie Excel.

Variable cells

Cell	Name	Final Value	Reduced Cost	Coefficient Objective	Allowable Increase	Allowab Decrea
$A$1	y1	0.66666667	0	200	0	
$B$1	y2	0	100	300	1E+30	1
$C$1	y3	0	0	1800	1E+30	
$D$1	y4	0.33333333	0	1000	0	2
$E$1	y5	0.33333333	0	600	100	
$F$1	y6	0	0	200	0	1E+

Constraints

Cell	Name	Final Value	Price dual	Constraint R. H. Side	Allowable Increase	Allowab Decrea
$H$3	Left Side Constraint 1	3	66.66666667	3	0.5	
$H$4	Left Side Constraint 2	4	66.66666667	4	0.285714286	
$H$5	Left Side Constraint 3	3	66.66666667	3	0.5	0

Tabela 4. Wektor $\boldsymbol{y}_c^*$ cen dualnych i odpowiadające mu granice dopuszczalnego wzrostu i spadku wartości $\boldsymbol{b}$ prawej strony warunków ograniczających w MP; na podstawie ostatniej macierzy metody simplex uzyskanej wg. [65]

Objective Cell (Min)

Cell	Name	Original Value	Final Value
$E$10	Total Cost	666.6666667	650

Variable Cells

Cell	Name	Original Value	Final Value	Integer
$A$2	x1	66.66666666	50	Contin
$B$2	x2	66.66666666	125	Contin
$C$2	x3	66.66666666	3.55271E-14	Contin

Constraints

Cell	Name	Cell Value	Formula	Status	Slack
$E$4	Left Side Constraint 1	175	$E$4>=$G$4	Binding	0
$E$5	Left Side Constraint 2	175	$E$5<=$G$5	Not Binding	125
$E$6	Left Side Constraint 3	1825	$E$6<=$G$6	Binding	0
$E$7	Left Side Constraint 4	1000	$E$7>=$G$7	Binding	0
$E$8	Left Side Constraint 5	600	$E$8>=$G$8	Binding	0
$E$9	Left Side Constraint 6	250	$E$9>=$G$9	Binding	0

Nieco odmienny wektor cen dualnych (por tabela 4):

$$\boldsymbol{y}_C^* = \left[y_{C_1}^*, y_{C_2}^*, y_{C_3}^*, y_{C_4}^*, y_{C_5}^*, y_{C_6}^*\right]^T = [0 \quad 0 \quad -1.(3) \quad 3 \quad 0 \quad 0.(3)]^T = 1/3 \cdot \tilde{\boldsymbol{y}}_C^*$$

uzyskano za pomocą aplikacji dostępnej na stronie internetowej [65] dla modelu *MP*. Dodatkowo wówczas z uzyskanej tam ostatniej macierzy metody simplex odczytano macierz odwrotną $\boldsymbol{B}^{-1}$ do tzw. Bazy (uzupełnionej dodatkowymi zmiennymi, zgodnie z regułą metody simplex) oraz wyznaczono zawarte w tabeli 4 granice dopuszczalnego wzrostu i spadku wartości $\boldsymbol{b}$ prawej strony warunków ograniczających *MP* poprzez rozwiązanie odpowiedniej nierówności:

$$[x_3^* \quad x_5^* \quad x_1^* \quad x_8^* \quad x_4^* \quad x_2^*]^T = \boldsymbol{B^{-1}}(\boldsymbol{b} + \boldsymbol{\Delta b}) \geq 0$$

ze względu na wektor $\boldsymbol{\Delta b}$ zmian prawych stron warunków ograniczających [49, p. 79].

Dla wszystkich $\alpha, \beta, \gamma \in \langle 0; 1\rangle$ takich, że $\alpha + \beta + \gamma = 1$ dowolna wypukła kombinacja liniowa $\alpha \cdot \boldsymbol{y}_A^* + \beta \cdot \boldsymbol{y}_B^* + \gamma \cdot \boldsymbol{y}_C^*$ trzech wektorów $\boldsymbol{y}_A^*, \boldsymbol{y}_B^*, \boldsymbol{y}_C^*$ również stanowi wektor rozwiązań dualnych. Nadto dla wektora kolumnowego $\boldsymbol{b}$ prawych stron warunków ograniczających *MP* zgodnie z twierdzeniem Gale'a – Kuhna – Tuckera [20] mamy:

$$\begin{gathered}\boldsymbol{c} \cdot \boldsymbol{x}^* = \boldsymbol{b}^T \cdot \boldsymbol{y}_A^* = \boldsymbol{b}^T \cdot \boldsymbol{y}_B^* = \boldsymbol{b}^T \cdot \boldsymbol{y}_C^* = \\ = \boldsymbol{b}^T \cdot [\alpha \cdot \boldsymbol{y}_A^* + \beta \cdot \boldsymbol{y}_B^* + \gamma \cdot \boldsymbol{y}_C^*] = \\ = 1/3 \cdot \boldsymbol{b}^T \cdot [\alpha \cdot \tilde{\boldsymbol{y}}_A^* + \beta \cdot \tilde{\boldsymbol{y}}_B^* + \gamma \cdot \tilde{\boldsymbol{y}}_C^*] = \\ = 1/3 \cdot \boldsymbol{b}^T \cdot [\tilde{\boldsymbol{y}}_B^* + \gamma \cdot (\tilde{\boldsymbol{y}}_C^* - \tilde{\boldsymbol{y}}_A^*) - (\alpha + \gamma) \cdot (\tilde{\boldsymbol{y}}_B^* - \tilde{\boldsymbol{y}}_A^*)] = \\ = 1/3 \cdot \boldsymbol{b}^T \cdot [\tilde{\boldsymbol{y}}_B^* + \gamma \cdot \boldsymbol{w} + \lambda \cdot \mathbf{v}] = \boldsymbol{b}^T \cdot [\boldsymbol{y}^*(\gamma, \lambda)] = 666.(6)\ .\end{gathered}$$

Przy czym $\lambda = \beta - 1 = -(\alpha + \gamma)$, $\boldsymbol{w} = \tilde{\boldsymbol{y}}_C^* - \tilde{\boldsymbol{y}}_A^*$, $\boldsymbol{v} = \tilde{\boldsymbol{y}}_B^* - \tilde{\boldsymbol{y}}_A^*$, nadto $\boldsymbol{b}^T \cdot \boldsymbol{w} = 0$ oraz $\boldsymbol{b}^T \cdot \boldsymbol{v} = 0$ dla $\boldsymbol{w}^T = [0 \quad 0 \quad -2 \quad 4 \quad -1 \quad 1]$ oraz $\boldsymbol{v}^T = 2 \cdot [1 \quad 0 \quad 1 \quad -2 \quad 0 \quad 0]$ oznacza ortogonalność wektorów $\boldsymbol{w}$ i $\boldsymbol{v}$ do wektora $\boldsymbol{b}$ prawej strony warunków ograniczających w *MP* będącego równocześnie wektorem współczynników funkcji celu modelu dualnego. Zauważmy, że obydwa wektory $\boldsymbol{w}$ i $\boldsymbol{v}$ ortogonalne do wektora $\boldsymbol{b}$ można wskazać jako te wiersze macierzy $\boldsymbol{B}^{-1}$ odwrotnej do macierzy bazy $\boldsymbol{B}$, które są ortogonalne do wektora $\boldsymbol{b}$ i poprzez to w iloczynie

$$\boldsymbol{B}^{-1} \cdot \boldsymbol{b} = [x_3^* \quad x_5^* \quad x_1^* \quad x_8^* \quad x_4^* \quad x_2^*]^T = [66.(6) \quad 100 \quad 66.(6) \quad 0 \quad 0 \quad 66.(6)]^T$$

tworzą zerowe wartości zmiennych dualnych pozostających w bazie w ostatniej macierzy metody simplex. Oznacza to, że dotychczasowy wysiłek na uzyskanie jednego z wektorów dualnych $\boldsymbol{y}_B^*$ oraz wektorów $1/3 \cdot \boldsymbol{w}$ i $1/3 \cdot \boldsymbol{v}$ można uprościć wykorzystując jedynie ostatnią macierz metody simplex [58].

Zatem rozwiązaniu optymalnemu $[66.(6) \quad 100 \quad 66.(6) \quad 0 \quad 0 \quad 66.(6)]^{\mathrm{T}}$ w *MP* metodą simplex odpowiada nieskończenie wiele rozwiązań optymalnych $\boldsymbol{y}^*(\gamma, \lambda)$ w *MD*, które w przestrzeni $\boldsymbol{R}^6$ tworzą 2 wymiarowy wypukły fragment tej przestrzeni o postaci parametrycznej

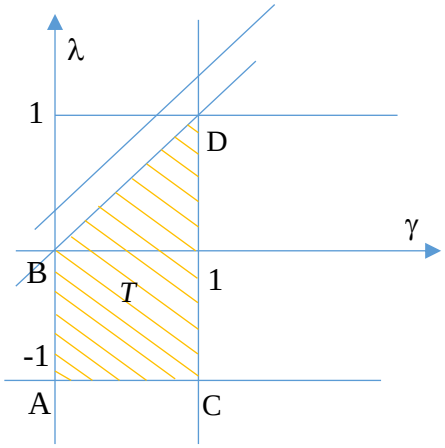
$$\boldsymbol{y}^*(\gamma, \lambda) = 1/3 \cdot (\tilde{\boldsymbol{y}}_B^* + \gamma \cdot \boldsymbol{w} + \lambda \cdot \mathbf{v})\ ,$$

gdzie dla każdej pary parametrów (γ,λ) wektor $\boldsymbol{y}^*(\gamma,\lambda)$ spełnia WB_{MD}. Stąd rozwiązując układ 6 nierówności

$$\frac{1}{3}\cdot\left(\begin{bmatrix}2\\0\\0\\1\\1\\0\end{bmatrix}+\gamma\cdot\begin{bmatrix}0\\0\\-2\\4\\-1\\1\end{bmatrix}+\lambda\cdot 2\cdot\begin{bmatrix}1\\0\\1\\-2\\0\\0\end{bmatrix}\right)\begin{matrix}\geq 0\\ \leq 0\\ \leq 0\\ \geq 0\\ \geq 0\\ \geq 0\end{matrix}$$

wyznaczamy zakres dla parametrów (γ,λ) w postaci figury wypukłej *T*. I tak $T=\{(\gamma,\lambda)\in \boldsymbol{R}^2: 0\leq\gamma\leq 1, -1\leq\lambda\leq\gamma\}$. Zakres T ma kształt trapezu (rys. 2) o wierzchołkach ABCD, gdzie punkt B pokrywa się z początkiem układu współrzędnych $O\gamma\lambda$. Trapez T zawiera uprzednio wyznaczony zakres parametrów $T_0=\{(\gamma,\lambda)\in \boldsymbol{R}^2: 0\leq\gamma\leq 1, -1\leq\lambda\leq-\gamma\}$ w postaci trójkąta o wierzchołkach ABC, odpowiadający wypukłej kombinacji liniowej tylko trzech wektorów

$$\boldsymbol{y}^*(0;-1)=\boldsymbol{y}_A^*,\ \boldsymbol{y}^*(0;0)=\boldsymbol{y}_B^*,\ \boldsymbol{y}^*(1;-1)=\boldsymbol{y}_C^*\ .$$



Rys. 2 Obszar trapezu T o wierzchołkach ABCD wyznacza zakres parametrów (γ,λ), dla których dualny wektor $\boldsymbol{y}^*(\gamma,\lambda)$ jest dopuszczalny w MD

Stąd uzyskujemy dodatkowy węzłowy wektor cen dualnych

$$\boldsymbol{y}^*(1;1)=\boldsymbol{y}_D^*=\left[y_{D_1}^*,y_{D_2}^*,y_{D_3}^*,y_{D_4}^*,y_{D_5}^*,y_{D_6}^*\right]^T=[1.(3)\quad 0\quad 0\quad 0.(3)\quad 0\quad 0.(3)]^T=1/3\cdot\tilde{\boldsymbol{y}}_D^*$$

o nieznanych zakresach zmian parametrów prawej strony *WO* odpowiadających tym cenom.

Dla każdego $(\gamma,\lambda)\in T$ w *MD* wartość funkcji $FC_{MD}(\gamma,\lambda)$ jest stała i wynosi 666.67 zaś wektor $\boldsymbol{y}^*(\gamma,\lambda)$ jest wektorem dopuszczalnym tj. równocześnie spełnione są wszystkie warunki WB_{MD}, a warunki ograniczające modelu dualnego są wówczas warunkami wiążącymi w *MD*. Nadto dla $j=1,\ldots,6$ mamy

$$y_j^*(\gamma,\lambda)\in\langle\min\left(y_{A_j}^*,y_{B_j}^*,y_{C_j}^*,y_{D_j}^*\right),\max\left(y_{A_j}^*,y_{B_j}^*,y_{C_j}^*,y_{D_j}^*\right)\rangle\ .$$

A stąd oraz z tabel 2 – 4 ponieważ $y_2^*(\gamma,\lambda)\equiv 0$, zatem dokonanie zmiany górnej granicy zapotrzebowania na wodę poprzez zmniejszenie prawej strony w WO_2 o 100 jednostek ani dowolne zwiększenie jej nie wpłynie na zmianę optymalnego planu poboru wody $\boldsymbol{x}^*$ i na łączny koszt.

Analizując tabele 2 – 4 wnioskujemy, że istnieją dwa alternatywne sposoby na zmniejszenie łącznych kosztów: a) poprzez wykorzystanie wartości zmiennej dualnej $y_{B_1}^*=0.(6)$ przy tylko pozornie niedozwolonym spadku b_1 lub b) korzystając z dwa razy większej wartości $y_{D_1}^*=1.(3)$ przy nieznanym tu zakresie dozwolonego spadku b_1.

Chociaż $y_1^*(\gamma,\lambda)\in\left(0;1.(3)\right)$, to być może największy spadek łącznych kosztów w wysokości $0.(6)\cdot 25\approx 16.67$ jednostek pieniężnych w drodze zmniejszenia dolnej granicy zapotrzebowania na wodę można uzyskać poprzez (jednak) dopuszczalne zmniejszenie wartości 200 prawej strony pierwszego warunku ograniczającego WO_1 aż o hipotetyczne 25 jednostek, gdy wykorzystamy wartość dualną $y_{B_1}^*=y_1^*(0{,}0)=0.(6)$ odczytaną z górnej

części tabeli 3 dla *MD*, gdy $[\alpha \quad \beta \quad \gamma] = [0 \quad 1 \quad 0]$ tj., gdy $[\gamma \quad \lambda] = [0 \quad 0]$. Lecz taka pojedyncza zmiana WO_1 może jednak nie wnieść nic z uwagi na pozostałe 4 wiążące warunki ograniczające. Informują o tym dwie tabele 2 i 4. Oznacza to, że pojedyncza zmiana prawej strony WO_1, jak w typowej analizie wrażliwości, nietypowo dodatkowo wymaga również zmiany prawych stron innych warunków ograniczających. Jak się dalej okazuje, taka słuszna sugestia wynika z faktu nienaturalnego zablokowania w tabeli 3 zarówno wzrostów jak i spadków parametrów prawej strony WO_1 odpowiadających niezerowej cenie dualnej. Podobne zastrzeżenia dotyczą nienaturalnego całkowitego zablokowania zmian parametrów prawej strony *WO* dla kilku niezerowych cen dualnych w tabelach 2 i 4.

Wyznaczenie wielkości zmiany łącznych kosztów poboru wody $\Delta FC_{MD}(\boldsymbol{\Delta b})$, w zależności od znanego dowolnego wektora $\boldsymbol{\Delta b} = [\Delta b_1 \quad \Delta b_2 \quad \Delta b_3 \quad \Delta b_4 \quad \Delta b_5 \quad \Delta b_6]^T$ zmian prawych stron warunków ograniczających w *MP*, najprościej wymaga powtórnego uruchomienia odpowiedniego oprogramowania. Lecz tu analityk stoi przed odwrotnym problemem, gdyż poszukuje najkorzystniejszego (najbardziej produktywnego) dla $y^*_{B_1}$ całego wektora zmian $\boldsymbol{\Delta b}$. Ponadto tu WO_1 nie jest jedynym wiążącym warunkiem ograniczającym lecz dla 3 zmiennych decyzyjnych jest aż 5 wiążących warunków ograniczających. Zatem dysponując tabelami 2 – 4 powstaje postoptymalizacyjne pytanie o charakterze nie tylko matematycznym lecz również menadżerskim: *Które zmiany Δb_j prawych stron warunków ograniczających, uwzględniające różne problemy techniczno-ekonomiczne poboru wody oraz zasadę nienaruszalności przepływu wody w rzece, muszą współtowarzyszyć:*

a) $y^*_{B_1} = 0.(6)$ *i spadkowi zapotrzebowania o* $\Delta b_1 = 25$ *jednostek wody tj. zaoszczędzeniu* $0.(6) \cdot 25 \approx 16.67$ *jednostek pieniężnych, a które*

b) *2 razy większemu* $y^*_{D_1} = 1.(3)$ *o nieznanym tu spadku zapotrzebowania* Δb_1?

Aby odpowiedzieć na powyższe pytania zauważmy, że $FC = FC_{MD} = \boldsymbol{c} \cdot \boldsymbol{x}^* = \boldsymbol{b}^T \cdot [\boldsymbol{y}^*(\gamma, \lambda)]$, tj. koszty są stałe dla dowolnych dozwolonych wartości parametrów γ oraz λ. Lecz jedynie dla właściwie obranego dozwolonego ustalonego wektora zmian $\boldsymbol{\Delta b}$ również nowe koszty tj. wartość $(\boldsymbol{b} + \boldsymbol{\Delta b})^T \cdot \boldsymbol{y}^*(\gamma, \lambda)$ jest stała dla dowolnych dozwolonych wyżej dobranych wartości parametrów γ oraz λ. Stąd należy tak dobierać wektor $\boldsymbol{\Delta b}$, by zmiana kosztów tj. wartość $\Delta FC_{MD}(\boldsymbol{\Delta b}; \gamma, \lambda) = (\boldsymbol{\Delta b})^T \cdot \boldsymbol{y}^*(\gamma, \lambda)$ nie zależała od doboru parametrów γ oraz λ. Ostatnia równość wskazuje jak dla wykorzystywanego tu całego wektora $\boldsymbol{y}^*(\gamma, \lambda)$ należy dobierać cały wektor $\boldsymbol{\Delta b}$. W szczególności, ponieważ również dla tak dobranych dowolnych parametrów $(\gamma, \lambda) \in T$ mamy $\Delta FC_{MD}(\boldsymbol{\Delta b}; \gamma, \lambda) = 1/3\,(\boldsymbol{\Delta b})^T \cdot [\tilde{\boldsymbol{y}}^*_B + \gamma \cdot \boldsymbol{w} + \lambda \cdot \boldsymbol{v}] = 1/3\,[(\boldsymbol{\Delta b})^T \cdot \tilde{\boldsymbol{y}}^*_B] = \Delta FC_{MD}(\boldsymbol{\Delta b}; 0, 0)$, gdy spełnione są dwa warunki: $(\boldsymbol{\Delta b})^T \cdot \boldsymbol{w} = 0$ oraz $(\boldsymbol{\Delta b})^T \cdot \boldsymbol{v} = 0$. Stąd ostatnie warunki ortogonalności opisują ten dozwolony właściwy sposób doboru składowych wektora zmian $\boldsymbol{\Delta b}$, odpowiadający operowaniu tu całym wektorem $\boldsymbol{y}^*_B$.

Ad a) *Co jednak, gdy podobnie jak w klasycznej analizie wrażliwości wykorzystać należy jedynie odczytaną z tabeli 3 pojedynczą składową* $y^*_{B_1} = y^*_1(0; 0) = 0.(6)$ *wektora* $\boldsymbol{y}^*_B$, *tzn. gdy:* $\Delta FC_{MD}(\boldsymbol{\Delta b}; \gamma, \lambda) = \Delta b_1 \cdot y^*_{B_1}$?

Ponieważ ostatnia równość jest równoważna warunkowi

$$\sum_{j=1}^{6} \left(\Delta \mathrm{b_j} \cdot y^*_{B_j}\right) = \Delta b_1 \cdot y^*_{B_1} ,$$

tzn. powstaje wymaganie aby oprócz $y^*_{B_1}$ pozostałe niezerowe elementy składowe $y^*_{B_j}$ wektora $\boldsymbol{y}^*_B$ nie wpłynęły na wyznaczaną wartość zmiany dualnej funkcji celu. W tym celu po pierwsze przyjmujemy $\Delta \mathrm{b_4} = \Delta \mathrm{b_5} = 0$. Po drugie dla tak dobranych dowolnych parametrów γ oraz λ musi zachodzić równanie

$$\Delta FC_{MD}(\boldsymbol{\Delta b}; \gamma, \lambda) = \tfrac{1}{3} \cdot (\boldsymbol{\Delta b})^T \cdot [\tilde{\boldsymbol{y}}^*_B + \gamma \cdot \boldsymbol{w} + \lambda \cdot \boldsymbol{v}] =$$

$$= \frac{1}{3} \cdot \begin{bmatrix} \Delta b_1 \\ \Delta b_2 \\ \Delta b_3 \\ 0 \\ 0 \\ \Delta b_6 \end{bmatrix}^T \cdot \left(\begin{bmatrix} 2 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \gamma \cdot \begin{bmatrix} 0 \\ 0 \\ -2 \\ 4 \\ -1 \\ 1 \end{bmatrix} + \lambda \cdot \begin{bmatrix} 2 \\ 0 \\ 2 \\ -4 \\ 0 \\ 0 \end{bmatrix} \right) = \Delta FC_{MD}(\boldsymbol{\Delta b}; 0{,}0) = \frac{2}{3} \cdot \Delta b_1 \text{ .}$$

Stąd warunki: $(\boldsymbol{\Delta b})^T \cdot \boldsymbol{w} = 0$ oraz $(\boldsymbol{\Delta b})^T \cdot \boldsymbol{v} = 0$ przyjmują postać $-2\Delta b_3 + \Delta b_6 = 0$ oraz $2\Delta b_1 + 2\Delta b_3 = 0$. Zatem jeżeli $-25 \leq \Delta b_1 \leq 28$ oraz $-100 \leq \Delta b_2$ to koszt poboru wody ulegnie zmianie o $\frac{2}{3} \cdot \Delta b_1$ gdy

$$\boldsymbol{\Delta b} = \Delta b_1 \cdot [1 \quad 0 \quad -1 \quad 0 \quad 0 \quad -2]^T + \Delta b_2 \cdot [0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]^T.$$

Oznacza to, że spadek łącznych kosztów o wysokość $0.(6) \cdot 25 \approx 16.67$ jednostek pieniężnych w drodze dopuszczalnego zmniejszenia minimalnego zapotrzebowania na wodę z 200 do 175 jednostek można uzyskać poprzez równoczesny (dopuszczalny) wzrost prawej strony trzeciego (technicznego) warunku ograniczającego WO_3 o 25 jednostek oraz wzrost prawej strony szóstego (również technicznego) warunku ograniczającego WO_6 o 50 jednostek, przy dopuszczalnym spadku prawej strony drugiego warunku ograniczającego WO_2 o 100 jednostek. Słuszność przytoczonego analitycznego wnioskowania potwierdzają tabele 5 i 6 dla *MP* po zmianie 4 spośród 6 parametrów prawych stron warunków ograniczających. Słuszność ta istnieje pomimo wątpliwości, które mogły wzbudzać w tabeli 3 zerowe wartości dopuszczalnego wzrostu dla WO_3 i dla WO_6. Lecz wskazane w tabeli 3 zakresy zmian Δb_j prawych stron dla WO_3 i dla WO_6 dotyczą zmiany tego pojedynczego parametru w połączeniu z odpowiednio zidentyfikowanym zespołem kilku parametrów opisanych wektorem $\boldsymbol{\Delta b}$. Słuszność dokonanych równoczesnych zmian powyżej zidentyfikowanego zespołu parametrów w $\boldsymbol{\Delta b}$ potwierdzono przeprowadzając analizę prowadzącą do tabeli 5 jak i teoretyczne rozważania w rozdz. 3.1.

Tabela 5. Optymalny pobór wody dla aglomeracji miejsko-przemysłowej z trzech ujęć wody. MP po zmianie 4 parametrów prawych stron warunków ograniczających. Raport wyników w programie Excel

Objective Cell (Min)

Cell	Name	Original Value	Final Value
$E$10	Total Cost	666.6666667	650

Variable Cells

Cell	Name	Original Value	Final Value	Integer
$A$2	x1	66.66666666	50	Contin
$B$2	x2	66.66666666	125	Contin
$C$2	x3	66.66666666	3.55271E-14	Contin

Constraints

Cell	Name	Cell Value	Formula	Status	Slack
$E$4	Left Side Constraint 1	175	$E$4>=$G$4	Binding	0
$E$5	Left Side Constraint 2	175	$E$5<=$G$5	Not Binding	125
$E$6	Left Side Constraint 3	1825	$E$6<=$G$6	Binding	0
$E$7	Left Side Constraint 4	1000	$E$7>=$G$7	Binding	0
$E$8	Left Side Constraint 5	600	$E$8>=$G$8	Binding	0
$E$9	Left Side Constraint 6	250	$E$9>=$G$9	Binding	0

Tabela 6. Raport wrażliwości w programie Excel dla MP po zmianie 4 parametrów prawych stron warunków ograniczających

Variable cells

Cell	Name	Final Value	Reduced Cost	Coefficient Objective	Allowable Increase	Allowable Decrease
$A$2	x1	50	0	3	0.5	1
$B$2	x2	125	0	4	0.8	1
$C$2	x3	3.55271E-14	0	3	0.5	0.4

Constraints

Cell	Name	Final Value	Price dual	Constraint R. H. Side	Allowable Increase	Allowable Decrease
$E$4	Left Side Constraint 1	175	1.333333333	200	75	7.10543E-15
$E$5	Left Side Constraint 2	175	0	300	1E+30	125
$E$6	Left Side Constraint 3	1825	0	1800	1E+30	2.66454E-14
$E$7	Left Side Constraint 4	1000	0.333333333	1000	1.33227E-14	150
$E$8	Left Side Constraint 5	600	0	600	1.42109E-14	1E+30
$E$9	Left Side Constraint 6	250	0.333333333	200	75	1.42109E-14

Zauważmy, że samo zmniejszenie minimów dostawy wody z trzech źródeł wyrażone zmniejszeniem wyrazu wolnego w WO_1 nie spowoduje spadku łącznych kosztów poboru wody. Dopiero zespołowe działanie polegające na jednoczesnym wykonaniu trzech wspomnianych wyżej czynności przynosi zamierzony efekt zmniejszenia łącznych kosztów zużycia wody przez całą aglomerację z tego systemu. Najbardziej efektywnym posunięciem jest utrzymanie w technicznych warunkach ograniczających proporcji $\Delta b_3/\Delta b_1 = -1$, $\Delta b_6/\Delta b_1 = -2$ pozwalających na maksymalne ograniczenie zużycia wody $\Delta b_1 = 25$. Przeprowadzona modyfikacja systemu poboru wody prowadzi do wyłączania punktu poboru P_3 w czasie minimalnego zapotrzebowania na wodę i wymaga równocześnie unowocześnienia punktu poboru P_2, by zwiększyć jego zdolność zaopatrywania aglomeracji w wodę.

Ad b) *Jak analogiczna skorygowana procedura analizy wrażliwości wygląda, gdy wykorzystać należy wyrażoną kwotowo jedynie najwyższą pierwszą składową* $y^*_{D_1} = y^*_1(1;1) = 1.(3)$ *wektora* $\boldsymbol{y}^*_D$*, tzn. gdy:* $\Delta FC_{MD}(\boldsymbol{\Delta b};\gamma,\lambda) = \Delta b_1 \cdot y^*_{D_1}$*?*

Wówczas dwuwymiarowy zbiór wektorów dualnych należy przeparametryzować do postaci

$$\boldsymbol{y}^*(\gamma,\lambda) = \boldsymbol{y}^*(1+\gamma_1, 1+\lambda_1) = 1/3 \cdot (\tilde{\boldsymbol{y}}^*_D + \gamma_1 \cdot \boldsymbol{w} + \lambda_1 \cdot \boldsymbol{v})$$

skonstruowanej na podstawie wektorów $\boldsymbol{y}^*_D$, $\boldsymbol{w}$, $\boldsymbol{v}$, gdzie teraz punkt D trapezu ABCD (rys. 2) pokrywa się z początkiem układu współrzędnych $O\gamma_1\lambda_1$. Wtedy podobnie jak poprzednio przyjmujemy $\Delta b_4 = \Delta b_6 = 0$ oraz rozwiązujemy ostatnią z poniższych równości:

$$\Delta FC_{MD}(\boldsymbol{\Delta b};\gamma_1,\lambda_1) = \frac{1}{3} \cdot (\boldsymbol{\Delta b})^T \cdot [\tilde{\boldsymbol{y}}^*_D + \gamma_1 \cdot \boldsymbol{w} + \lambda_1 \cdot \boldsymbol{v}] =$$

$$= \tfrac{1}{3} \cdot \begin{bmatrix} \Delta b_1 \\ \Delta b_2 \\ \Delta b_3 \\ 0 \\ \Delta b_5 \\ 0 \end{bmatrix}^T \cdot \left(\begin{bmatrix} 4 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} + \gamma_1 \cdot \begin{bmatrix} 0 \\ 0 \\ -2 \\ 4 \\ -1 \\ 1 \end{bmatrix} + \lambda_1 \cdot \begin{bmatrix} 2 \\ 0 \\ 2 \\ -4 \\ 0 \\ 0 \end{bmatrix} \right) = \tfrac{4}{3} \cdot \Delta b_1 \,,$$

gdzie warunki $(\boldsymbol{\Delta b})^T \cdot \boldsymbol{w} = 0$ oraz $(\boldsymbol{\Delta b})^T \cdot \boldsymbol{v} = 0$ przyjmują postać $-2\Delta b_3 - \Delta b_5 = 0$ oraz $\Delta b_1 + \Delta b_3 = 0$. Stąd $\boldsymbol{\Delta b} = \Delta b_1 \cdot [1 \quad 0 \quad -1 \quad 0 \quad 2 \quad 0]^T + \Delta b_2 \cdot [0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]^T$.

Oznacza to, że spadek łącznych kosztów aż o wysokość $1.(3) \cdot 20 \approx 26.67$ jednostek pieniężnych w drodze zmniejszenia minimalnego zapotrzebowania na wodę o dozwolone jedynie $\Delta b_1 = 20$ jednostek z 200 do 180 jednostek można uzyskać poprzez równoczesny (dopuszczalny) wzrost prawej strony trzeciego (technicznego) warunku ograniczającego WO_3 o 20 jednostek oraz spadek prawej strony piątego warunku ograniczającego WO_5 o 40 jednostek, przy dopuszczalnym spadku prawej strony drugiego warunku ograniczającego WO_2 o 100 jednostek. Tym razem należy zachować proporcje zmian $\Delta b_3/\Delta b_1 = -1$ oraz $\Delta b_5/\Delta b_1 = 2$.

Wybór pomiędzy tymi dwoma rozwiązaniami a) i b) wymaga dodatkowych informacji o postoptymalizowanym systemie zaopatrzenia w wodę.

3. Dyskusja nad alternatywnymi scenariuszami

3.1. Dwa scenariusze

Zauważamy, że jak w a) zawarte w tabeli 3 zera dla cen dualnych $y^*_{B_3}$ oraz $y^*_{B_6}$ wskazują, że jedynie techniczne wiążące warunki ograniczające WO_3 oraz WO_6 współtowarzyszą łącznie z wiążącym WO_1 przy wyłącznej eksploatacji niezerowej wartości $y^*_{B_1} = 0.(6)$. Zaś jak w b) wykorzystanie wyłącznie $y^*_{D_1} = 1.(3)$ oznacza większą korzyść finansową przy mniejszym spadku minimalnego zapotrzebowania na wodę, lecz wymaga to jednocześnie nieco innego zakresu interwencji związanej z technicznym wiążącym warunkiem ograniczającym WO_3 oraz ograniczeniem parametru związanego z prawą stroną innego warunku ograniczającego WO_5, gdyż $y^*_{D_3} = y^*_{D_5} = 0$. Zatem wybór pomiędzy tymi dwoma rozwiązaniami wymaga pełnego rozważenia informacji o postoptymalizowanym systemie zaopatrzenia w wodę zarówno ujętych w modelu jak i dodatkowych informacji spoza modelu.

Fakt wskazanej powyżej konieczności jednoczesnej proporcjonalnej zmiany kilku parametrów systemu zaopatrzenia w wodę wynika z poniższej szerszej uwagi.

3.2. Interpretacja zerowych cen dualnych

Usunięcie warunku ograniczającego w *MP* skutkuje usunięciem odpowiadającej mu zmiennej dualnej z *MD*. A pozostawienie tej zmiennej dualnej w MD_1 równoważnym z *MD*

wymaga uznania, że w rozwiązaniu optymalnym dla MD_1 wartość tej zmiennej dualnej wynosi zero. Również niewiążącym warunkom ograniczającym w *MP* odpowiadają zerowe wartości zmienne dualne w *MD*. Zatem zerowanie się pewnych zmiennych dualnych wynika z istniejących już luzów w warunku niewiążącym lub wskazuje, że aby wykorzystać pozostałe niezerowe wartości w całym wektorze dualnym zachodzi ewentualna konieczność dokonania niezerowych zmian w prawej stronie warunku ograniczającego tam, gdzie wektor dualny się zeruje. Innymi słowy, zerowej optymalnej wartości zmiennej dualnej odpowiada 1) warunek niewiążący w *MP* albo 2) brak tej zmiennej dualnej w *MD*. Z kolei brak zmiennej dualnej w *MD* wynika z usunięcia skutków odpowiedniego warunku ograniczającego w *MP* co odpowiada pełnemu rozluźnieniu wiązań wynikających z tego warunku ograniczającego lub należnej zmianie wartości prawej strony tego warunku ograniczającego. Oznacza to, że zerowej wartości zmiennej dualnej przy warunku niewiążącym w *MP* odpowiada możliwość samodzielnej zmiany wartości prawej strony tego warunku ograniczającego. Zaś zerowej wartości zmiennej dualnej przy warunku wiążącym w *MP* odpowiada ewentualna konieczność łącznej zmiany wartości prawej strony tego warunku ograniczającego w przypadku próby zmiany prawej strony innego wiążącego warunku w *MP* o odpowiadającej mu niezerowej wartości ceny dualnej.

3.3. Dodatkowe scenariusze techniczne

Na podstawie ostatniego stwierdzenia zauważamy, że spadek zużycia wody (tj. spadek wartości prawej strony WO_1) może być również konsekwencją sześciu dalszych scenariuszy zmieniających: c) prawe strony wiążących warunków WO_3, WO_4, WO_5 w *MP* przy wykorzystaniu odpowiadającym im niezerowych wartości $y^*_{A_3}, y^*_{A_4}, y^*_{A_5}$ wektora $\boldsymbol{y}^*_A$ cen dualnych; d) prawe strony wiążących warunków WO_3, WO_4, WO_6 w *MP* przy wykorzystaniu odpowiadającym im niezerowych wartości $y^*_{C_3}, y^*_{C_4}, y^*_{C_6}$ wektora $\boldsymbol{y}^*_C$ cen dualnych. Konstrukcja tych scenariuszy wymaga m.in. analogicznych reparametryzacji przesuwających początek układu współrzędnych odpowiednio do wierzchołków A oraz C trapezu. Każdy z tych razem ośmiu scenariuszy zmniejszenia zużycia wody przynosi odrębne zmiany w łącznych kosztach zużywanej wody. Scenariusze te są konsekwencjami wykorzystania pojedynczo ośmiu z 12 niezerowych wartości cen dualnych występujących dla węzłów zbioru ABCD. Cztery pozostałe niezerowe wartości nie przynoszą scenariuszy z efektem w postaci zmniejszenia zużycia wody. O wyborze z tych 8 najlepszego scenariusza $\boldsymbol{\Delta b}$ lub kombinacji liniowej dwóch najlepszych z nich zdecyduje szczegółowa analiza m.in. konsekwencji finansowych, technicznych, biologicznych, ekologicznych, społecznych powiązanych z tymi scenariuszami. Wśród rozważanych warunków ograniczających te, które wynikają z zasady nienaruszalności przepływu wody w rzece albo konsekwentnie nie ulegają one zmianie lub należy je też zmodyfikować na drodze prawnej, ekonomicznej lub poprzez dodatkowe kroki techniczne.

Dla wektorów $\boldsymbol{\Delta b}$ położonych na jednostkowej sferze w $\mathbf{R^6}$ tj. gdy

$$|\boldsymbol{\Delta b}| = \sqrt[2]{\sum_{j=1}^{6}\left(\Delta b_j\right)^2} = 1,$$

wówczas zmiana wartości funkcji celu jako iloczyn skalarny:

$$\Delta FC_{MD}(\boldsymbol{\Delta b};\gamma,\lambda) = (\boldsymbol{\Delta b})^T \cdot \boldsymbol{y}^*(\gamma,\lambda)$$

osiąga wartość: i) maksymalną, gdy wektory $\boldsymbol{\Delta b}$ oraz $\boldsymbol{y}^*(\gamma,\lambda)$ mają identyczny kierunek i zwrot; ii) minimalną, gdy wektory $\boldsymbol{\Delta b}$ oraz $\boldsymbol{y}^*(\gamma,\lambda)$ mają identyczny kierunek i przeciwny zwrot; iii) zerową, gdy wektory $\boldsymbol{\Delta b}$ oraz $\boldsymbol{y}^*(\gamma,\lambda)$ są ortogonalne. Fakty te decydują o wyborze scenariuszy o ekstremalnych lub zerowych zmianach łącznych kosztów zużywanej wody i wymagają odrębnej analizy podobnie jak wymienione w rozdziale 1.5 pozostałe 7 typów analizy postoptymalizacyjnej.

O znaczeniu uzyskanego tu wyniku może świadczyć fakt, że dotychczas część badaczy z uwagi na trudności analityczne i interpretacyjne w ogóle odradzała opracowywanie podobnej

analizy [8, p. 44]. Autorzy oprócz *MP* wspomagają się *MD* i różnorodnością oprogramowania, by uniknąć konieczności budowy algorytmu permutacyjnego wspierającego programowanie liniowe celem generowania wszystkich rozwiązań węzłowych w trapezie ABCD. O złożoności obliczeń analitycznych niech świadczy brak możliwości odtworzenia na podstawie tabel 5 i 6 kroków powrotnych do tabeli 3 pozwalających powrócić do początkowego *MP*. Zaś udostępniona w tabeli 6 analiza postoptymalizacyjna, wskazuje możliwość dalszej rekurencyjnej modyfikacji rozważanego systemu poboru wody. Lecz wymaga to powtarzania kroków analitycznych podobnych do tych opisanych tu wyżej. Oczywiście każdorazowe zmiany prawych stron warunków ograniczających w modelu w rozsądnym zakresie winny odzwierciedlać potrzeby eksploatacyjne aglomeracji i możliwości środowiskowe systemu hydrologicznego.

4. Wnioski

1. W swojej klasycznej postaci modele programowania liniowego dla eksploatacji złożonych systemów technicznych najczęściej w raportach wrażliwości zawierają niejednoznaczne odpowiedzi. Stąd w rozdziale 1.4 krytycznie oceniono niektóre wyniki prezentowane w literaturze a odnoszące się do ceny dualnej wody. Ponieważ praca ta poświęcona jest usunięciu zawodności standardowych idealistycznych procedur analizy wrażliwości łączących wartość każdej ceny dualnej wyłącznie ze zmianą tylko pojedynczego parametru prawych stron warunków ograniczających, dlatego zaproponowana tu metodyka postoptymalizacyjna może być szeroko wykorzystana przez inżynierów nie tylko w zarządzaniu procesami eksploatacji.
2. Wnioskowanie o rynkowej cenie wody w oparciu o niejednoznaczną cenę dualną (a takie sytuacje występują w zagadnieniach technicznych prawie zawsze [29]) prowadzi do:
   a) braku rozpoznania wielu scenariuszy modyfikacji systemu wodociągowego, a ewentualny wybór przypadkowego scenariusza może wymagać konieczności przeprowadzenia trudnej lub nawet niemożliwej do osiągnięcia modernizacji systemu wodociągowego.
   b) uzyskania mniejszych, niż dających się osiągnąć na innej drodze, oszczędności w funkcjonowaniu systemu wodociągowego.

   Prezentując zręby scenariuszy technicznych dla uproszczonego przykładu autorzy artykułu prezentują metodykę:

   ad. a) identyfikacji, poprzez warunki ograniczające, części systemu wodociągowego, podlegających zmianie przy wdrażaniu różnych scenariuszy technicznych,

   ad. b) znajdowania różnych oszczędności powiązanych z wdrażaniem alternatywnych scenariuszy, co pozwala wybrać najlepszą modyfikację systemu.
3. Opracowana w pracy metodyka umożliwia wygenerowanie ośmiu alternatywnych scenariuszy technicznych oszczędzania wody, z których nakreślono zręby dwóch.
4. Wyznaczane w pracy proporcje składowych $\boldsymbol{\Delta b}$ służą do uproszczenia i poprawienia procedur jedno i wieloparametrycznego programowania liniowego [54]. A procedury te to również analiza postoptymalizacyjna przy funkcjonalnej zależności zmian parametrów prawych stron warunków ograniczających lub zmian współczynników funkcji celu i są analogiczne do przypadku postoptymalizacji typu 6 z rozdziału 1.5. Problemy te nie są opisane w literaturze, a ulegają dalszej komplikacji przy większej ilości zmiennych decyzyjnych.
5. Praca zawiera zastosowanie wektorów dualnych $\boldsymbol{y}^*$ oraz wektorów $\boldsymbol{w}$, $\boldsymbol{v}$ ortogonalnych do $\mathbf{b}$ przy wyznaczania efektywnych proporcjonalnych zmian $\boldsymbol{\Delta b}$. Tym samym uogólnia znaczenie wektorów bilansujących w modelu transportowym a ortogonalnych do $\mathbf{b}$ i pozwala uogólnić metodę macierzy cen dualnych [24, 55-58] na modele z nierównościowymi warunkami ograniczającymi.

6. Przedstawiona powyżej metodyka może być bezpośrednio dostosowana do postoptymalizacji różnych procesów eksploatacji złożonych systemów technicznych np.:
   a. w systemie poboru wody dla aglomeracji miejskiej z kilku źródeł przy różnych warunkach eksploatacyjnych [9-10, 57];
   b. do kosztów eksploatacji innych wieloźródłowych systemów np. dostaw gazu w miejskiej sieci gazowej [34, 50];
   c. do kosztów eksploatacji aparatury medycznej i infrastruktury technicznej [51];
   d. w zagadnieniach doskonalenia rejonizacji stacji ratownictwa medycznego [39, 52];
   e. do średnich czasów przebywania systemów w podzbiorach stanów niezawodnościowych [31-32].
7. Obecnie autorzy prowadzą prace nad pełną algorytmizacją prezentowanej tu metodyki.

Literatura

1. Abdy Sayyed M A H, Gupta R, Tanyimboh T T. Flow-based penalty in GA for opitmal design of water distribution network. Hydro-2017 International, L.D.College of Engineering, Ahmedabad, India, 2017; 1439 – 1446.
2. Agyei E, Munch D, Burger P. Application of Optimization Modeling to Water Resource Planning in East-Central Florida. Technical Publication SJ2005-2. St. Johns River Water Management District, Palatka, Florida, 2005 ftp://secure.sjrwmd.com/technicalreports/TP/SJ2005-2.pdf.
3. Amani N, Ali N M, Mohammed A H, Samat R A. Maintenance and management of wastewater system components using the condition index system, prediction process and costs estimation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2013; 15(2):161–168.
4. Arsham H. Construction of the largest sensitivity region for general linear programs. Appl. Math. Comput. 2007; 189(2):1435–1447.
5. Arsham H, Oblak M. Perturbation analysis of general LP models: a unified approach to sensitivity, parametric, tolerance, and more-for-less analysis. Math Comput. Model. 1990; 13(8):79–102.
6. Bradley S P, Hax A C, Magnanti T L. Applied mathematical programming. Addison-Wesley Publishing Company, Boston, 1977.
7. Brown L R. 2030, Who will feed China? World watch, 1994.
8. Busłowski A. Stabilność rozwiązania optymalnego zadania programowania liniowego. Wyd. Uniwersytetu w Białymstoku, 2000.
9. Chmielowski W. Zastosowania optymalizacji w gospodarce wodnej. Wyd. Politechniki Krakowskiej, 2005.
10. Chmielowski W. Management of Complex Multi-reservoir Water Distribution Systems using Advanced Control Theoretic Tools and Techniques. Springer-Verlag, 2013.
11. Conradie B I, Hoag D L. A review of mathematical programming models of irrigation water values. Water SA 2004; 30(3):287-292.
12. Dantzig G B, Thapa M N. Linear programming. 1: Introduction. New York, Berlin, Heidelberg: Springer 1997.
13. Datta B, Harikrishna V. Optimization Applications in Water Resources Systems Engineering. Research Journal of IIT Kanpur 2005:57-64.
14. Drobny N L. Linear programming applications in water resources 1. Journal of the American Water Resources Association 1971; 7(6):1180-1193.
15. Easter K W, Rosegrant M W, Dinar A. Markets for water: Potential and performance. Springer Science & Business Media, 1998.

16. Ehrgott M, Wiecek M M. Multiobjective programming. International Series in Operations Research and Science 2005:667-722
17. European Commission. Report on the Review of the European Water Scarcity and Droughts Policy; European Commission: Brussels, Belgium, 2012.
18. Freire-González J, Decker C A, Hall J W. A Linear Programming Approach to Water Allocation during a Drought. Water 2018; 10(4):363.
19. Frizzone J A, Coelho R D, Dourado-Neto D, Soliani R. Linear programming model to optimize the water resource use in irrigation projects: An application to the Senator Nilo Coelho Project. Scientia Agricola 1997; 54(SPE):136-148.
20. Gale D, Kuhn H W, Tucker A W. Linear programming and the theory of games, Activity Analysis of Production and Allocation, Ed. TC Koopmans, 1951.
21. Gass S I, Saaty T L. Parametric objective function (part 2)-generalization. Oper. Res. 1955; 3(4):395–401.
22. Gibbons D. The Economic Value of Water. Resources for the Future. Publisher: Resources for the Future, Washington, 1986.
23. Hanafizadeh P, Ghaemi A, Tavana M. Local perturbation sensitivity analysis of linear programming with functional relation among parameters. Int. J. Oper. Res. Inf. Syst. 2011; 2(1):42–65.
24. Ignaciuk S. Post-optimization procedures for the selected transport issues in agricultural engineering. PhD thesis. Department of Applied Mathematics and Computer Science, Faculty of Production Engineering, University of Life Sciences in Lublin (in Polish), 2018.
25. Jacovkis P M, Gradowczyk H, Freisztav A M, Tabak E G. A linear programming approach to water-resources optimization. Zeitschrift für Operations Research 1989; 33(5):341-362.
26. Jansen B, de Jong J J, Roos C, Terlaky T. Sensitivity analysis in linear programming: just be careful!. European Journal of Operational Research 1997; 101(1):15-28.
27. Keshari A K, Datta B. Multiobjective Management of a Contaminated Aquifer for Agricultural Use. Water Resources Management 1996; 10(5):373-395.
28. Klabjan D, Adelman D. Existence of optimal policies for semi-Markov decision processes using duality for infinite linear programming. Siam Journal on Control and Optimization 2006; 44(6): 2104-2122.
29. Koltai T, Terlaky T. The difference between the managerial and mathematical interpretation of sensitivity analysis results in linear programming. Int. J. Production Economics 2000; 65: 257-274.
30. Koltai T, Tatay V. A practical approach to sensitivity analysis in linear programming under degeneracy for management decision making. Int. J. Production Economics 2011; 131: 392-398.
31. Kołowrocki K, Soszyńska J. Optymalizacja procesów eksploatacji złożonych systemów technicznych. Problemy Eksploatacji 2010; 31-40.
32. Kolowrocki K, Soszynska-Budny J. Reliability and Safety of Complex Technical Systems and Processes: Modeling – Identification – Prediction- Optimization. Springer Science & Business Media, 2011.
33. Kołowrocki K, Soszyńska J. Integrated safety and reliability decision support system. Journal of KONBiN 2011; 20(1): 17-28.
34. Li J, Yan M, Yu J. Evaluation on gas supply reliability of urban gas pipeline network. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20(3): 471–477, http://dx.doi.org/10.17531/ein.2018.3.17.
35. Lipinski P, Filipiak P, Rychlikowski P, Stanczyk J, Kajewska-Szkudlarek J, Lomotowski J, Konieczny T. Discovering weekly seasonality for water demand

prediction using evolutionary algorithms. In Proceedings of the Genetic and Evolutionary Computation Conference Companion 2017.07; 33-34.
36. Liu X, Chen X, Wang S. Evaluating and Predicting Shadow Prices of Water Resources in China and its Nine Major River Basins. Water Resour. Manag. 2009; 23(8):1467–1478.
37. Madadpour S, Hanafizadeh P, Habibi R. Sensitivity Analysis of Asset Allocation: In The Presence of Correlation. Journal of Accounting, Finance and Auditing Studies 2018; 4(1):151-166.
38. Malinowski J. A newly developed method for computing reliability measures in a water supply network. Operations Research and Decisions 2016; 26(4):49–64.
39. Manikowski A. Optymalizacja rejonizacji Stacji Ratownictwa Medycznego z wykorzystaniem zmodyfikowanego zagadnienia transportowego na przykładzie Warszawy. Problemy Zarządzania 2011; 9(3): 227 – 241.
40. Mavrotas G. Effective implementation of the ε-constraint method in Multi-Objective Mathematical Programming problems. Applied Mathematics and Computation 2009; 213(2): 455-465.
41. McKee P W, Clark B R, Czarnecki J B. Conjunctive-use optimization model and sustainable-yield estimation for the Sparta aquifer of southeastern Arkansas and north-central Louisiana. Water-Resources Investigations Report 2003; 2003-4231, https://doi.org/10.3133/wri034231
42. Nieuwoudt W L, Backeberg G R. A review of the modelling of water values in different use sectors in South Africa. Water SA 2011; 37(5): 703-710.
43. Połomska J, Chmielowski W. Optimization of water management system operation while quality and quantity of water determined. Annals of Warsaw University of Life Sciences-SGGW. Land Reclamation 2012; 44(1): 15-24.
44. Romaniuk M. Optimization of maintenance costs of a pipeline for a V-shaped hazard rate of malfunction intensities. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20(1):46–56.
45. Romaniuk M. Application of Markov chain and interest rate process for forecasting of costs of maintenance of pipelines. In: Simulation in Umwelt-und Geowissenschaften. Workshop Müncheberg. 2015; p. 147-156.
46. Romaniuk M. On simulation of maintenance costs for water distribution system with fuzzy parameters. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18(4):514–527, https://doi.org/10.17531/ein.2016.4.6.
47. Saaty T L, Gass S I. Parametric objective function (part 1). Oper. Res. 1954; 2(3):316–319.
48. Shahin A., Hanafizadeh P, Hladík M. Sensitivity analysis of linear programming in the presence of correlation among right-hand side parameters or objective function coefficients. Central European Journal of Operations Research 2016; 24(3): 563-593.
49. Sioshansi R, Conejo A J. Optimization in Engineering: Models and Algorithms (Vol. 120). Springer, 2017.
50. Słowiński B. Inżynieria eksploatacji maszyn. Wyd. Politechniki Koszalińskiej, 2014.
51. Sukharev M G, Karasevich A M. Reliability models for gas supply systems. Automation and Remote Control 2010; 71(7): 1415–1424, https://doi.org/10.1134/S0005117910070155.
52. Taha H A. Operations research: An introduction. Prentice Hall, 2007.
53. Wang Z. The linear programming model of water resource utilization in the irrigated area of Huaibei, Institute of Water Resource of Huai River Committee and Anhui Province, Bengbu, 2018

www.hyarc.nagoya-u.ac.jp/game/6thconf/html/abs_html/pdfs/T7MWZ30Jul04203305.pdf.
54. Ward J E, Wendell R E. Approaches to sensitivity analysis in linear programming. Ann. Oper. Res. 1990; 27(1):3–38.
55. Wawrzosek J, Ignaciuk S. Use of extended sensitivity reports of linear programming in emergency medicinal services issues. Logistyka (Poznan) 2015; No 4 attachment CD ROM no 2 part 5: 8473-8481 (in Polish).
56. Wawrzosek J, Ignaciuk S. Selected postoptimization problems of the logistics of supplying urban-industrial agglomerations with water. Logistyka (Poznan) 2016; No 1 attachment CD ROM part 2: 425-431 (in Polish).
57. Wawrzosek J, Ignaciuk S. Dual model for classic transportation problem as a tool for dynamizing management in a logistics company. Econtechmod 2016; 5(3): 95-100.
58. Wawrzosek J, Ignaciuk S. Postoptimization of the model of water supply for urban and industrial agglomeration. XLVIII Seminar of Applied Mathematics, ITM Web of Conferences 23, 00035, 1-5, https://doi.org/10.1051/itmconf/20182300035.
59. Wendell R E. A preview of a tolerance approach to sensitivity analysis in linear programming. Discrete Math. 1982; 38(1):121–124.
60. Wendell R E. Using bounds on the data in linear programming: the tolerance approach to sensitivity analysis. Math. Program. 1984; 29(3):304–322.
61. Wendell R E. The tolerance approach to sensitivity analysis in linear programming. Manag Sci. 1985; 31(5):564–578.
62. Wendell R E. Sensitivity analysis revisited and extended. Decision Sciences 1992; 23(5): 1127-1142.
63. Wondolowski F R. A generalization of Wendell's tolerance approach to sensitivity analysis in linear programming. Decis Sci 1991; 22(4):792–811.
64. World Bank. High and Dry: Climate Change, Water, and the Economy; World Bank: Washington, DC, USA, 2016.
65. http://www.maslowski.pl/index.php?id=simplex
66. Zio E. An introduction to the basics of reliability and risk analysis. World Scientific Publishing Co. Pte. Ltd, 2006.

Marek Sikora[1], Krzysztof Szczyrba[1,2], Łukasz Wróbel[1], Marcin Michalak[1]

[1]Instytut Informatyki, Politechnika Śląska

ul. Akademicka 16, 44-100 Gliwice

{marek.sikora, lukasz.wrobel, marcin.michalak}@polsl.pl

[2]Somar S.A

ul. Karoliny 4, 40-186 Katowice

k.szczyrba@somar.com.pl

Monitorowanie i utrzymanie suwnicy bramowej na podstawie bezprzewodowego systemu pomiaru i analizy poziomu drgań

***Słowa kluczowe**: czujnik drgań, system monitorowania, utrzymanie, suwnica, utrzymanie predykcyjne, eksploracja danych, analiza trendu*

Streszczenie: W artykule przedstawiono system monitorowania i diagnostyki suwnicy bramowej. Głównym zadaniem systemu jest akwizycja, wizualizacja i monitorowanie poziomu drgań newralgicznych elementów suwnicy. System wyposażony jest również w część obliczeniowo-analityczną umożliwiającą realizację zadań predykcyjnego utrzymania ruchu (ang. predictive maintenance) związanych z oceną poziomu drgań. Architektura systemu umożliwia wykorzystanie go również do innych zastosowań, w których dla realizacji zadań diagnostyki wymagana jest bezprzewodowa sieć czujników drgań.

1. Wstęp

Obecne systemy informatyczne monitorowania i diagnostyki obiektów przemysłowych oferują bogate możliwości przechowywania i przetwarzania danych. Od kilkunastu lat – zgodnie ze standardem ISA95 [11] – rysuje się coraz wyraźniejsza specjalizacja pozwalająca wyróżnić systemy: akwizycji danych pomiarowych, bieżącego monitorowania i wizualizacji procesów (HMI – Human Machine Interface; SCADA – Supervisory Control and Data Acquisition), realizacji procesów (MES – Manufacturing Execution System) oraz systemy ERP (Enterprise Resource Planning) zapewniające kompleksowe wspomaganie zarządzania przedsiębiorstwem.

Dynamicznie rozwijającym się obszarem zastosowań systemów informatycznych jest obszar analityki danych i bazujące na nim nowe funkcjonalności pozwalające dostawcom i producentom uzyskiwać przewagę konkurencyjną. Z analizą danych gromadzonych przez systemy SCADA i MES związane są dwie grupy aplikacji: CMMS (Computerised Maintenance Management Systems) i DAP (Data Analytics Platforms). Aplikacje te w architekturze ISA95 należałoby umieścić pomiędzy MES i ERP. Aplikacje DAP zapewniają narzędzia i metody do budowy modeli diagnostycznych urządzeń lub analizy gromadzonych danych w celu np. poprawy procesów produkcyjnych.

W przedsiębiorstwach zdecydowanie rośnie świadomość dotycząca konieczności zatrudniania specjalistów z obszarów analizy danych (tzw. data scientist) budujących i utrzymujących modele diagnostyczne stosowane na przykład w predykcyjnym utrzymaniu ruchu (PdM - Predictive Maintenance [18]).

W niniejszym artykule przedstawiono system monitorowania i diagnostyki suwnicy bramowej. Suwnica będąca przedmiotem diagnostyki jest elementem linii transportu surowca stalowni konwertorowej jednego z zakładów hutniczych. Ze względu na fakt, że suwnica jest jednym z kluczowych elementów linii produkcyjnej, konieczne jest monitorowanie stanu suwnicy oraz – przynajmniej przybliżone – identyfikowanie momentu w czasie, w którym użytkowanie suwnicy stanie się niebezpieczne lub niemożliwe.

Źródłem danych jest bezprzewodowa sieć czujników drgań. Dane przetwarzane są za pomocą dedykowanego systemu typu SCADA, a część analityczna bazuje na środowisku analitycznym R [21].

Niniejszy artykuł zorganizowany jest w następujący sposób. W rozdziale 2 przedstawiono przykłady prac pokrewnych związanych z systemami dedykowanymi do diagnostyki urządzeń bazującej na analizie danych historycznych, w szczególności analizie poziomu drgań. W rozdziale 3 przedstawiono obiekt podlegający monitorowaniu oraz problem jego diagnostyki. Rozdział 4 zawiera opis obejmujący sprzętową i programową część systemu monitorowania. W rozdziale 5 przedstawiono proces analizy danych i wynikającą z niego procedurę diagnostyczną, której praktyczne zastosowanie przedstawiano w rozdziale 6. Rozdział 7 stanowi podsumowanie, przedstawiono w nim również koncepcję rozwoju systemu.

2. Prace pokrewne

Systemy komputerowe dedykowane do definiowania, monitorowania i prognozowania stanów diagnostycznych urządzeń można kategoryzować z różnych punktów widzenia. Jednym z takich sposobów jest ich podział na systemy szkieletowe [13] oraz dedykowane do konkretnego rozwiązania lub grupy rozwiązań (np. [17]). Metodyki realizacji zadań diagnostyki, utrzymania maszyn i urządzeń [29], predykcyjnego utrzymania ruchu [14,26] czy diagnostyki uszkodzeń [20,25,10] pozwalają na coraz efektywniejsze wykorzystanie parku maszynowego będącego w zasobach przedsiębiorstw.

W literaturze można spotkać dużą liczbę publikacji dotyczących diagnostyki. W zależności od stopnia skomplikowania urządzenia, jego budowy, przeznaczenia i sposobu eksploatacji, diagnostykę przeprowadza się w oparciu o zróżnicowane metody badawcze. Przykładowo, diagnostyka silników spalinowych pod kątem ich zużycia bardzo często przeprowadzana jest w oparciu o analizę spektrometryczną składu oleju pobranego z tych silników [12]. Znane są także prace, w których stosuje się analityczne podejście do modelowania procesu spalania oleju w tłokach w celu opisania zależności zawartości zanieczyszczeń od upływu czasu [16]. Przykład ten ilustruje dwa główne podejścia do budowy modeli diagnostycznych, jedno bazuje na analizie zgromadzonych danych pomiarowych lub innych danych mających związek z monitorowaniem pracy urządzenia (np. obrazami termicznymi [9]), drugie na modelowaniu/symulacji działania diagnozowanego obiektu.

Ze względu na tematykę niniejszego artykułu na szczególną uwagę zasługują prace, w których wejściowym sygnałem będącym podstawą diagnostyki, są drgania generowane przez pracujące urządzenie. W pracy [6] monitorowano zachowanie pompy nurnikowej w żurawiu samochodowym poprzez akwizycję jej drgań w trzech wymiarach. Znane były widma pracy urządzenia w stanie poprawnym oraz pięciu niepoprawnych stanach diagnostycznych. Do klasyfikacji bieżącego stanu pompy zastosowano techniki

wnioskowania bayesowskiego, której parametry były strojone z wykorzystaniem algorytmu optymalizacji chmurą cząstek (ang. *particle swarm optimization*). W pracy [8] opisano zastosowanie zaawansowanej ekstrakcji cech z sygnałów drgań opisujących pracę silnika komutatorowego w dziedzinie częstotliwości.

Do celu analizy danych, których źródłem są czujniki drgań, stosuje się również transformatę falkową. Przegląd zastosowań tej metodyki do diagnostyki urządzeń przedstawiono w [19,27]. W szczególności metodę tę stosowano do wykrywania uszkodzeń wirnika silnika indukcyjnego [28], wykrywania uszkodzeń w sieciach przesyłowych [23], czy też uszkodzeń łożysk [15].

Przebiegi amplitudy drgań zazwyczaj nie są stacjonarne, z tego powodu do ich analizy stosuje się często tak zwaną kurtozę spektralną [1]. Metodę tę wykorzystano między innymi do wykrywania uszkodzeń w układach napędowych turbin wiatrowych [7], czy też oceny stanu łożysk [5].

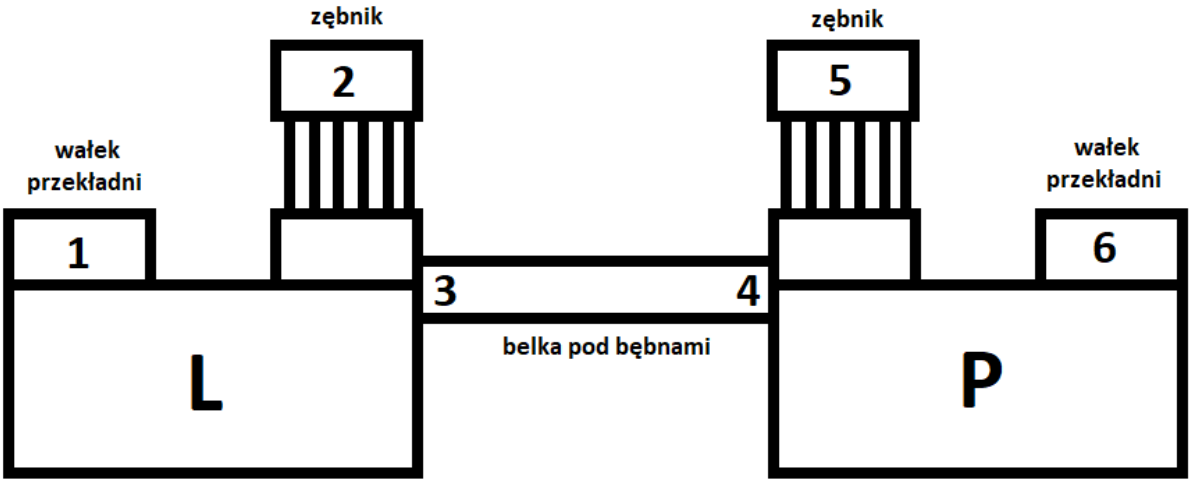
Kompleksowe ujęcie problematyki budowy modeli diagnostycznych na podstawie analizy drgań stanowi praca Samuela [22], w której przedstawiono wiele wariantów wykorzystania pomiarów drgań (począwszy od propozycji prostych wskaźników statystycznych w dziedzinie czasu i częstotliwości, poprzez transformację falkową, zastosowanie sieci neuronowych, aż po modelowanie matematyczne) do budowy modeli diagnostycznych urządzeń.

Bartelmus i inni w pracach [2], [3] podejmują z kolei temat uwzględniania podczas analizy drgań obciążenia, pod jakim w danej chwili pracuje urządzenie (np. przenośnik lub koparka kołowa). Wyodrębnienie składowej drgań wywołanych zmiennym obciążeniem urządzenia i odjęcie jej od oryginalnie zaobserwowanego szeregu drgań pozwala – w szczególnych przypadkach – na uzyskanie właściwych danych wejściowych dla procesu diagnostycznego.

3. Obiekt i cel monitorowania

Monitorowanym obiektem są dwie suwnice bramowe, każda o udźwigu 500 ton, pracujące w hali stalowni zakładu hutniczego. Suwnice są elementem linii transportu surowca w stalowni konwertorowej. Suwnica pobiera pełne kadzie z surówką z wózków transportowych, przemieszcza je w okolice gardzieli jednego z konwertorów, wlewa do nich surówkę, przechylając kadź odlewniczą, po czym odstawia puste kadzie na wózki transportowe.

Rysunek 1 przedstawia schemat układu napędowego odpowiedzialnego za podnoszenie i opuszczanie kadzi. Układ ten jest głównym miejscem, w jakim instalowane są czujniki drgań. Każda z suwnic posiada dwa napędy – lewy oznaczony na rysunku jako L oraz prawy oznaczony jako P. Napędy połączone są przez wałki napędowe z przekładniami zębatymi redukującymi prędkość obrotową bębnów głównego mechanizmu podnoszenia. Na rysunku przedstawiono również lokalizację czujników drgań, które umieszczane są na: wałkach napędowych (1, 6), łożyskach zębnika (2, 5) oraz na belce wsporczej (3, 4). Czujniki o numerach 3 i 4 pełnią w zasadzie rolę pomocniczą, ich zadaniem jest monitorowanie poziomu drgań tła, mogą one stanowić punkt odniesienia dla pozostałych czujników. Istotne dla działania systemu diagnostyki (w szczególności dla fazy przygotowania danych) jest to, że dla różnych suwnic (nawet tego samego typu) bardzo trudne jest umiejscowienie czujników drgań dokładnie w tym samym miejscu.



Rysunek 1. Schemat układu napędowego suwnicy z naniesionymi miejscami zamontowania czujników drgań.

Poprawna praca suwnic jest jednym z kluczowych elementów zachowania ciągłości produkcji stali. Awaria suwnic oznacza brak możliwości zasilania konwertorów surówką dostarczaną z wielkiego pieca, co skutkuje niemożnością przekazania surowca do kolejnych etapów produkcji (ciągłego odlewania stali i walcowni), a w skrajnych przypadkach nawet do konieczności wygaszania pieca.

Suwnica pracuje w trudnych warunkach środowiskowych, jakim jest hala produkcyjna (Rys. 2). Mechanizm przekładni pracuje stale w obecności pyłu tlenków metalicznych, dostających się do jego wnętrza. Wpływ pyłu uwidacznia się znaczącym wzrostem poziomu drgań całej suwnicy, gdyż drgania przekładni przenoszone są na inne elementy, w tym elementy konstrukcji hali przemysłowej. W skrajnej sytuacji praca suwnicy staje się niemożliwa lub operatorzy zmuszeni są do redukowania biegów silników, co niekorzystanie wpływa na wydajność produkcji. W sytuacjach takich niezbędne jest wykonanie generalnego remontu przekładni.

Rysunek 2. Fragment obudowy układu napędowego suwnicy oraz widok hali produkcyjnej z poziomu pomostu operatorskiego.

Kluczowym zadaniem diagnostycznym jest analiza poziomu drgań i określenie na tej podstawie czy w horyzoncie kilku lub kilkunastu dni możliwe będzie dalsze normalne użytkowanie suwnicy. Procedura diagnostyczna powinna ułatwiać służbom utrzymania ruchu, podejmowanie decyzji dotyczących wykonania remontu czy też doraźnych – poza zaplanowanym harmonogramem – prac utrzymaniowych (np. wymianę oleju/smaru przekładniowego, kontrolę uzębienia wałów).

Z praktycznego punktu widzenia pożądane jest aby procedura diagnostyczna była parametryzowalna, gdyż np. niemożność rozmieszczenia czujników drgań za każdym razem w tym samymi miejscu powoduje, że poziomy drgań nieakceptowalne w jednym warunkach eksploatacyjnych mogą okazać się akceptowalne w innych warunkach. W przedstawionej w rozdziale 5 analizie danych dużą rolę odegrał fakt posiadania informacji pochodzących z tzw. dziennika zdarzeń, w którym zapisywane były wszystkie prace konserwatorskie i remontowe wykonywane na monitorowanym obiekcie. Informacje te wykorzystano do opracowania i oceny efektywności przedstawionej metody. Pozwoliło to m.in. odnieść poziom drgań do wykonanych prac remontowych i, w szczególności, nie traktować nagłych spadków poziomu drgań jako błędów danych.

4. System monitorowania

Głównym zadaniem systemu monitorowania jest akwizycja, monitorowanie i analiza poziomu drgań rejestrowanych przez czujniki. Śledzenie poziomu i dynamiki zmian drgań pozwala – poza bieżącą wizualizacją – na określenie momentu osiągnięcia ostrzegawczych i alarmowych poziomów drgań.

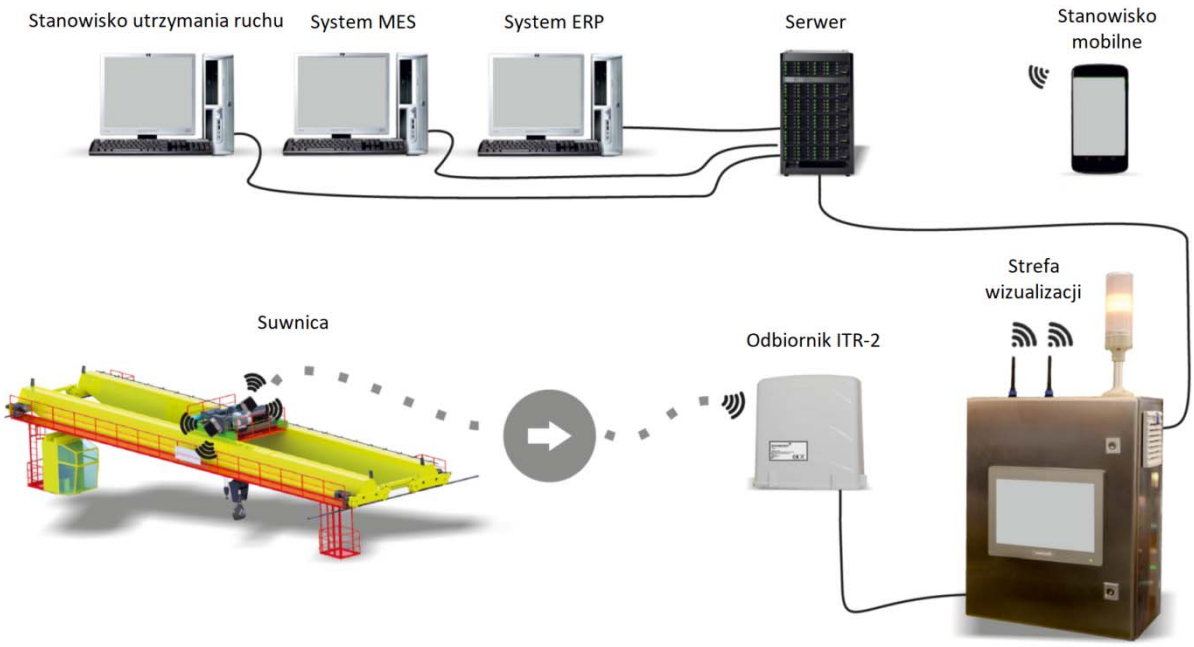
Głównym źródłem danych dla prezentowanego systemu są czujniki drgań montowane w newralgicznych punktach monitorowanych obiektów. Praktyka pokazuje, że punkty te są często zlokalizowane w miejscach trudnodostępnych. Równie często monitorowaniu podlegają ruchome części maszyn.

Warunki te zdecydowały o następujących głównych cechach omawianego systemu:

- czujniki montowane są na stałe w punktach pomiarowych,
- czujniki są bezprzewodowe (zasilane bateryjnie),
- czujniki są odporne na tzw. udary temperaturowe, gdyż okresowo mogą być narażone na kontakt z płomieniami,
- czujniki są pyłoszczelne, co pozwala je zabezpieczyć przed działaniem pyłu metalicznego,
- komunikacja z siecią czujników odbywa się drogą radiową z wykorzystaniem sieci transceiverów,
- część informatyczna systemu zrealizowana jest w architekturze klient-serwer:
    - serwer zapisuje, przechowuje i udostępnia pomiary gromadzone w bazie danych,
    - aplikacja kliencka o cechach systemu SCADA umożliwia definiowanie plansz i kokpitów menadżerskich (pozwalających na wizualizację pracy urządzenia, w tym rejestrowanego poziomu drgań) oraz generowanie raportów podsumowujących,
- integralną częścią systemu jest także serwer analityczny wykorzystujący środowisko R [21], na serwerze uruchamiane są procedury diagnostyczne związane z analizą danych, serwer dostarcza informacji o aktualnym stanie urządzenia.

Rysunek 3 przedstawia architekturę systemu dedykowanego do monitorowania suwnicy. Na komputerze *Serwer* zainstalowane jest oprogramowanie komunikacyjne (dokonujące akwizycji danych), serwer baz danych (PostgreSQL) oraz środowisko analityczne R. Komputer *Stanowisko utrzymania ruchu* reprezentuje aplikację klienta

(aplikacji takich może być wiele). Dodatkowo, system może udostępniać dane innym aplikacjom np. systemom klasy MES czy ERP.



Rysunek 3. Schemat systemu monitorowania drgań i diagnostyki maszyn z zastosowaniem bezprzewodowych czujników drgań.

W dalszej części rozdziału zostanie krótko opisany czujnik drgań wybrany do realizacji zadania monitorowania suwnicy oraz wybrane elementy aplikacji klienckiej.

4.1 Czujnik drgań

Przez drgania rozumie się ruchy oscylacyjne ciał fizycznych o określonej masie w założonym układzie odniesienia. Drgania mogą być opisane za pomocą jednego z trzech parametrów: przemieszczenia, prędkości lub przyspieszenia. Do diagnostyki wystarczającym jest pomiar jednego z tych parametrów – pozostałe można uzyskać na podstawie znanych zależności fizycznych. Najłatwiejszym sposobem pomiaru drgań dla wyższych częstotliwości jest pomiar przyspieszenia, ponieważ amplituda tej wielkości jest mierzalna w szerokim zakresie widma drgań.

Z punktu widzenia wymagań diagnostycznych czujnik powinien charakteryzować się jak najlepszą dokładnością i rozdzielczością pomiaru, szerokim pasmem przenoszenia i liniowością pomiaru w tym paśmie.

Zestawienie przykładowych czujników rozważanych do użycia jako źródła danych pomiarowych zamieszczono w Tabeli 1 (producentem pierwszego czujnika jest firma Banner Engineering, drugiego KCF Technologies, trzeciego Swift Sensors, czwartego EMAG-Servis, a ostatniego firma Somar).

Typ czujnika	QM42VT1	SD-VSN-2	201R	Pantera	WS-VT1
Odporność na trudne warunki środowiskowe (wilgotność, zapylenie, udary termiczne)	-	+	-	-	+
Możliwość pracy w atmosferach zagrożonych wybuchem gazów i pyłów	-	-	-	-	+
Pasmo komunikacyjne	2,4GHz	2,4GHz	860MHz	Brak	860MHz
Czas działania w trybie podtrzymania bateryjnego	+	+	+	-	+

Tabela 1. Zestawianie czujników umożliwiających pomiar przyspieszenia.

Z wymienionej listy najlepszym kandydatem do zastosowania – poza ostatecznie wybranym czujnikiem WS-VT1 – wydawał się czujnik typu SD-VSN-2. Jednakże ze względu na problemy z komunikacją, jakie odnotowano dla tego czujnika w hali produkcyjnej, w której pracuje suwnica podczas wdrożenia systemu zastosowano opracowany w Somar S.A czujnik WS-VT1. Jak widać w tabeli 1, czujnik ten charakteryzuje się wszystkimi wymaganymi dla rozważanego zastosowania cechami. W szczególności, jego pasmo komunikacyjne jest wystarczające do poprawnej transmisji danych w hali produkcyjnej. Dodatkowo, czujnik został zintegrowany z nadajnikiem radiowym.

WS-VT1 pozwala na pomiar przyspieszeń w trzech osiach, w zakresach ±2g/±4g/±8g/±16g, w paśmie do 2,5 kHz. Czujnik pozwala także mierzyć temperaturę otoczenia. Za pomiar temperatury odpowiedzialny jest cyfrowy przetwornik temperatury dokonujący pomiaru z rozdzielczością 12-to bitową i z dokładnością 1°C w zakresie –55÷+85°C. Oba przetworniki można wprowadzić w tryb uśpienia w celu oszczędzania energii. Pomiarami zarządza energooszczędny, szesnastobitowy mikrokontroler wyposażony w transceiver radiowy, pracujący w paśmie 860 MHz. Całość zasilana jest z baterii litowo-jonowej o rozmiarze ½ AA i pojemności 1,2 Ah.

W podstawowym trybie pracy czujnik jest cyklicznie wybudzany ze stanu uśpienia w celu dokonania pomiarów przyspieszeń, temperatury oraz poziomu naładowania baterii. Drogą radiową czujnik wysyła następujące informacje:

- swój numer indentyfikacyjny,
- wartość skuteczną i maksymalną drgań,
- wartość temperatury,
- poziom naładowania baterii.

Cykliczne usypianie i wybudzanie czujnika nie jest jedynym sposobem wykonywania pomiarów. Opracowane oprogramowanie wbudowane pozwala m.in. na zmianę częstotliwości akwizycji pomiarów w zależności do obserwowanego poziomu drgań. Rozwiązanie to pozwala na oszczędzanie energii w chwilach przestoju maszyny (brak lub niski poziom drgań), a jednocześnie na zwiększoną częstotliwość akwizycji danych podczas rejestrowania drgań powyżej pewnej predefiniowanej przez użytkownika wartości granicznej. Teoretycznie czujnik może być zaprogramowany w taki sposób, aby zwiększać częstotliwości pomiarów wraz z przekraczaniem kolejnych – coraz wyższych – poziomów drgań.

4.2 Stanowisko utrzymania ruchu

Czujniki wraz z nadajnikami stanowią źródło danych dla aplikacji monitorującej stan pracy urządzenia. Dane z czujników pobierane są z wykorzystaniem transceiverów radiowych. Transceivery pracują w sieci Ethernet w standardach 10Base-T lub 100Base-Tx.

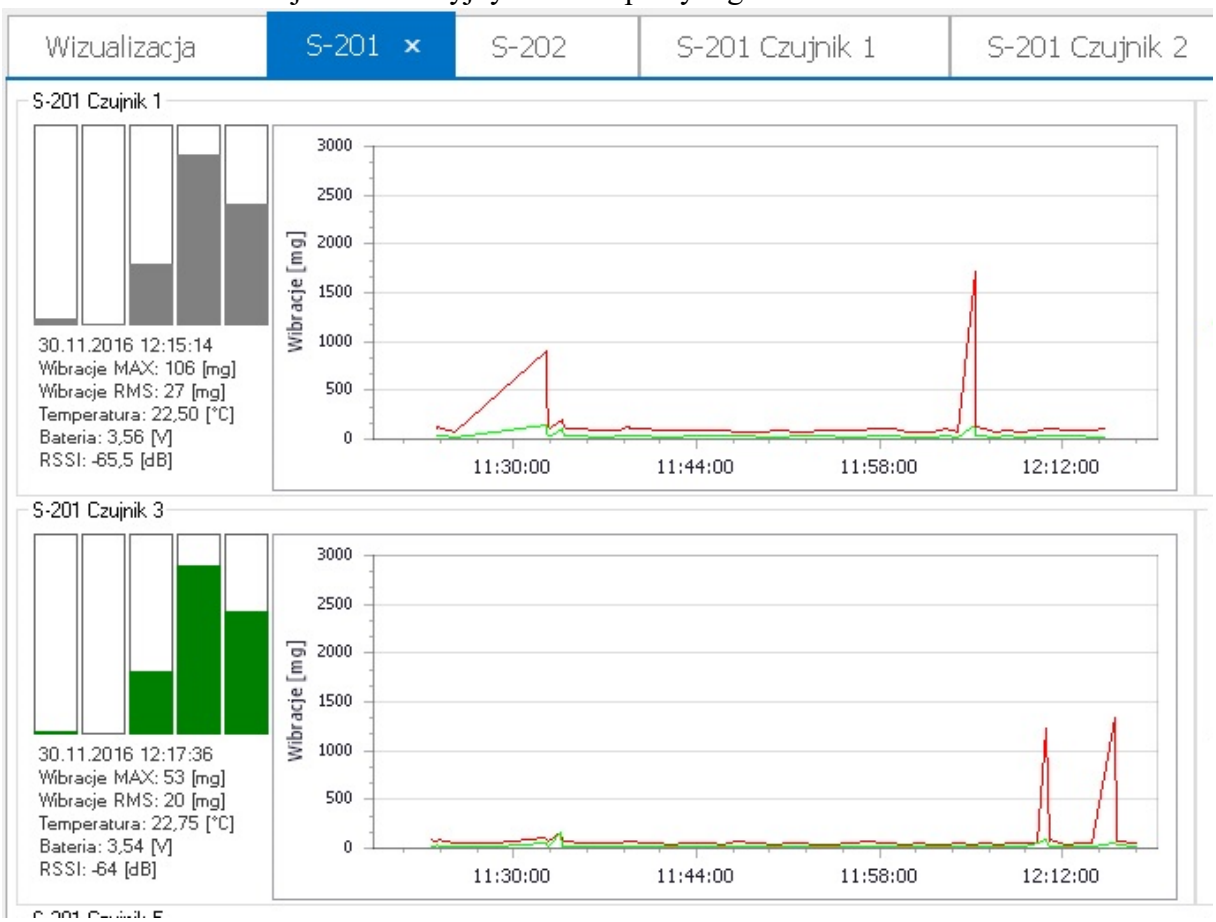
Ich podstawowym zadaniem jest odbiór danych pomiarowych z czujników będących w ich zasięgu. Dane te są uzupełniane o wskaźnik siły sygnału odbiornika i zapisywane w odpowiednich tabelach bazy danych.

Serwer bazy danych gromadzi dane pomiarowe oraz udostępnianie je aplikacji stanowiska utrzymania ruchu będącej programem typu SCADA oraz serwerowi środowiska R, w którym wykonywana jest analiza danych.

Aplikacja stanowiska utrzymania ruchu pozwala m.in. na:

- prezentację bieżących wartości drgań, temperatury i poziomu naładowania baterii,
- konfigurację i ustawianie prawidłowych, ostrzegawczych i alarmowych poziomów drgań (generalnie parametrów procedury diagnostycznej) oraz poziomów temperatury i naładowania baterii,
- prezentację danych bieżących i historycznych na wykresach (Rysunek 4),
- prezentację wybranych wyników analizy danych, wykonywanych przez analityczną część systemu,
- generowanie raportów okresowych (zmianowych, dobowych, długoterminowych).

Aplikacja umożliwia również prowadzenie dziennika awarii i prac remontowo-konserwatorskich. Jak już wspominano, jest to bardzo ważna cecha systemu pozwalająca na wykorzystanie informacji zawartych w dzienniku podczas analizy danych. W szczególności, informacja o wymianie dowolnego z monitorowanych elementów suwnicy powoduje zresetowanie informacji o bezawaryjnym czasie pracy tego elementu.



Rysunek 4. Fragment okna wizualizacji pomiarów dla grupy czujników – widoczne są bieżące pomiary (wykres słupkowy) oraz pomiary historyczne (na rysunku pomiary z ostatniej godziny).

5. Procedura diagnostyczna

5.1 Wstępne przetwarzanie danych

Dwie suwnice były monitorowane za pomocą zestawu sześciu czujników każda. Do analizy wykorzystano informację o zarejestrowanych wartościach drgań – maksymalnej (VibrMAX) oraz skutecznej (VibrRMS).

Pierwszy etap przetwarzania danych polegał na przygotowaniu dwóch zbiorów danych (po jednym dla każdej suwnicy), w których każdy kolejny wiersz zawiera informacje o maksymalnym i skutecznym poziomie drgań rejestrowanych przez każdy z 6 czujników w kolejnych sekundach (Tabela 2).

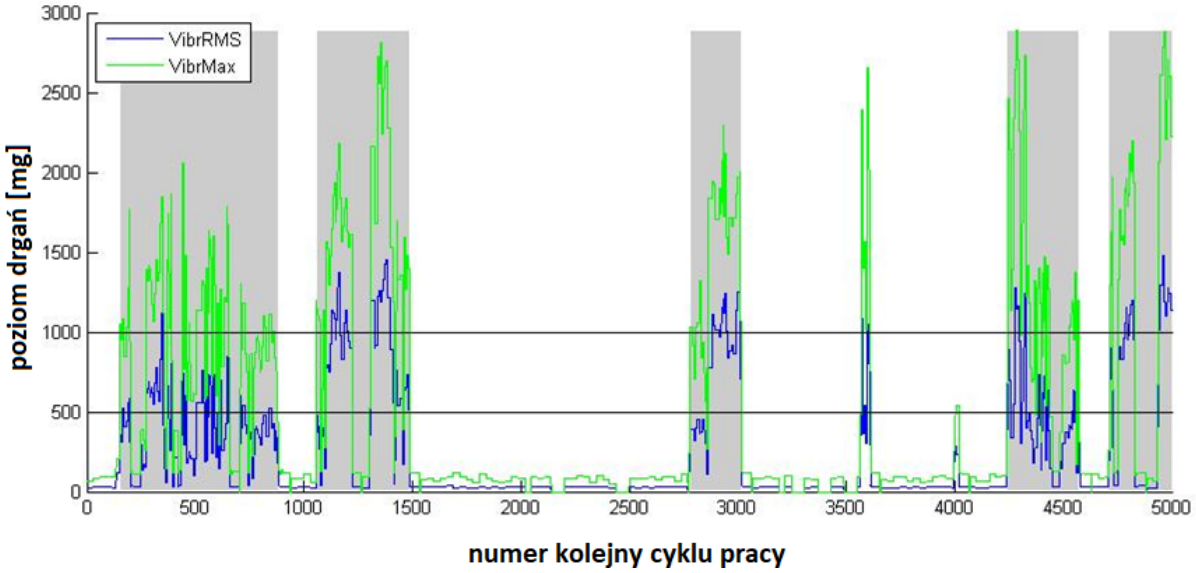
Data/czas	VibrRMS1	…	VibrRMS12	VibrMAX1	…	VibrMAX12
2017-05-05 15:12:00	28	…	29	71	…	74
2017-05-05 15:12:01	28	…	30	70	…	72

Tabela 2. Zbiór danych podstawowych opisujących pracę każdej z suwnic.

W sytuacji, w której z danego czujnika w danej sekundzie nie odnotowano pomiarów, a przerwa w danych wynosiła mniej niż 30 sekund, brakujące dane uzupełniano ostatnią z zarejestrowanych wartości. Okres 30 sekund wynikał z konfiguracji transmisji, w której wymuszano przesyłanie danych z czujników nie rzadziej niż 30 sekund. Przerwy dłuższe niż 120 sekund traktowano jako braki w danych.

Procedura diagnostyczna bazuje na analizie wektora zmiennych, który opisuje pojedynczy cykl pracy suwnicy. Cykl pracy powinien odzwierciedlać działanie suwnicy polegające na podniesieniu przenoszonego elementu przemieszczeniu go w inne miejsce i jego opuszczeniu. Ponieważ nie dysponowano informacjami o pracy tych elementów, które odpowiadają za podnoszenie i przemieszczanie się suwnicy, cykl zdefiniowano jako okres czasu, w którym czujniki rejestrowały podwyższone poziomy drgań. Ponadto, w celu identyfikacji cykli pracy przyjęto następujące dodatkowe założenia:

- podstawą identyfikacji cykli były pomiary, dla których zarejestrowana wartość skuteczna drgań przekraczała 150 mg; wartości niższe traktowane były jako przerwa w pracy suwnicy,
- zidentyfikowane cykle o długości krótszej niż 70 sekund nie były brane pod uwagę,
- jeśli przerwa pracy trwała mniej 60 sekund, dwa kolejne cykle były łączone.



Rysunek 5. Przykład identyfikacji cykli pracy suwnicy – obszary szare oznaczają zidentyfikowane cykle pracy.

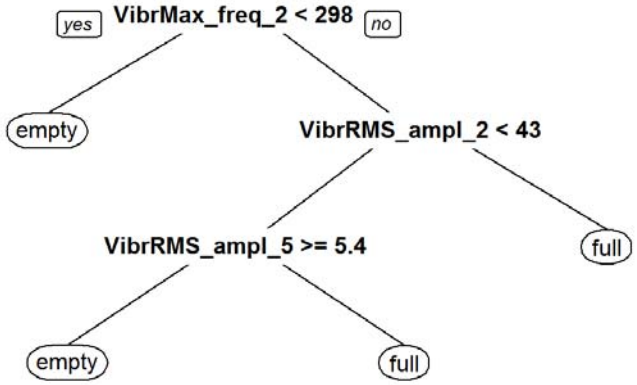
Dla każdego ze zidentyfikowanych cykli pracy obliczano szereg parametrów opisujących charakterystyki przebiegów zmiennych VibrMAX oraz VibrRMS. Obliczeń dokonano w całym zakresie rejestrowanych wartości oraz dla odfiltrowanych wartości tych cech w przedziałach 2000-2500mg, 2500-3000mg, 3000-3500mg. Dla VibrMAX i VibrRMS wyznaczano m.in.:

- wartość minimalną,
- wartość średnią i jej piąty (ozn. P5) oraz dziewięćdziesiąty piąty (ozn. P95) percentyl,
- liczbę pomiarów powyżej 3500mg,
- wyniki analizy harmonicznej sygnału wartości skutecznej drgań: częstotliwość 10-ciu pierwszych składowych harmonicznych (łącznie ze składową stałą) oraz ich amplitudę,
- wyniki analizy harmonicznej sygnału wartości maksymalnej drgań: częstotliwość 10-ciu pierwszych składowych harmonicznych (łącznie ze składową stałą) oraz ich amplitudę,
- pierwszy kwartyl (ozn. Q1),
- medianę (ozn. Q2),
- trzeci kwartyl (ozn. Q3),
- wartość maksymalną,
- rozstęp międzykwartylowy (Q3-Q1),
- rozstęp percentylowy (P95-P5),
- modę zarejestrowanych wartości.

Efektem końcowym przeprowadzonej analizy był, dla każdego czujnika, wektor 115 cech charakteryzujących każdy cykl pracy suwnicy.

Pierwsza z wykonanych analiz polegała na sprawdzeniu czy na podstawie powyższych parametrów da się jednoznacznie rozpoznawać cykle pracy, w których suwnica obciążona jest ciężarem pełnej kadzi. Do badań dysponowano bardzo ograniczonym zbiorem treningowym składającym się z 48 przejazdów suwnicy, które zostały przez obsługę

kwalifikowane jako „przejazd pełny” (24 przykłady), „przejazd pusty” (24 przykłady). Na podstawie dostępnych przykładów wygenerowano – stosując algorytm RPART [24] – drzewo decyzyjne załączone na Rysunku 6.



Rysunek 6. Drzewo decyzyjne do identyfikacji pełnych i pustych przejazdów suwnicy.

Cechy bazujące na analizie Fouriera pozwoliły na uzyskanie drzewa o zdolności klasyfikacji wynoszącej odpowiednio: dokładność klasy empty 0,792 oraz dokładność klasy full 0,958. Niestety, po uruchomieniu klasyfikatora na danych gromadzonych przez dłuższy okres czasu zaobserwowano, że większość przykładów (5467 z 6439) została zaklasyfikowana jako empty). Pomimo tego, że przykłady nie posiadały etykiet klas, taki wynik klasyfikacji uznano za zły, gdyż ze specyfiki pracy suwnicy rozkład przykładów powinien być zbliżonych do równomiernego. Ostatecznie, dalszej analizie podawano cykle pracy suwnicy bez rozgraniczania czy oznaczają one przejazd pusty czy pełny.

W dalszej części badań do określenia procedury diagnostycznej użyto danych z jednej suwnicy, dane z drugiej suwnicy wykorzystano do weryfikacji przeprowadzonych badań.

W celu identyfikacji cech mających największe znaczenie diagnostyczne, zastosowano metodę selekcji cech bazującą na rankingu cech klasyfikatora utworzonego za pomocą metody lasów losowych [4]. Wektor kolejnych cykli z okresu 4 miesięcy pracy suwnicy (w sumie 11 818 cykli) poetykietowano w ten sposób, że cyklom jakie zarejestrowano w okresie do dwóch tygodniu bezpośrednio przed awarią i po jej usunięciu nadano etykietę 0 (suwnica sprawna), a cyklom jakie zarejestrowano w okresie dwóch tygodni przed awarią aż do wystąpienia awarii nadano etykietę 1 (suwnica w stanie przedawaryjnym). Uzyskano w ten sposób zbiór, w którym 9275 przykładów wskazywało na klasę 0 i 2543 przykłady wskazywały na klasę 1. Na przygotowanym – dla każdego czujnika drgań oddzielnie – zbiorze danych trenowano klasyfikator złożony z 1000 drzew decyzyjnych. Głównym kryterium podziału i oceny węzła było kryterium Gini [4]. Uzyskano następującą jakość klasyfikacji:

- czujnik 1 – specyficzność 0,984, czułość 0,898,
- czujnik 2– specyficzność 0,991, czułość 0,722,
- czujnik 3 – specyficzność 0,986, czułość 0,828.

Na podstawie analizy drzew tworzących las losowy oceniano istotność każdej z rozważanych cech z punktu widzenia jej ważności w utworzonym klasyfikatorze. Ważność cechy X na podstawie utworzonego lasu losowego została wyznaczona według wzoru:

$$Imp(X) = \frac{1}{M}\sum_T \sum_{t \in T: v(t)=X} \frac{N_t}{N} \cdot \{G(t) - [\frac{N_{t_L}}{N_t} G(t_L) + \frac{N_{t_R}}{N_t} G(t_R)]\} \qquad (1)$$

gdzie:

- M – liczba wszystkich drzew T w lesie losowym (w tym przypadku 1000),
- t – węzeł drzewa T,
- t_L, t_R – lewy i prawy potomek węzła t,
- $v(t)$ – cecha pojawiająca się w warunku węzła t,
- N – liczba wszystkich obserwacji w zbiorze treningowym (w tym przypadku 11818),
- N_t – liczba obserwacji w węźle t,
- $G(t)$ – indeks Giniego dla węzła t; $G(t) = 1 - \frac{N_t^+}{N_t} - \frac{N_t^-}{N_t}$, gdzie N_t^+ i N_t^- to odpowiednio liczba przykładów pozytywnych (klasa 1) i negatywnych (klasa 0) w węźle t.

Im wyższa wartość miary (1), tym cecha opisująca cykl pracy suwnicy lepiej rozróżnia stany diagnostyczne, a tym samym jest bardziej pożądana jako cecha diagnostyczna. Jeżeli cecha nie występowała w żadnym z drzew tworzących las losowy, jej ważność wynosiła 0.

Rankingi cech według miary (1) zostały utworzone dla każdego z czujników z osobna. Uzyskano w ten sposób 3 rankingi cech. Ostateczna ranga danej cechy była średnią arytmetyczną pozycji, jaką cecha zajmowała w każdym z tych rankingów.

Pierwsze pięć cech w rankingu wykorzystano do dalszych prac nad opracowaniem procedury diagnostycznej. Cechami tymi były:

- VibrMax2000_Q3 – wartość 3 kwartyla cechy VibrMAX; w obliczeniach uwzględniono jedynie wartości VibrMAX>2000,
- VibrMax3000_Q3 – analogicznie jak wyżej,
- VibrMax2500_Q3 – analogicznie jak wyżej
- VibrMax_P95 – wartość 95 centyla cechy VibrMAX;

VibrMax2000_P95 – wartość 95 centyla cechy VibrMAX; w obliczeniach uwzględniono jedynie wartości VibrMAX>2000.

5.2 Analiza trendu

Podstawą procedury diagnostycznej jest analiza trendu zmian wartości cech (nazwijmy je zmiennymi diagnostycznymi) zidentyfikowanych w poprzednim podrozdziale jako kluczowe. Analiza trendu wykonywana jest jednocześnie na podstawie danych historycznych dla *h1*, *h2*, *h3* cykli pracy suwnicy. Przy czym *h1* opisuje krótki okres czasu, a *h2*=2*h1*, *h3*=3*h1*. Pozwala to na wychwycenie dynamicznych (ale być może krótkotrwałych), jak również stałych tendencji zmian w analizowanym szeregu czasowym.

W procedurze uwzględniane są również następujące parametry:

- stopień wygładzania (ozn. *sm*) – ze względu na bardzo dużą wariancję wartości zmiennych diagnostycznych, przed wykonaniem analizy trendu wartości są wygładzane średnią ruchomą uwzględniającą 100 wcześniejszych wartości (100 cykli pracy suwnicy),
- graniczny poziom wartości zmiennej diagnostycznej (ozn. *cvl*) – poziom ten jest poziomem odniesienia pozwalającym na estymację liczby cykli pracy suwnicy pozostałych do przekroczenia tego poziomu; w przypadku badania wzrostu wartości zmiennej diagnostycznej poziom ten nie musi być utożsamiany z awarią suwnicy.

Ekstrapolacja trendu pozwala na oszacowanie liczby cykli pozostałych do przekroczenia poziomu *cvl*. Zamiast liczbą cykli, można posługiwać się liczbą dni za jaką zostanie przekroczony poziom *cvl*, w tym celu przyjmuje się, że ustalana jest (arbitralnie) lub obliczana (na podstawie danych historycznych) średnia dzienna liczba cykli pracy (ozn. *awc*), która jest ekwiwalentem jednego dnia pracy suwnicy. Wartość ta może być modyfikowana w miarę gromadzenia informacji o pracy konkretnej suwnicy.

Niech S oznacza szereg czasowy kolejnych wartości zmiennej diagnostycznej w dziedzinie cykli pracy. Niech S_{sm} oznacza szereg czasowy wartości S wygładzony za pomocą średniej ruchomej z ostatnich *sm* wartości S. Analiza trendu liniowego w szeregu S_{sm}, którym do estymacji parametrów regresji wykorzystano hi ($i \in \{1, 2, 3\}$) ostatnich wartości szeregu S_{sm} prowadzi do uzyskania modelu postaci:

$$\hat{S}_{sm,h}(i) = Ai + B \tag{2}$$

gdzie: i to numer cyklu pracy suwnicy, A i B to parametry modelu estymowane na podstawie h ostatnich wartości szeregu S_{sm}.

Oznaczając przez $\hat{S}_{sm,h}(i)$ aktualną wartość modelu (2), liczbę cykli pracy pozostałych do przekroczenia wartości progowej *cvl* obliczamy według wzoru:

$$\Delta i = \begin{cases} \frac{cvl - \hat{S}_{sm,h}(i) - B}{A} & A \neq 0 \\ \infty & A = 0 \end{cases} \tag{3}$$

Przyjmując, że średnia dzienna liczba cykli pracy suwnicy wynosi *awc*, możliwe jest określenie liczby dni pracy suwnicy, po której wartość rozważanej zmiennej diagnostycznej przekroczy poziom *cvl*. Wartości wskaźnika (3) można pogrupować tak, aby ich interpretacja była łatwiej zrozumiała dla końcowego użytkownika. Uzyskamy w ten sposób meta-wskaźnik oznaczany jako Δi_+, którego wartości interpretowane są w sposób następujący:

- $\Delta i_+ = 0$ wtedy i tylko wtedy, gdy $\Delta i < 0$ lub $\Delta i = \infty$ – odpowiada to sytuacji, gdy mamy do czynienia z trendem malejącym lub bocznym,
- $\Delta i_+ = 1$ wtedy i tylko wtedy, gdy $\Delta i \in [0,1]$ – odpowiada to sytuacji, w której poziom *cvl* został przekroczony lub zostanie przekroczony za mniej niż *awc* cykli pracy suwnicy,
- $\Delta i_+ = (\frac{1}{awc})\Delta i$ w pozostałych przypadkach.

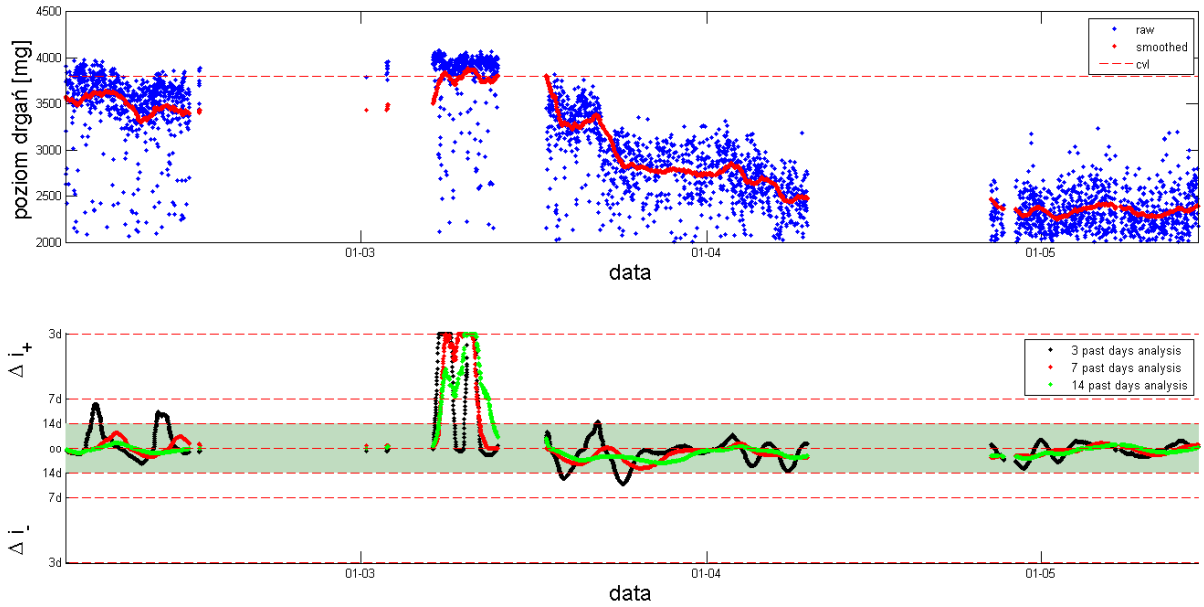
Ze względu na sposób zdefiniowania Δi_+ wskaźnik ten nie odzwierciedla spadku wartość zmiennej diagnostycznej. Dokładniej, w przypadku trendu bocznego lub malejącego wartość $\Delta i_+ = 0$, w szczególności będzie tak w sytuacji, gdy wartości badanej zmiennej ustabilizują się na jakimś poziomie (również na poziomie >*cvl*).

Meta-wskaźnik Δi_- pozwala wskazać przedziały czasu, w których mamy do czynienia z malejącym trendem monitorowanej wartości. Wartość progową dla tego wskaźnika ustalono na 0. Inne rozwiązanie to ustalenie tej wartości jako średniego poziomu drgań w okresie, w którym mamy pewność, że suwnica była sprawna. Interpretacja Δi_- – po modyfikacji wzoru (3) polegającej na zamianie wartości *cvl* wartością *nvl* – jest następująca:

- $\Delta i_- = 0$ wtedy i tylko wtedy, gdy $\Delta i < 0$ lub $\Delta i = \infty$ – odpowiada to sytuacji, gdy mamy do czynienia z trendem rosnącym lub bocznym,

- $\Delta i_- = 1$ wtedy i tylko wtedy, gdy $\Delta i \in [0,1]$ – odpowiada to sytuacji, w której poziom 0 został osiągnięty lub zostanie osiągnięty za mniej niż *awc* cykli pracy suwnicy,
- $\Delta i_- = (\frac{1}{awc})\Delta i$ w pozostałych przypadkach.

Wartości Δi_+ oraz Δi_- można wizualizować na wykresie. Wracając do analizowanego zbioru danych, w pierwszej połowie lutego obserwujemy dwa momenty znaczącego pogarszania się stanu diagnostycznego urządzenia (analizę przeprowadzono dla *h1*=7 dni, *awc*=40, *cvl*=4000). Analizując wykres Δi_+, pierwszy z nich osiąga maksimum wskazujące na niecałe 7 dni pozostałych do przekroczenia dopuszczalnego poziomu wartości zmiennej diagnostycznej, a w przypadku drugim jest to około 10 dni. Widać to szczególnie na wykresach trendów budowanych na podstawie analizy ostatnich 3 (kolor czarny) i 7 (kolor czerwony) dni pracy suwnicy. Z kolei wykres Δi_- nie pokazuje znaczącej poprawy stanu diagnostycznego pomiędzy tymi okresami, chociaż analiza wygładzonego przebiegu zmiennej diagnostycznej (rys. 7, wykres górny) uwidacznia drobny chwilowy spadek wartości zmiennej diagnostycznej.



Rysunek 7. Wykres monitorowania wartości VibrMax2000_Q3 – suwnica 1. Wykres górny: wartości oryginalne (niebieski), wartości wygładzone (czerwony) oraz poziom *cvl*. Wykres dolny: wartości wskaźnika Δi generowane na podstawie analizy trendu wykonanej w oparciu o 3, 7 i 14 poprzednich dni pracy suwnicy.

Na wykresie dolnym Rysunku 7 – ze względu na jego czytelność – oś Y obcięto do wartości odpowiadającej 3 dniom do przekroczenia *cvl*. Ponadto, jako wartość *nvl* przyjęto wartość 0, co obniża granice progów ostrzegawczych i alarmowych, a tym samym zawyża wartości umieszczone w raporcie.

W rozważanym okresie, w dzienniku napraw nie odnotowano żadnych informacji na temat podjętych działań naprawczych – zapisano jednak, że operator odnotowuje podwyższony poziom drgań. Przerwa w danych pomiarowych wynika z przerwy pracy suwnicy, następnie około 10 marca suwnica ponownie pracowała, ale jak widać w zasadzie do początku jej użytkowania wskaźnik Δi_+ osiągnął bardzo szybko wartość graniczną, co potwierdza podążający za zwiększającymi się wartościami wskaźnika Δi_+ zwiększający się poziom drgań. Następnie suwnica przeszła kilkudniowy remont i ponownie została oddana do użytkowania. Fakt wykonania remontu znajduje odzwierciedlenia w spadku wartości Δi_+

i wzroście wartości Δi_{-}, a następnie stabilizacji trendów 7 i 14-dniowych wartości wskaźnika Δi_{+}.

Interpretację zmian wartości wskaźników Δi_{+} oraz Δi_{-} oraz wartości zmiennej diagnostycznej na podstawie, której wskaźniki te są obliczane można umieścić w raporcie, którego przykład zamieszczono w Tabeli 3.

Wartości umieszczane w raporcie wyznaczane są według następujących zasad:

- OK! jeśli wskaźnik Δi_{+} jest równy 0 oraz wartości zmiennej diagnostycznej są poniżej poziomu *cvl*0,7; dodatkowo jeśli trend jest rosnący, to podawana jest liczba „dni" za jakie zostanie przekroczony poziom *cvl* o ile liczba ta jest mniejsza od 35;
- W jeśli Δi_{+} jest równy 0 oraz wartości zmiennej diagnostycznej w przedziale (*cvl*0,7, *cvl*0,8]; dodatkowo jeśli trend jest rosnący, to podawana jest liczba „dni" za jakie zostanie przekroczony poziom *cvl* o ile liczba ta jest mniejsza od 28; jeśli liczba ta jest większa od 28, podawana jest jedynie informacja o kierunku trendu,
- A jeśli wartości zmiennej diagnostycznej są powyżej poziomu *cvl*0,8; dodatkowo jeśli trend jest rosnący, to podawana jest liczba „dni" za jakie zostanie przekroczony poziom *cvl* o ile liczba ta jest mniejsza od 21; jeśli liczba ta jest większa od 21, podawana jest jedynie informacja o kierunku trendu.

Dla przypomnienia, 1 „dzień" to średnia liczba uruchomień suwnicy w trakcie dnia, wartość ta może być ustalona arbitralnie albo może zostać obliczona na podstawie historii cykli pracy suwnicy.

Suwnica SW1 trendy zmian wartości Δi_{+}				
Zmienna diagnostyczna Vibr_Max2000_Q3	Analiza trendu na podstawie ostatnich:			
Data	3 dni	7 dni	14 dni	Zgodnych
2017-02-04	A (8)	A(↑)	A (↑)	3
2017-02-05	A (10)	A(↑)	A(↑)	3
…	…	…	…	
2017-02-08	A(↓)	A(↑)	A(↑)	2
…	…	…	…	…
2017-03-24	W(↓)	W(↓)	W(↓)	3
…	…	…	…	
2017-05-09	OK!	OK!	OK!	3

Tabela 3. Raport diagnostyczny trendów wartości zmiennej diagnostycznej opisującej poziom drgań w poszczególnych cyklach pracy suwnicy (suwnica 1).

Ostania kolumna raportu zawiera liczbę jednakowych decyzji dla analiz trendu przeprowadzonych na podstawie ostatnich 3, 7 i 14 „dni". Kolumna ta pozwala na śledzenie stabilności zmian wartości zmiennej diagnostycznej.

W miarę potrzeb użytkownika raport może dotyczyć jednego lub większej liczby czujników, a także większej liczby zmiennych diagnostycznych.

Raport nie podaje w sposób jednoznaczny sugestii dotyczących czasu, w jakim wystąpi awaria lub konieczność remontu suwnicy. Pozwala natomiast w uporządkowany sposób obserwować trendy zmian cechy (lub cech) zidentyfikowanych jako zmienne diagnostyczne.

5.3 Faza wdrożeniowa

Praktyczne wykorzystanie przedstawionej powyżej metody wymaga:

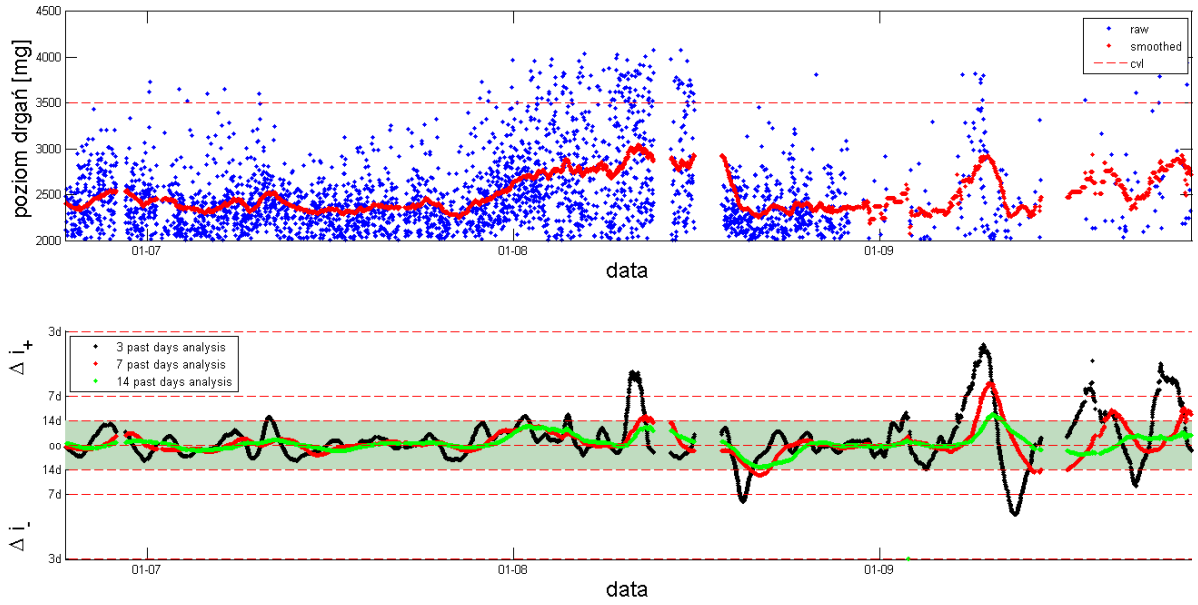
- rozmieszczenia czujników na monitorowanej suwnicy,
- identyfikacji cykli pracy suwnicy zgodnie z metodą przedstawioną na początku sekcji 5,2,
- obliczenia dla każdego z czujników wartość cechy VibrMax2000_Q3 i ew. kolejnych cech z rankingu zamieszczonego w sekcji 5,2,
- ustaleniu wartości *cvl*, *awc*,
- cyklicznym – raz na dobę – obliczaniu wartości wskaźników Δi_+, Δi_-; podczas obliczania wartości tych wskaźników wartości parametrów *sm* i *h1* ustawione są odpowiednio na 100 i 7*awc*
- generowaniu wykresów wartości Δi_+, Δi_- oraz raportu przedstawionego w Tabeli 3.

Z praktycznego punktu widzenia należy również doprecyzować warunki postępowania w przypadku: wykonania prac remontowych (czy historia wartości zmiennej diagnostycznej ma zostać wyzerowana?) oraz pracy suwnicy na tzw. biegu rozruchowym. Zdarzają się często sytuacje, w których suwnica pomimo tego, że nie powinna już być użytkowana, jest użytkowana w trybie biegu rozruchowego – w takich sytuacjach poziom drgań jest znacząco niższy.

6. Badanie weryfikacyjne

Badania weryfikacyjne przeprowadzono na drugiej identycznej suwnicy, pracującej w tym samym zakładzie hutniczym. Z dziennika remontów wynikało, że suwnica ta przechodziła remont w okresie 16-18.08.2017. Przy czym remont ten został wykonany nieprawidłowo i suwnica w krótkim czasie musiała przechodzić kolejny remont polegający na wymianie układu zębnika (remont ten wykonano w październiku). W okresie od początku września do terminu drugiego remontu suwnica pracowała zdecydowanie rzadziej niż w sytuacji jej pełnej sprawności.

Na rysunku 8 przedstawiono oryginalne i wygładzone wartości zmiennej diagnostycznej VibrMAx_Q3 oraz wartości wskaźników Δi_+ i Δi_-. W tabeli 4 przedstawiono raport dla kilku początkowych dni sierpnia. Warto dodać, że dwa tygodnie przed remontem suwnicy pracowała ona jedynie na biegu rozruchowym – co osłabiło nieco dynamikę wzrostu poziomu drgań. Jest to widoczne na wykresie wartości zmiennej diagnostycznej VibrMax2000_Q3.

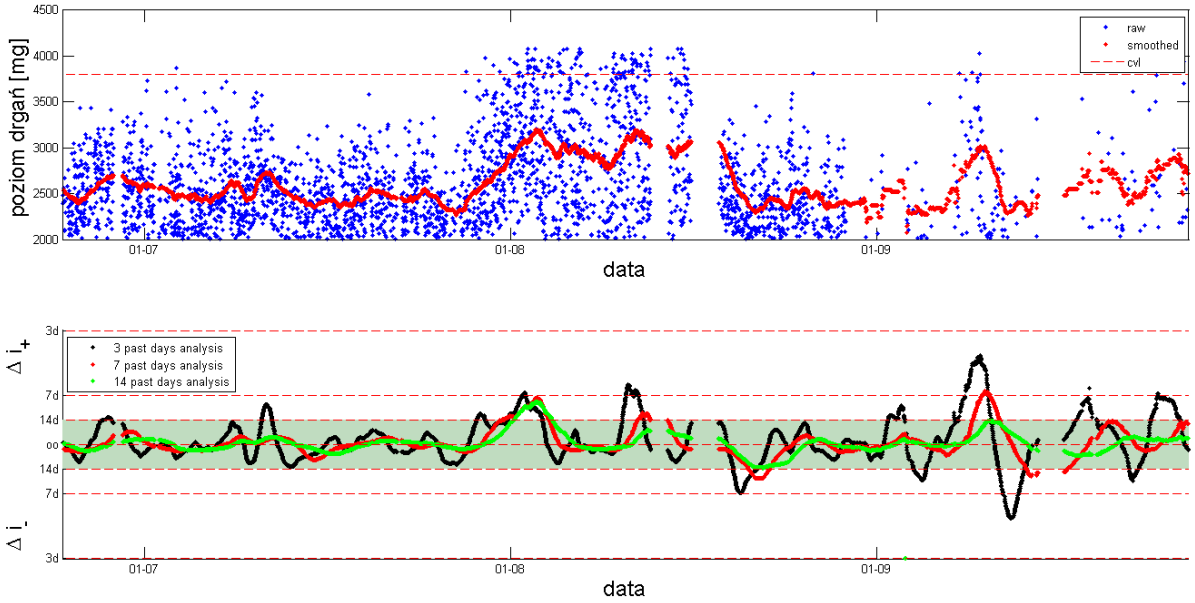


Rysunek 8. Wykres monitorowania wartości VibrMax2000_Q3 – suwnica 2. Wykres górny: wartości oryginalne (niebieski), wartości wygładzone (czerwony) oraz poziom *cvl*. Wykres dolny: wartości wskaźnika Δ*i* generowane na podstawie analizy trendu wykonanej w oparciu o 3, 7 i 14 poprzednich dni pracy suwnicy.

Suwnica SW2 trendy zmian wartości Δi_+				
Zmienna diagnostyczna Vibr_Max2000_Q3	Analiza trendu na podstawie ostatnich:			
Data	3 dni	7 dni	14 dni	Zgodnych
2017-07-25	OK!	OK!	OK!	3
…	…	…	…	…
2017-08-07	W (↓)	W(↑)	W(↑)	2
…	…	…	…	
2017-08-11	A(5)	A(20)	A(↑)	3

Tabela 4. Raport diagnostyczny trendów wartości zmiennej diagnostycznej opisującej poziom drgań w poszczególnych cyklach pracy suwnicy (suwnica 2).

Na rysunku 9 przedstawiono również wykres zmiennej diagnostycznej VibrMax2000_P95, jest to ostatnia ze zmiennych istotnych zidentyfikowanych w podrozdziale 5.2.



Rysunek 9. Wykres monitorowania wartości VibrMax2000_P95 – suwnica 2. Wykres górny: wartości oryginalne (niebieski), wartości wygładzone (czerwony) oraz poziom *cvl*. Wykres dolny: wartości wskaźnika Δ*i* generowane na podstawie analizy trendów wykonanej w oparciu o 3, 7, 14 poprzednich dni pracy suwnicy.

Zmienna ta bardzo dobrze obrazuje proces pogarszania się stanu suwnicy, gdyż wyraźne trendy wzrostowe obserwowane są w początkowym okresie sierpnia – wtedy, gdy suwnica pracowała jeszcze w normalnych warunkach (na biegu roboczym zamiast rozruchowego). Następnie – pomimo obniżenia biegu pracy – prawidłowo prognozowany jest moment, w którym konieczne było wykonanie remontu. Po nieprawidłowo wykonanym remoncie również sygnalizowane są nieprawidłowości w pracy urządzenia.

Przykład ten obrazuje jak istotne jest generowanie wykresów i raportów diagnostycznych dla wszystkich zidentyfikowanych w rozdziale 5.2 pięciu zmiennych diagnostycznych, dla wszystkich monitorowanych czujników. Z praktycznego punktu widzenia analiza i interpretacja tak dużej liczby wykresów i raportów jest – w wykonaniu dyspozytora – niemożliwa, dlatego rekomendacje dotyczące tego, którą zmienną diagnostyczną i jakie poziomy drgań *cvl* i *nvl* mają być podstawą do generowania raportu powinne być wypracowywane przez analityka danych. Praca taka jest coraz częściej realizowana jako usługa utrzymania systemu wspomagania decyzji przez działy data science firm dostarczających tego typu systemy.

7. Podsumowanie i wnioski

Monitorowanie i diagnostyka urządzeń technicznych jest zadaniem niełatwym, wymagającym każdorazowo indywidualnego podejścia do analizowanego obiektu.

W artykule przedstawiono system monitorowania drgań wykorzystujący sieć bezprzewodowych czujników. Przedstawiono także zastosowanie pomiarów gromadzonych przez czujniki do monitorowania stanu diagnostycznego suwnic bramowych.

W szczególności, w drugiej części artykułu przedstawiono metodę pozwalającą na analizę trendów zmian poziomu drgań i powiązanie tego poziomu z pogarszającym się stanem suwnicy. Przedstawiono także wymagania dotyczące praktycznego stosowania metody. Ze względu na brak możliwości symulowania różnego rodzaju uszkodzeń i awarii, metodyka diagnostyki jest stosunkowo prosta, bazuje ona na założeniu, że o pogarszającym

się stanie suwnicy świadczy zwiększający się poziom drgań (co jest zgodne z intuicją operatorów tych urządzeń). Kluczowymi problemami dla procedury diagnostycznej są: proces identyfikacji cykli pracy, identyfikacja statystyki najlepiej odzwierciedlającej wpływ poziomu drgań na stan urządzenia oraz analiza trendów (krótko i średnioterminowych) zmian wartości tej statystyki. W artykule przedstawiono sposoby rozwiązania tych problemów, zaproponowano także budowę raport diagnostycznego prezentowanego dyspozytorowi utrzymania ruchu.

Zaprezentowany system pomiarowy wdrożono w zakładzie hutniczym na terenie Górnego Śląska.

Dalsze prace polegać będą na zastosowaniu systemu do monitorowania i diagnostyki innych urządzeń – planowane jest monitorowanie drgań krążników, po których przesuwa się taśma przenośnika transportowego. W celu wdrożenia konieczne będzie opracowanie nowej procedury diagnostycznej dedykowanej do tego typu zagadnienia, pozostałe część systemu (hardware i software – część monitorująca) są gotowe do zastosowania na dowolnym obiekcie.

Opracowany system jest w znacznym stopniu zgodny z architekturą i zasadami działania tzw. szkieletowych – bazujących na danych (ang. data driven) – systemów wspomagania decyzji. W systemach tych modele diagnostyczne tworzone są dla konkretnych potrzeb na podstawie analizy danych historycznych. Natomiast moduły akwizycji, przechowywania i wizualizacji danych są modułami gotowymi do użycia.

Finansowanie

Praca została częściowo wykonana w ramach prac statutowych Instytutu Informatyki BK-213/RAU2/2018. Udział drugiego autora sfinansowano ze środków Ministerstwa Nauki I Szkolnictwa Wyższego (umowa numer 10/DW/201701/1).

Literatura

1. Antoni J. The spectral kurtosis: a useful tool for characterising non-stationary signals. Mechanical Systems and Signal Processing 2006; 20(2):282-307.
2. Bartelmus W, Zimroz R. Vibration condition monitoring of planetary gearbox under varying external load. Mechanical Systems and Signal Processing 2008; 23:246-257.
3. Bartelmus W, Zimroz R. A new feature for monitoring the condition of gearboxes in non-stationary operating conditions. Mechanical Systems and Signal Processing 2009; 23(5):1528-1534.
4. Breiman L. Random forests. Machine Learning 2001; 45(1):5-32.
5. Chen B, Yin P, Gao Y, Peng F. Use of the correlated EEMD and time-spectral kurtosis for bearing defect detection under large speed variation. Mechanism and Machine Theory 2018; 129:162-174.
6. Du W, Li A, Ye P, Liu C. Fault diagnosis of plunger pump in truck crane based on relevance vector machine with particle swarm optimization algorithm. Shock and Vibration 2013; 20(4):781-792.
7. Elforjani M, Bechhoefer E. Analysis of extremely modulated faulty wind turbine data using spectral kurtosis and signal intensity estimator. Renewable Energy 2018; 127:258:268.
8. Głowacz A, Głowacz W. Vibration-Based Fault Diagnosis of Commutator Motor. Shock and Vibration 2018; art. id 7460419.
9. Głowacz A, Głowacz Z. Diagnosis of the three-phase induction motor using thermal imaging. Infrared Physics & Technology 2017; 81: 7-16.

10. Henao H, Capolino G, Manes F. Trends in fault diagnosis for electrical machines: A review of diagnostic techniques. IEEE industrial electronics magazine 2014; 8(2): 31-42.
11. ISA95 – Enterpise-Control System Integration Standard (https://www.isa.org/isa95/)
12. Jingwei G, Niaoqin H, Lehua J, Jianyi F. A New Condition Monitoring and Fault Diagnosis Method of Engine Based on Spectrometric Oil Analysis. Advances in Intelligent and Soft Computing 2011, 110:117-124.
13. Korbicz J, Kościelny M (eds.). Modeling, Diagnostics and Process Control. Implementation in the DiaSter System. Springer-Verlag Berlin, Heidelberg 2011.
14. Korbicz J, Kościelny M, Kowalczuk Z, Cholewa W (eds.). Fault Diagnosis. Models, Artificial Intelligence, Appications. Springer-Verlag Berlin Heidelberg 2004.
15. Li Y, Liang X, Xu M, Huang W. Early fault feature extraction of rolling bearing based on ICD and tunable Q-factor wavelet transform. Mechanical Systems and Signal Processing 2017, 86(Part A):204-223.
16. Macián V, Tormos B, Olmeda P, Montoro L. Analytical approach to wear rate determination for internal combustion engine condition monitoring based on oil analysis. Tribology International 2003; 36:771–776.
17. Mazurkiewicz, D. Computer-aided maintenance and reliability management systems for conveyor belts. Eksploatacja i Niezawodność - Maintenance and Reliability 2014; 16(3):377-382.
18. Mobley R. An Introduction to Predictive Maintenance. Second Edition. Butterworth-Heinemann 2013.
19. Peng Z, Chu F. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. Mechanical Systems and Signal Processing 2004; 18(2):199-221
20. Przystałka P, Moczulski W. Methodology of neural modelling in fault detection with the use of chaos engineering. Engineering Applications of Artificial Intelligence 2015; 41:25-40.
21. R Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014, http://www.R-project.org
22. Samuel P, Pines D. A review of vibration-based techniques for helicopter transmission diagnostics. Journal of Sound and Vibration 2005; 282(1-2):475-508.
23. Silva S, Costa P, Gouvea M, Lacerda A, Alves F, Leite D. High impedance fault detection in power distribution systems using wavelet transform and evolving neural network. Electric Power Systems Research 2018, 154:474-483.
24. Therneau T, Atkinson B. Package: rpart (http://cran.r-project.org/web/packages/rpart/rpart.pdf)
25. Wachla D, Moczulski W. Identification of dynamic diagnostic models with the use of methodology of knowledge discovery in databases. Engineering Applications of Artificial Intelligence 2007; 20(5):699-707.
26. Wu S, Zuo M. Linear and Nonlinear Preventive Maintenance Models. IEEE Transactions on Reliability 2010; 59(1):242-249.
27. Yan R, Gaob R, Chen X. Wavelets for fault diagnosis of rotary machines: A review with applications. Signal Processing 2014; 96(A):1-15.
28. Ye Z, Wu B, Zargari N.: Online mechanical fault diagnostics of induction motor by wavelet artificial neural network using stator current. IECON Proceedings 2000; 2:1183–1188.
29. Zio E. Some challenges and opportunities in reliability engineering. IEEE Transactions on Reliability 2016; 65(4):1769-1782.

Bolesław Karwat, Ryszard Machnik, Jerzy Niedźwiedzki, Magdalena Nogaj, Piotr Rubacha, Emil Stańczyk[1]

AGH Akademia Gómiczo-Hutnicza im. Stanisława Staszica w Krakowie, al. Mickiewicza 30, 30-059 Kraków.
[1]ESOS Projekt, Cieśle 109, 29-105 Krasocin.

e-mail: karwat@agh.edu.pl, machnik@agh.edu.pl, jniedz@agh.edu.pl, nogaj@agh.edu.pl, rubacha.piotr@gmail.com, stanczykemil@gmail.com

Kalibracja modelu materiału sypkiego w Metodzie Elementów Dyskretnych na przykładzie perlitu D18-DN

Calibration of bulk material model in Discrete Element Method on example of perlite D18-DN

Słowa kluczowe: *metoda elementów dyskretnych, modelowanie, symulacje komputerowe, kalibracja, materiały sypkie, perlit.*

Keywords: *discrete element method, modelling, computer simulations, calibration, bulk materials, perlite.*

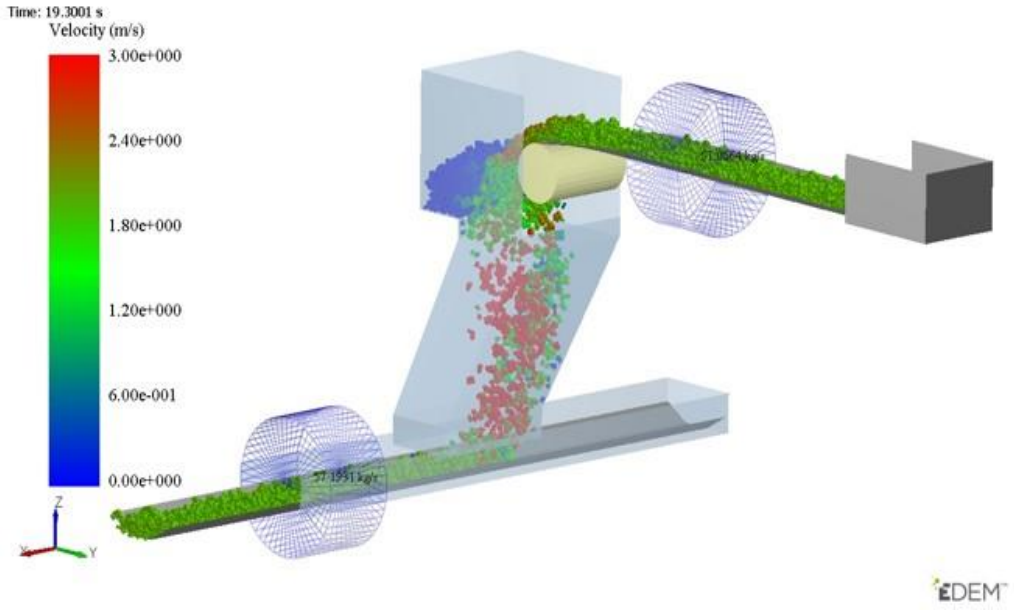
Streszczenie: Analityczne metody obliczeniowe parametrów konstrukcyjnych maszyn i urządzeń transportowych są często niewystarczające, zawłaszcza w przypadku transportu nietypowych materiałów sypkich. Pomocnym narzędziem numerycznym wspierającym proces projektowania i optymalizacji urządzeń do transportu materiałów sypkich jest Metoda Elementów Dyskretnych (DEM). Uzyskanie wiarygodnych wyników symulacji wymaga kalibracji parametrów wejściowych modelowanego materiału wykorzystując wyniki badań laboratoryjnych właściwości fizykochemicznych rzeczywistych materiałów. W artykule przedstawiono metodologię kalibracji modelu DEM na przykładzie perlitu D18-DN. W oparciu o przeprowadzoną kalibrację zaprezentowano możliwości zastosowania metody DEM do symulowania transportu materiału przenośnikiem ślimakowym.

Abstract: Analytical methods for calculations of the transport machinery are often insufficient especially when untypical granular materials are considered. Discrete Element Method (DEM) is a very useful numerical tool supporting designing and optimization of the transport equipment. However, to obtain confidence DEM simulation results an accurate set of input parameters values is needed. The most common calibration approach is to make use of a procedure where laboratory tests are performed and then the same experiments are numerically replicated in DEM. The article presents calibration of the DEM input parameters on the example of perlite D18-DN. Based on the calibrated material model simulation of perlite transport in screw conveyor has been shown.

1. Wstęp

Optymalizacja konstrukcji maszyn i urządzeń do transportu materiałów sypkich ma na celu obniżenie kosztów ich produkcji przy zachowaniu jakości i funkcjonalności. Proces ten polega na analizie funkcji cech użytkowych urządzenia oraz poszukiwaniu możliwości jego tańszego wytwarzania lub poprawy funkcjonalności poprzez usprawnienia konstrukcyjne, oszczędność lub substytucję materiałów konstrukcyjnych. Zmniejszenie masy urządzenia, ograniczenie zużycia energii oraz wzrost wydajności to jedne z głównych kryteriów optymalizacji przyjmowanych przez biura projektowe oraz firmy produkujące przemysłowe urządzenia transportowe. Analityczne podejście do projektowania maszyn i urządzeń transportowych pozwala z zadowalającą dokładnością dobrać wymiary charakterystyczne komponentów, przewidzieć wydajność i zapotrzebowanie mocy, a także pozwala oszacować parametry wytrzymałościowe i trwałość urządzenia. Jednak podejście to nie dostarcza

informacji o zachowaniu się materiału o określonych własnościach fizykalnych w czasie transportu oraz jego interakcji z podzespołami urządzenia. Wiedza ta jest niezbędna, aby móc przystąpić do procesu optymalizacji konstrukcji. W tej kwestii pomocne okazują się programy komputerowego wspomagania projektowania. Coraz większe moce obliczeniowe komputerów pozwalają na stosowanie metod numerycznych do rozwiązywania skomplikowanych problemów inżynierskich. W zakresie symulacji materiałów sypkich szeroko wykorzystuje się Metodę Elementów Dyskretnych DEM (ang. Discrete Element Method) [2-3,6,9,12]. Pozwala ona na symulowanie zachowania materiałów ziarnistych o dowolnych własnościach fizycznych, poddanych różnym wymuszeniom wynikającym z warunków pracy rozpatrywanego urządzenia. Przykład zastosowania DEM w symulacji przenośnika taśmowego zaprezentowano na rysunku 1.



Rys. 1. Przykład zastosowania metody DEM w symulacjach przenośnika taśmowego

Uwzględnienie w modelu symulacyjnym własności fizykochemicznych i mechanicznych transportowanego materiału drobnoziarnistego wymaga skalibrowania parametrów wejściowych modelu (między innymi określenie: współczynnika tarcia, gęstości nasypowej, kształtu i rozmiaru ziaren) w oparciu o wyniki badań laboratoryjnych [4-5,10-11,13]. W niniejszej pracy przedstawiono metodykę badań materiałów sypkich pod kątem ich wykorzystania do kalibracji modelu tworzonego w środowisku DEM oraz sam proces kalibracji na przykładzie materiału ziarnistego perlit D18-DN.

2. Badania właściwości fizykochemicznych materiału: perlitu D18-DN

Model symulacyjny procesu transportu materiału wymaga wprowadzenia parametrów opisujących własności fizykochemiczne transportowanego materiału. Parametry te wyznaczono na podstawie wyników badań właściwości wybranego jako materiał modelowy perlitu D18-DN. Zakres badań obejmował wyznaczenie:

a) gęstości względnej materiału – ρ_{mat} [g/cm^3],
b) ciężaru nasypowego – ρ_b [g/cm^3],
c) kąta naturalnego usypu – α_u [°],
d) współczynnika tarcia zewnętrznego pary ciernej stal-perlit – φ_z [–].

2.1. Oznaczenie gęstości względnej materiału

Gęstość względną materiału wyznaczono metodą piknometryczną. Pomiarów dokonywano piknometrem Gay-Lussaca o objętości 50 cm^3. W metodzie piknometrycznej jako ciecz wzorcową najczęściej stosuje się wodę destylowaną. Ze względu na własności chemiczne perlitu (możliwość jego ługowania w środowisku wodnym) jako cieczy wzorcowej użyto bezwodnego alkoholu etylowego. Wartość gęstości badanego materiału wyznaczano z zależności (1):

$$\rho_{mat} = \rho_{alk}^{t} \cdot \frac{m_3 - m_1}{(m_2 - m_1) - (m_4 - m_3)} \tag{1}$$

gdzie:

ρ_{mat} – gęstość materiału [g/cm^3],
ρ^t_{alk} – gęstość alkoholu etylowego w temperaturze t [g/cm^3],
m_1 – masa pustego piknometru [g],
m_2 – masa piknometru z alkoholem [g],
m_3 – masa piknometru z materiałem [g],
m_4 – masa piknometru z alkoholem i materiałem [g].

W obliczeniach uwzględniono zmianę gęstości względnej alkoholu etylowego w funkcji temperatury. Jako wartość gęstości względnej badanych materiałów przyjęto wartość średnią uzyskaną z trzech pomiarów.

2.2. Oznaczenie gęstości nasypowej materiału

Ciężar nasypowy stanowi stosunek ciężaru materiału do jego całkowitej objętości z uwzględnieniem porów, przestrzeni między ziarnami oraz ewentualnych zanieczyszczeń. Ciężar nasypowy materiału w stanie luźnym jest zależny od ciężaru właściwego materiału, jego wilgotności oraz stopnia wypełnienia nim danej objętości. Pomiaru dokonywano w pojemniku pomiarowym o objętości 1 dm^3, wyznaczając gęstość nasypową materiału w stanie luźnym ρ_b z zależności (2):

$$\rho_b = \frac{m_{2b} - m_1}{V} \qquad [g/cm^3] \tag{2}$$

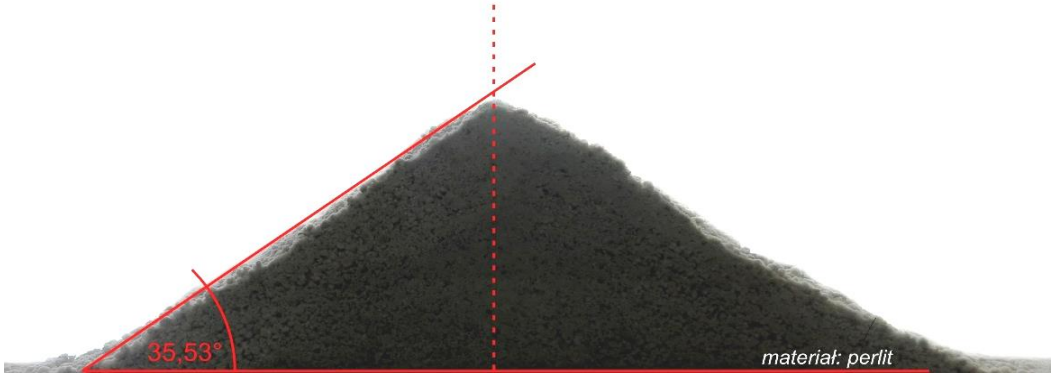
gdzie:

ρ_b – gęstość nasypowa w stanie luźnym [g/cm^3],
m_{2b} – masa cylindra pomiarowego i próbki w stanie luźnym [g],
m_1 – masa cylindra pomiarowego [g],
V – pojemność cylindra pomiarowego [cm^3].

Jako wartość gęstości nasypowej badanych materiałów przyjęto wartość średnią uzyskaną z trzech pomiarów.

2.3. Kąt naturalnego usypu

Kąt naturalnego usypu α_u jest to kąt nachylenia do poziomu płaszczyzny tworzącej stożka powstałego przez swobodne nasypywanie materiału na płaszczyznę poziomą. Badania wykonano w stanowisku pomiarowym przy luźnym zasypie badanego materiału. Wartość kąta usypu materiału wyznaczono poprzez odczyt bezpośrednio na stanowisku, wykonywano także zdjęcia do dalszej analizy. Stosując analizę obrazu określano kąty ułożenia materiału względem podłoża (rys. 2.).



Rys. 2. Pomiar kąta naturalnego usypu perlitu z wykorzystaniem analizy obrazu

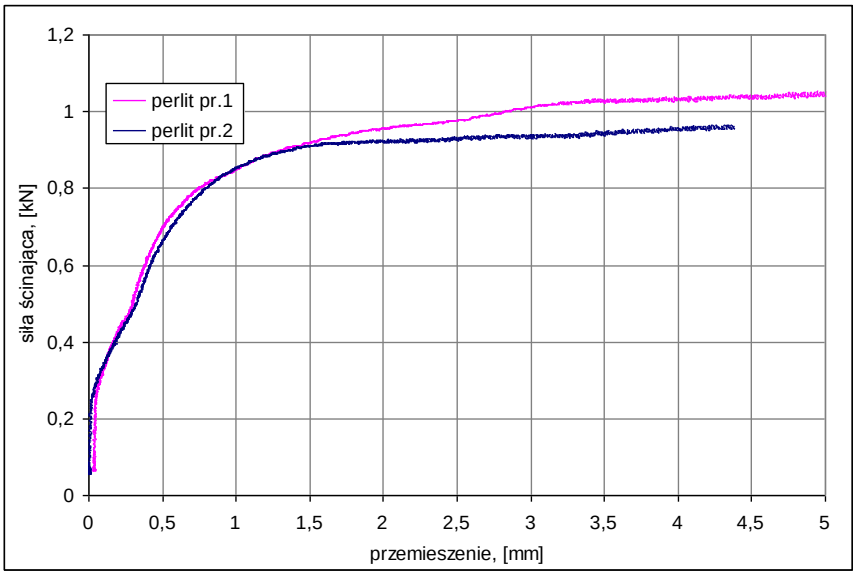
2.4. Współczynnik tarcia zewnętrznego

Wyznaczenie tarcia zewnętrznego prowadzono w aparacie do bezpośredniego ścinania typu AB-2A stosowanym do badań spójności gruntu. Poddana naprężeniu normalnemu próbka materiału ulegała, pod wpływem działającej poziomo siły, przemieszczeniu po stalowej płytce umieszczonej na dnie komory pomiarowej. Badania wykonano w skrzynce o wymiarach 60 x 60 mm przy naprężeniu normalnym wynoszącym 500 kPa. Na podstawie zarejestrowanych charakterystyk czasowych lub zmian wartości siły ścinającej w funkcji przemieszczenia wyznaczono współczynnik tarcia zewnętrznego w zależności (3):

$$\varphi_z = \frac{F_S}{F_N} \tag{3}$$

gdzie:

φ_Z – współczynnik tarcia zewnętrznego [–],
F_S – siła ścinająca [kN],
F_N – siła normalna [kN].



Rys. 3. Zmiana wartości siły ścinającej w funkcji przemieszczenia dla pary ciernej stal-perlit

Przykład zarejestrowanych zmian wartości siły ścinającej w funkcji przemieszczenia przedstawiono na rysunku 3. Obliczenia przeprowadzono przyjmując maksymalną wartość naprężenia stycznego lub wartość naprężenia stycznego odpowiadającą przemieszczeniu klatki o 6 mm (10% wymiaru klatki).

Wyniki badań właściwości fizykochemicznych materiału modelowego – perlitu D18-DN przedstawiono w tabeli 1.

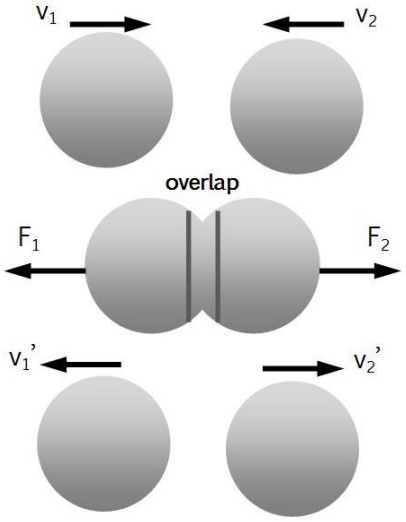
Tabela 1. Własności fizykochemiczne perlitu D18-DN

Własność fizykochemiczna	Wartość	
Gęstość względna ρ_{mat} [g/cm^3]	1,24	
Gęstość nasypowa ρ_b [g/cm^3]	0,10	
Kąt naturalnego usypu [°]	35,50	
Współczynnik tarcia zewnętrznego (para cierna stal-perlit D18-DN)	0,54	

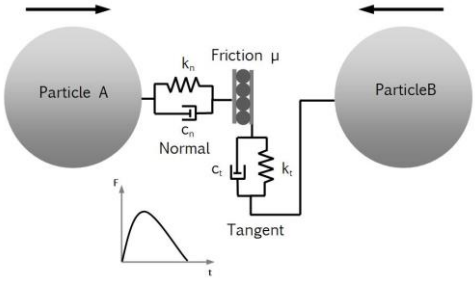
3. Opis metody DEM

Komputerowe metody wspomagania projektowania stały się w ostatnich latach istotnymi narzędziami wykorzystywanymi przez inżynierów w procesie projektowania nowych maszyn i urządzeń. Szereg zagadnień inżynierskich oraz naukowych, zbyt skomplikowanych do opisania przez modele analityczne lub trudnych do przebadania w laboratorium, zostało rozwiązanych dzięki metodom numerycznym [1,7-8,14-15,17-18]. Dodatkowo relatywnie tani i łatwy dostęp do wysokich mocy obliczeniowych oraz szybki rozwój programów komputerowych implementujących określone metody numeryczne sprawia, że zjawiska fizyczne modelowane są z niebywałą dokładnością w stosunku do rzeczywistości. Fundamentalnym przykładem jest Metoda Elementów Skończonych, powszechnie wykorzystywana do rozwiązywania problemów mechaniki ośrodków ciągłych (continuum). U podstaw metody MES leży założenie, że badany materiał jest ciągły w sensie makroskopowym, a więc charakteryzuje się ciągłym rozkładem materii (masy) w przestrzeni, natomiast jego własności wynikające z budowy atomowej (lub cząsteczkowej) są pomijane. Ideą Metody Elementów Skończonych jest podział rozpatrywanego obszaru ciągłego na skończoną liczbę podobszarów (elementów) połączonych ze sobą w punktach węzłowych. Dodatkowo zakłada się, że ciągłość ośrodka zachowana jest pod obciążeniem, stąd koncepcja MES wyklucza jego zastosowanie w symulowaniu zachowania materiałów sypkich. Dlatego na potrzeby mechaniki discontinuum opracowano Metodę Elementów Dyskretnych, w której badany materiał modelowany jest jako zbiór pojedynczych cząstek wchodzących ze sobą we wzajemną interakcję dzięki odpowiednim modelom kontaktowym. Stąd też makroskopowe cechy materiału są determinowane przyjętymi własnościami cząstek oraz ich wzajemnymi oddziaływaniami [19].

Algorytm metody DEM opiera się na mechanice kontaktu, dzięki której poprzez odpowiednie modele obliczane są siły działające na poszczególne cząstki. Na podstawie równań ruchu Newtona wyliczane są przyspieszenia cząstek. Prędkości oraz położenia cząstek wyznaczane są poprzez całkowanie w czasie tychże równań. Z założenia cząstki są bryłami sztywnymi. Jednak metoda DEM uwzględnia deformacje cząstek w sposób sztuczny, a mianowicie podczas kolizji wyliczana jest głębokość wzajemnej penetracji, jak przedstawiono na rysunku 4.



Rys. 4. Kolizja sztywnych cząstek



Rys. 5. Model kontaktowy dwóch cząstek w czasie kolizji

Zadaniem modelu kontaktowego jest powiązanie głębokości penetracji (w kierunku normalnym i stycznym do kierunku ruchu zderzających się cząstek) z wartością siły kontaktowej. Na rysunku 5 przed-stawiono model kontaktowy dwóch cząstek w czasie kolizji. Zwykle pakiety do symulacji DEM oferują szereg różnych modeli kontaktowych pozwalających na modelowanie zderzeń sprężystych, plastycznych z uwzględnieniem lub pominięciem kohezji.

Na zachowanie badanego ośrodka sypkiego duży wpływ mają parametry poszczególnych cząstek (rozmiar, kształt, gęstość) oraz parametry kontaktowe (współczynniki tarcia, opory toczenia, współczynnik restytucji) [4]. Podstawowym kształtem modelowej cząstki jest sfera, jednakże ze względu na fakt, że rzeczywiste kształty pojedynczych ziaren rzadko przybierają idealny, kulisty kształt istnieje możliwość tworzenia bardziej skomplikowanych brył w wyniku trwałego scalenia sferycznych cząstek. Należy jednak mieć świadomość, że użycie bardzo złożonych kształtów brył wiążę się z wydłużeniem czasu obliczeń. Stąd istotne jest znalezienie kompromisu pomiędzy rzeczywistym kształtem i rozmiarem brył, a tym użytym w symulacjach DEM. Aby odzwierciedlić realne zachowanie się materiału w symulacjach DEM, należy poddać kalibracji parametry wejściowe modelu (tj. współczynniki tarcia, gęstość, itp.).

4. Kalibracja parametrów DEM

W analizach DEM wszystkie parametry fizyczne cząstek oraz parametry kontaktowe definiuje się na poziomie mikroskopowym cząstki [16]. Kalibracja polega, na doborze wartości parametrów w skali mikro, aby uzyskać zbieżny wynik modelowanego złoża z zachowaniem złoża rzeczywistego (skala makro). Zatem ważniejsze jest dokładne odwzorowanie zachowania całego złoża aniżeli pojedynczych interakcji międzycząsteczkowych. Kalibracji parametrów dokonuje się poprzez wykonanie odpowiedniego eksperymentu badawczego np. badania kąta naturalnego usypu lub próby ścinania. Następnie ten sam eksperyment modeluje się w środowisku DEM. Zakładając określony kształt i rozmiar cząstek dokonuje się iteracyjnego doboru parametrów wejściowych modelu, aż do momentu uzyskania zbieżnego wyniku z przeprowadzonym eksperymentem [19]. W ramach niniejszej pracy przedstawiony został proces kalibracji parametrów wejściowych DEM na przykładzie wyników pomiarów laboratoryjnych własności fizykochemicznych perlitu D18-DN.

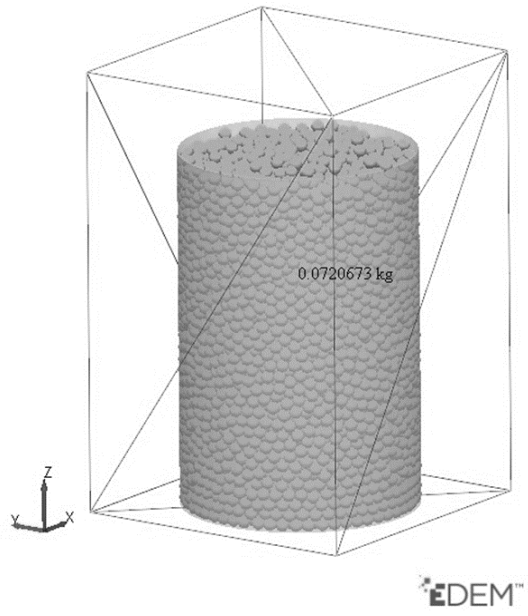
Kalibrację modelu DEM rozpoczyna się od doboru kształtu i rozmiaru cząstek. Nie ma jednoznacznej reguły mówiącej, w jakim stopniu uprościć modelowany kształt bryły oraz jaki przyjąć jej wymiar. Szereg badań symulacyjnych wykonanych przez wielu badaczy wskazuje, że dobór ten jest podyktowany konkretnym przeznaczeniem symulacji. Zarówno skomplikowane bryły jak i cząstki sferyczne są w stanie odzwierciedlić realny kąt naturalnego usypu. W przypadku tych drugich należy wprowadzić do modelu DEM dodatkowy współczynnik

ograniczający toczenie się sfer. Istotnym czynnikiem symulacji DEM jest czas obliczeń, na który wpływa między innymi ilość użytych cząstek oraz ich rozmiar (zgodnie z teorią Rayleigha). W przypadku materiałów drobnoziarnistych lub proszków, rozmiary cząstek zostają przeskalowane tak, aby zapewnić rozsądny kompromis między dokładnością obliczeń, a czasem ich wykonania. Zwykle rozmiar cząstek modelowanego materiału dobiera się na zasadzie proporcji wymiaru charakterystycznego urządzenia (np. skok ślimaka) do promienia cząstki.

Biorąc po uwagę regularny kształt perlitu, przyjęto sferyczne cząstki o promieniach 2, 5 i 8 mm. Rozmiary zostały przeskalowane względem rzeczywistych wymiarów ziaren w celu zredukowania ich liczby w złożu, a tym samym ograniczeniu czasu obliczeń. Ze względu na fakt, że perlit nie wykazuje własności kohezyjnych, w symulacji użyto nieliniowego modelu kontaktowego według teorii Hertza-Mindlina. Następnie dla każdego wymiaru cząstki wykonano kalibrację gęstości nasypowej. Polegała ona na swobodnym zasypaniu cylindra pomiarowego o określonej objętości i odczytaniu masy złoża zajmującego wypełnioną przestrzeń cylindra (rys. 6). Parametrem wejściowym modelu była gęstość rzeczywista pojedynczej cząstki. Gęstość tą dobierano w sposób iteracyjny tak, aby uzyskać wartość gęstości nasypowej zbieżną z uzyskaną w wyniku pomiarów laboratoryjnych. Wyniki powyższej analizy dostarczyły również informacji, w jaki sposób rozmiar dobranej cząstki wpływa na wartość gęstości usypowej. W tabeli 2 przedstawiono wyniki gęstości nasypowej w zależności od promieni cząstek oraz gęstości rzeczywistej.

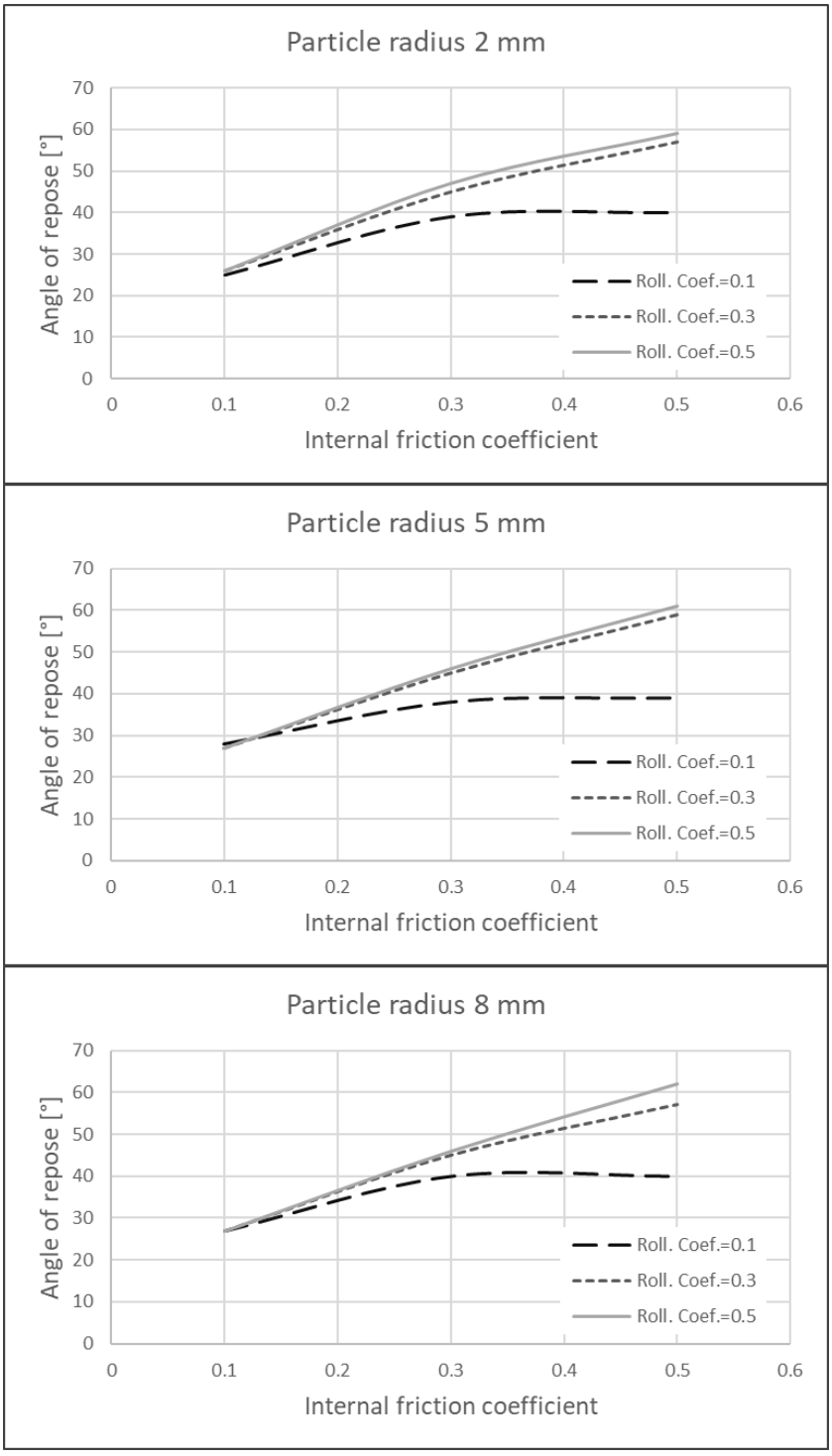
Tabela 2. Wyniki kalibracji gęstości nasypowej perlitu

Promień cząstki [mm]	Gęstość rzeczywista cząstki [kg/m^3]	Gęstość nasypowa [kg/m^3]
2	200	119,4
5	200	119,9
8	200	119,7



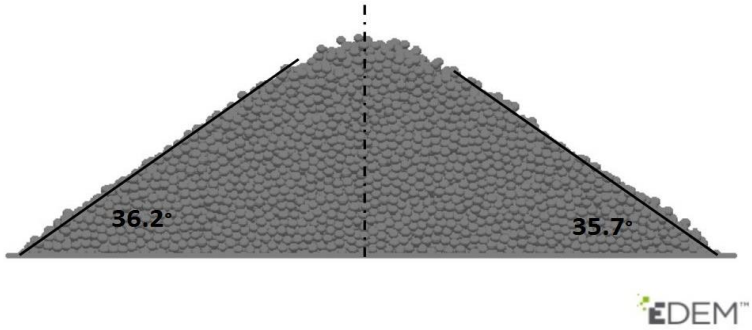
Rys. 6. Kalibracja gęstości nasypowej

Badanie naturalnego kąta usypu miało na celu skalibrowanie parametrów kontaktowych modelu, tj. współczynnika tarcia wewnętrznego oraz oporów tarcia tocznego. W symulacji DEM odwzorowano rzeczywiste stanowisko do badania naturalnego kąta usypu [10]. Kalibracja polegała na iteracyjnym doborze wartości obu współczynników, co pozwoliło uzyskać wartość kąta naturalnego usypu perlitu zbliżoną do rzeczywistej. W tym celu przyjęto plan badań polegający na zmianie wartości tylko jednego parametru wejściowego tj. współczynnika tarcia wewnętrznego przy stałej wartości współczynnika oporów toczenia. Seria badań symulacyjnych została powtórzona dla trzech wartości współczynnika oporów toczenia, dzięki czemu uzyskano charakterystyki łączące współczynniki tarcia wewnętrznego i oporów toczenia z wartością kąta naturalnego usypu. Rysunek 7 przedstawia wykresy zmian wartości kąta naturalnego usypu w zależności od przyjętych parametrów wejściowych.



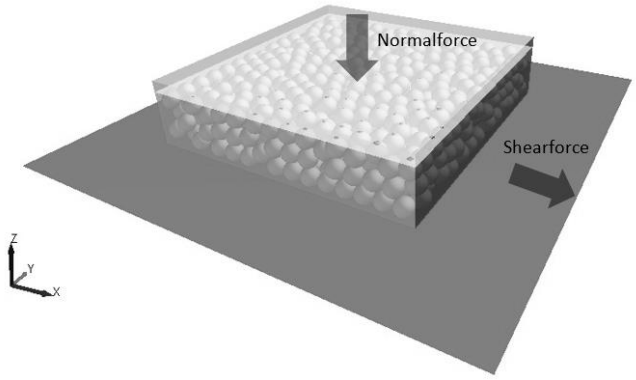
Rys. 7. Wpływ parametrów wejściowych na wartość kąta naturalnego usypu

Rysunek 8 przedstawia skalibrowany stożek usypowy. Podejście to przy założeniu trzech różnych promieni cząstek, trzech poziomów wartości współczynnika tarcia wewnętrznego i trzech wartości współczynnika oporów toczenia wymaga przeprowadzenia 27 symulacji. Badania te są bez wątpienia czasochłonne, dostarczają jednak istotnych informacji, które parametry wejściowe decydują o wartości kąta naturalnego usypu.



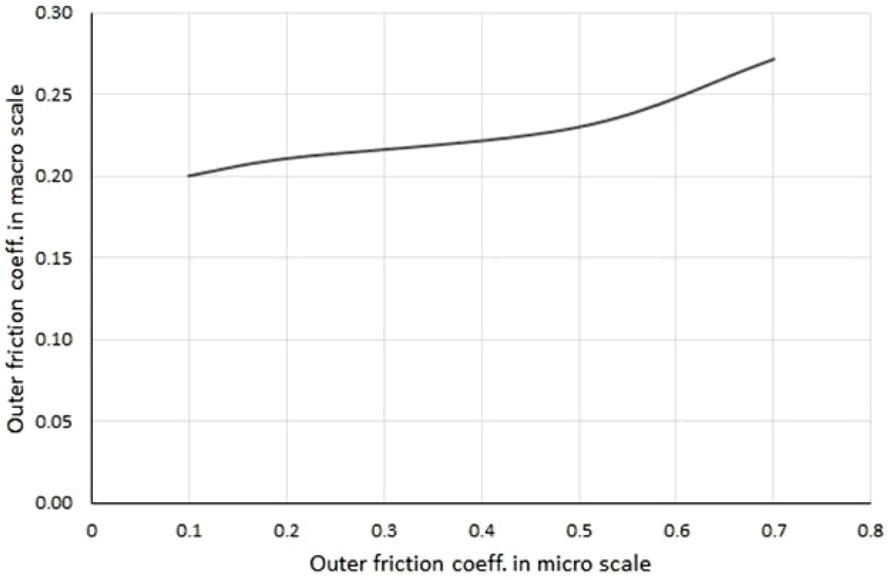
Rys. 8. Skalibrowany stożek naturalnego usypu

Kalibracja współczynnika tarcia zewnętrznego (materiał-stal) wykonana została na modelu aparatu skrzynkowego do badania bezpośredniego ścinania złoża. Model stanowiska pokazany został na rysunku 9.



Rys. 9. Kalibracja współczynnika tarcia zewnętrznego

Próbka kalibrowanego materiału została umieszczona wewnątrz skrzyni ograniczonej od dołu nieruchomą płytą i od góry sztywnym stemplem generującym nacisk na materiał. Poprzez iteracyjną zmianę współczynnika tarcia pary kontaktowej materiał-płyta oraz rejestrację zmiany siły normalnej oraz stycznej możliwe było oszacowanie (dzięki zależności 3) współczynnika tarcia zewnętrznego. Wpływ współczynnika tarcia w skali mikro na wartość współczynnika tarcia w skali makro przedstawiono na rysunku 10.



Rys. 10. Zależność współczynnika tarcia zewnętrznego złoża od współczynnika tarcia pojedynczej cząstki

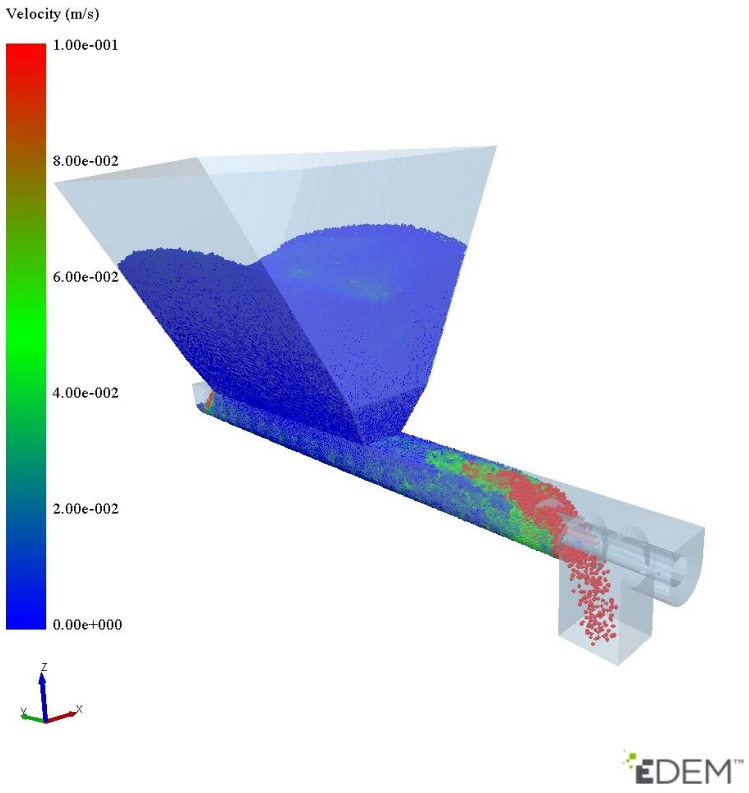
5. Symulacja transportu perlitu D18-DN w przenośniku ślimakowym

Dysponując odpowiednio skalibrowanymi parametrami wejściowymi modelu DEM można przystąpić do wykonywania właściwych symulacji. Mogą one posłużyć między innymi do określenia spodziewanej wydajności transportowej projektowanego urządzenia czy też wyznaczenia wymaganej mocy napędu. Poniżej przedstawiono przykładową symulację transportu perlitu D18-DN w przenośnikach ślimakowych. Analizie poddano dwa ślimaki o tych samych średnicach zewnętrznych oraz wewnętrznych, różniących się wyłącznie wymiarem skoku. Przyjęto wariant I – ślimak o skoku 250 mm oraz wariant II – ślimak o skoku 150 mm. Podstawowe wymiary geometryczne przedstawiono w tabeli 3.

Tabela 3. Parametry analizowanych przenośników ślimakowych

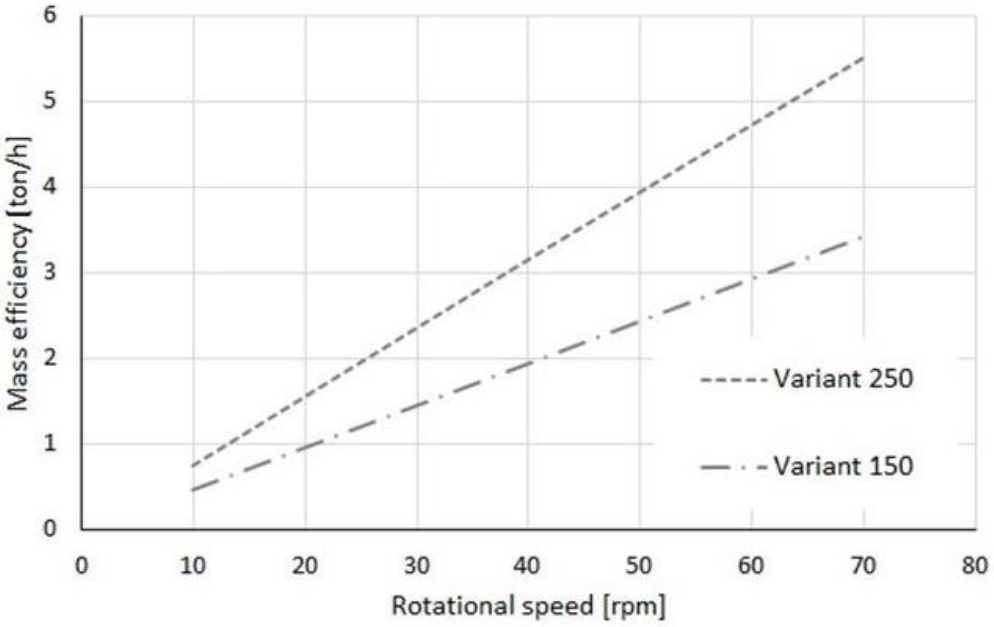
Wariant ślimaka	Średnica zewnętrzna [mm]	Średnica wewnętrzna [mm]	Skok ślimaka [mm]
Wariant I	250	139	250
Wariant II	250	139	150

Wyniki symulacji miały dostarczyć informacji jak skok ślimaka wpływa na wydajność oraz zapotrzebowanie na moc przy różnych prędkościach obrotowych wału. Na potrzeby analizy wybrano cząstki o wymiarach 5 mm. Symulację dla obu wariantów ślimaka wykonano zmieniając jego prędkość obrotową w zakresie 10-70 obr/min. Model przenośnika ślimakowego wraz z materiałem przedstawiono na rysunku 11.



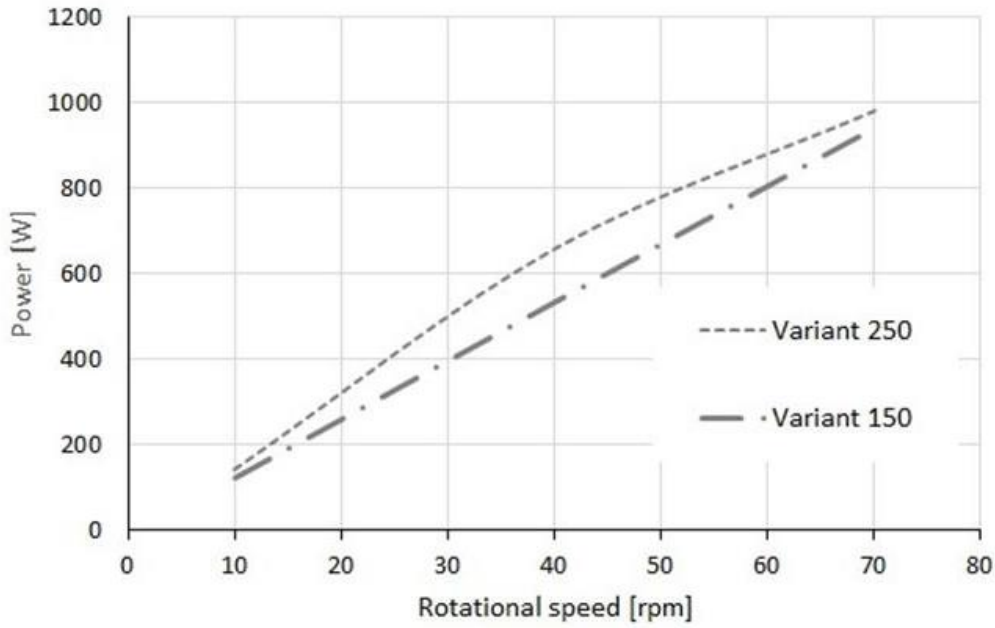
Rys.11. Model przenośnika ślimakowego

Wyniki symulacji dla obu wariantów przenośnika przedstawiające wydajność masową w zależności od prędkości obrotowej wału ślimaka przedstawiono na rysunku 12.



Rys. 12. Wydajność masowa przenośnika w zależności od prędkości obrotowej i skoku ślimaka

Zmiana skoku linii śrubowej przenośnika wpływa bezpośrednio na jego wydajność transportową. Wraz ze wzrostem skoku przy tej samej prędkości obrotowej wału rośnie wydajność masowa. Dysponując odpowiednio skorelowanym modelem materiału w prosty sposób można zbadać funkcjonalność danego urządzenia transportowego lub dobrać jego odpowiednie wymiary geometryczne. Oprócz określenia cech geometrycznych można również określić wymaganą moc napędu. Na rysunku 13 przedstawiono zapotrzebowanie na moc przenośnika ślimakowego w zależności od prędkości obrotowej i skoku ślimaka w obu przyjętych wariantach.



Rys. 13. Zapotrzebowanie na moc przenośnika w zależności od prędkości obrotowej i skoku ślimaka

Z powyższego wykresu widać, że wraz ze wzrostem prędkości obrotowej wzrasta zapotrzebowanie przenośnika na moc. Dla tej samej prędkości obrotowej wału, zmniejszenie skoku ślimaka skutkuje obniżeniem zapotrzebowania mocy. Wzrost zapotrzebowania na moc nie jest wprost proporcjonalny do wzrostu wydajności, co szczególnie widoczne jest przy prędkościach wału powyżej 50 obr/min. Przykładowo przy prędkości obrotowej 70 obr/min i zmianie skoku ze 150 mm do 250 mm, wydajność masowa wzrosła o 60 %, natomiast zapotrzebowanie na moc wzrosło tylko o 6,5 %. Dysponując charakterystykami pracy urządzenia można dokonać optymalizacji wymiarów i masy ślimaka czy też określić korzystny zakres prędkości obrotowej. Przyspiesza to proces projektowania i pozwala na wirtualne testowanie różnych rozwiązań.

6. Podsumowanie

Proces kalibracji parametrów wejściowych modelu DEM jest czynnikiem mocno determinującym wiarygodność końcowych wyników symulacji przepływu materiałów sypkich w urządzeniach transportowych. Badania laboratoryjne określające własności fizykochemiczne materiałów są stosunkowo łatwe i tanie do wykonania, zwłaszcza w przypadku materiałów drobnoziarnistych. Jest to bez wątpienia ogromny atut, gdyż na ich podstawie można dokonać kalibracji modelu numerycznego. Jak wykazano sam proces kalibracji DEM, ze względu na iteracyjny dobór wartości poszczególnych parametrów, jest procesem czasochłonnym. Badania symulacyjne dowiodły, że promień cząstki kulistej nie wpływa na gęstość nasypową złoża. Wobec tego, można oznaczyć ją używając jednego rozmiaru cząstek. Podobnie, rozmiar cząstek nie wpływa zasadniczo na wartość kąta naturalnego usypu. Z powyższych badań wynika, że tą samą wartość kąta stożka nasypowego można uzyskać stosując różne zestawy parametrów współczynnika tarcia wewnętrznego oraz współczynnika oporów toczenia. Wartości współczynników oporów toczenia $F = 0{,}3$ oraz $F = 0{,}5$ nieznacznie wpływają na wartość kąta naturalnego usypu. Odmienne wyniki uzyskano przyjmując współczynnik oporów toczenia $F = 0{,}1$. W tym przypadku dla wszystkich trzech przyjętych w symulacjach średnic cząstek (2, 5 i 8 mm) ze wzrostem współczynnika tarcia wewnętrznego μ_w wartość kąta naturalnego usypu α_u początkowo rośnie, a następnie przyjmuje stałą wartość $\alpha_u \approx 40°$dla $\mu_w \geq 0{,}3$. Zjawisko to wytłumaczyć można zbyt niskim momentem oporów toczenia cząstek powodującym ich staczanie po zboczu stożka. Oznacza to, że nawet przy bardzo wysokiej wartości współczynnika tarcia wewnętrznego kąt naturalnego usypu będzie mały. Stąd, w przypadku cząstek sferycznych konieczne jest używanie modelu oporów toczenia o odpowiedniej wartości (skalibrowanej). Współczynnik tarcia zewnętrznego złoża zależy nieliniowo od wartości współczynnika pojedynczej cząstki.

Tak skalibrowany model materiałowy DEM powinien zostać poddany weryfikacji poprzez symulację konkretnej aplikacji np. transportu w przenośniku ślimakowym. Jako cechy porównawcze powinny zostać wyznaczone parametry pracy urządzenia tj. wartość momentu skręcającego wał ślimaka, wydajność masowa czy pobór mocy, które można skonfrontować z wynikami uzyskanymi na rzeczywistym obiekcie. Dodatkowo dalsze badania symulacyjne powinny skupić się na cząstkach niesferycznych ze względu na fakt, że rzeczywiste bryły mają bardziej złożone kształty.

Bez wątpienia istotną zaletą symulacji DEM jest możliwość szybkiego porównania kilku konstrukcji w aspekcie przyjętych kryteriów np. wydajności transportowej czy mocy. Ilościowa i jakościowa ocena tych różnic pozwala na właściwy dobór cech geometrycznych przenośnika ślimakowego transportującego określony typ materiału sypkiego. Usprawnia to proces projektowania i pozwala na zbadanie różnych konfiguracji bez konieczności wykonywania kosztownych prototypów urządzeń.

Literatura

1. Balac M, Grbovic A, Petrovic A, Popovic V. FEM analysis of pressure vessel with an investigation of crack growth on cylindrical surface. Eksploatacja i Niezawodność – Maintenance and Reliability 2018; 20 (3): 378–386.
2. Campello M.B E. A computational model for the simulation of dry granular materials. International Journal of Non-Linear Mechanics 2018; (106): 89-107.
3. Chen W, Roberts A, Katterfeld A, Wheeler C. Modelling the stability of iron ore bulk cargoes during marine transport. Powder Technology 2018; (326): 255-264.
4. Coetzee C J. Review: Calibration of the discrete element method. Powder Technology 2017; 310: 104-142.
5. Coetzee C J. Calibration of the discrete element method and the effect of particle shape. Powder Technology 2016; 297: 50-70.
6. Elskamp F, Kruggel-Emden H. DEM simulations of screening processes under the influence of moisture. Chemical Engineering Research and Design 2018; (136): 593-609.
7. Falkowicz K, Ferdynus M, Dębski H. Numerical analysis of compressed plates with a cut-out operating in the geometrically nonlinear range. Eksploatacja i Niezawodność – Maintenance and Reliability 2015: 17 (1): 222.
8. Ferdynus M, Kotełko M, Kral J. Energy absorption capability numerical analysis of thin-walled prismatic tubes with corner dents under axial impact. Eksploatacja i Niezawodność – Maintenance and Reliability 2018; 20 (1): 252–259.
9. Fernandes M.A. P, Chaves Pereira M. J, Pereira C.F. J. Discrete particle simulation in horizontally rotating drum: Uncertainty quantification of granular material physical parameters. Powder Technology 2018; (339): 199-210.
10. Grima A, Wypych P. Discrete element simulation of a conveyor impact-plate transfer: calibration, validation and scale-up. Australian Bulk Handling Review 2010.
11. Gröger T, Katterfeld T. On the numerical calibration of discrete element models for the simulation of bulk solids. Conveying and Handling of Particulate Solids. Italy 2006.
12. Gröger T, Katterfeld A. Application of the Discrete Element Method in Materials Handling: Basics and Calibration. Bulk Solid Handling 2007; 27.
13. Gröger T, Katterfeld A. Verified Discrete Element Simulations of Bulk Solids Handling Equipment. AIChE Spring Meeting and Global Congress on Process Safety 2006.
14. Labra C, Oñate E, Kargl H, Rojek J. Discrete Element Simulation of rock cutting. International Journal of Rock Mechanics and Mining Sciences 2011; 48: 996-1010.
15. Mitura A, Gawryluk J, Teter A. Numerical and experimental studies on the rotating rotor with three active composite blades. Eksploatacja i Niezawodność – Maintenance and Reliability 2017; 19 (3): 571-579.
16. Powder Calibration: Challenges, Advances & Guidelines. edem.com 2018.
17. Rubacha P, Kulinowski P, Furmanik K, Badania laboratoryjne i symulacyjne ruchu ziaren w rynnie przesypowej przenośnika taśmowego. Transport Przemysłowy i Maszyny Robocze 2014; (1): 25-29.

18. Śledziewski K. Experimental and numerical studies of continuous composite beams taking into consideration slab cracking. Eksploatacja i Niezawodność – Maintenance and Reliability 2016; 18 (3): 578-589.
19. WHAT IS DEM Theoretical background behind the Discrete Element Method (DEM). edem.com 2018.