Vol. 20. No 4, 2018

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

Editorial Board

Prof. Andrzej Niewczas
Chair of Editorial Board
President of the Board of the Polish Maintenance Society

Prof. Holm Altenbach
Otto-von-Guericke-Universität, Magdeburg, Germany

Prof. John Andrews
University of Nottingham, Nottingham, UK

Prof. Karol Andrzejczak
Poznań University of Technology, Poznań

Prof. Christophe Bérenguer
Institut Polytechnique de Grenoble, Grenoble, France

Prof. Gintautas Bureika
Vilnius Gediminas Technical University, Vilnius, Lithuania

Dr Alireza Daneshkhah
Warwick Centre for Predictive Modelling
University of Warwick, UK

Prof. Sławczo Denczew
The Main School of Fire Service, Warsaw, Poland

Prof. Luis Andrade Ferreira
University of Porto, Porto, Portugal

Prof. Mitra Fouladirad
Troyes University of Technology, France

Dr Ilia Frenkel
Shamoon College of Engineering, Beer Sheva, Israel

Prof. Olgierd Hryniewicz
Systems Research Institute of the Polish Academy of Science, Warsaw, Poland

Prof. Hong-Zhong Huang
University of Electronic Science and Technology of China, Chengdu, Sichuan, China

Prof. Vaclav Legat
Czech University of Agriculture, Prague, Czech Republic

Prof. Jerzy Merkisz
Poznań University of Technology, Poznań, Poland

Prof. Gilbert De Mey
University of Ghent, Belgium

Prof. Maria Francesca Milazzo
University of Messina, Italy

Prof. Tomasz Nowakowski
Wrocław University of Technology, Wrocław, Poland

Prof. Marek Orkisz
Rzeszów University of Technology, Rzeszów, Poland

Prof. François Pérès
Touluse University, Touluse, France

Prof. Jan Szybka
AGH University of Science and Technology, Cracow, Poland

Prof. Marcin Ślęzak
Motor Transport Institute, Warsaw, Poland, Poland

Prof. Katsumi Tanaka
Kyoto University, Kyoto, Japan

Prof. David Vališ
University of Defence, Brno, Czech Republic

Prof. Lesley Walls
University of Strathclyde, Glasgow, Scotland

Prof. Min Xie
City University of Hong Kong, Hong Kong

Prof. Irina Yatskiv
Riga Transport and Telecommunication Institute, Latvia

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

KLUZ R, KUBIT A, SĘP J, TRZEPIECINSKI T. **Effect of temperature variation on repeatability positioning of a robot when assembling parts with cylindrical surfaces**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 503–513, http://dx.doi.org/10.17531/ein.2018.4.1.

In this paper, the effect of the errors induced by temperature changes on the repeatability positioning error of an industrial robot is analysed. It has been shown that after the stabilization of the thermal conditions, these errors can be identified with the systematic errors. It has also been shown that if the ambient temperature cannot be sufficiently stabilized, the temperature errors can be described using a normal or uniform probability distribution. Depending on the choice of a point in the robot's workspace and temperature fluctuations, these errors can comprise a small share of the total error of the robot. Then the total repeatability positioning error can be approximated with sufficient accuracy by a normal probability distribution or it can comprise the dominant component of this error. In this case, the total error can be approximated using a flat normal distribution. It has been shown that, depending on the choice of location in the workspace in which the assembly operation is carried out, it is possible to obtain both different probabilities of assembling the parts correctly and a different effect of errors caused by slight temperature changes on the value of those probabilities. The results found indicate the potential possibility of increasing the reliability of the process by proposing the selection of the location in the robot workspace which has the lowest sensitivity to errors ascribed to changes in temperature.

KLUZ R, KUBIT A, SĘP J, TRZEPIECINSKI T. **Wpływ zmian temperatury na powtarzalność pozycjonowania robota przy montażu części o powierzchniach cylindrycznych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 503–513, http://dx.doi.org/10.17531/ein.2018.4.1.

W niniejszej pracy przeanalizowano wpływ zmiany temperatury na błąd powtarzalności pozycjonowania robota przemysłowego. Wykazano, że po ustabilizowaniu się warunków termicznych błędy te można sklasyfikować jako błędy systematyczne. Wykazano również, że jeżeli w trakcie eksploatacji zrobotyzowanego stanowiska montażowego temperatura otoczenia nie może być wystarczająco ustabilizowana, błędy temperatury można opisać za pomocą jednostajnego rozkładu prawdopodobieństwa i w ten sposób uwzględnić w strukturze całkowitego błędu powtarzalności pozycjonowania. Błędy te na ogół stanowią niewielki udział w całkowitym błędzie robota, wówczas całkowity błąd powtarzalności pozycjonowania robota z dostateczną dokładnością można aproksymować normalnym rozkładem prawdopodobieństwa. W przeciwnym przypadku błąd ten może być przybliżony przy użyciu rozkładu płasko-normalnego. Wykazano że w zależności od wyboru miejsca realizacji zabiegu montażowego w przestrzeni stanowiska można uzyskać zarówno odmienne wartości prawdopodobieństwa połączenia części jak i różny wpływ błędów wywołanych niewielkimi zmianami temperatury na wartość tego prawdopodobieństwa. Uzyskane wyniki badań wskazują na potencjalne możliwości zwiększenia niezawodności procesu poprzez wybór miejsca w przestrzeni roboczej stanowiska charakteryzującego się najniższą wrażliwością na błędy spowodowane zmianami temperatury.

OU D, TANG M, XUE R, YAO H. **Hybrid fault diagnosis of railway switches based on the segmentation of monitoring curves**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 514–522, http://dx.doi.org/10.17531/ein.2018.4.2.

Switches are one of the most important pieces of infrastructure in railway signal systems, and they significantly influence the efficiency and safety of train operation. Currently, the identification of switch failures mainly depends on the experience of railway staff and the use of simple thresholding methods. However, these basic methods are highly inaccurate and frequently result in false and missing alarms. This paper aims to develop a hybrid fault diagnosis (HFD) method for railway switches. The method is an intelligent diagnosis method that uses massive current curves collected by microcomputer monitoring systems. We first divide the switch operation current curves into three segments based on the three mechanical processes that occur during switch operation. Then, a standard curve is selected from the fault-free curves, and common typical faults are ascertained through a microcomputer monitoring system. Finally, derivative dynamic time warping and a quartile scheme are employed to identify fault curves. An experiment based on current curves collected from the Guangzhou Railway Bureau in China demonstrates that the HFD method is extremely accurate and has low false and missing alarm rates. HFD performs better than the studied support vector machine (SVM) and dynamic time warping (DTW) methods, which are widely used for fault diagnosis.

OU D, TANG M, XUE R, YAO H. **Hybrydowa diagnostyka uszkodzeń zwrotnic kolejowych w oparciu o segmentację krzywych prądowych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 514–522, http://dx.doi.org/10.17531/ein.2018.4.2.

Zwrotnice stanowią jeden z najważniejszych elementów infrastruktury systemów sygnalizacji kolejowej i mają znaczący wpływ na wydajność i bezpieczeństwo eksploatacji pociągów. Obecnie, identyfikacja awarii zwrotnic zależy głównie od doświadczenia personelu kolejowego i opiera się na stosowaniu prostych metod progowania. Jednakże te elementarne metody są wysoce niedokładne i często skutkują fałszywymi alarmami lub brakiem alarmu. Niniejszy artykuł ma na celu opracowanie hybrydowej metody diagnostyki błędów (HFD) dla zwrotnic kolejowych. Metoda ta jest inteligentną metodą diagnostyczną, która wykorzystuje wykresy przebiegu prądowego zebrane przez mikrokomputerowe systemy monitorowania. Najpierw krzywe prądowe działania zwrotnicy dzieli się na trzy segmenty w oparciu o trzy procesy mechaniczne, które zachodzą podczas jej działania. Następnie, spośród krzywych opisujących działanie bezusterkowe, wybiera się przebieg standardowy, a w dalszej kolejności ustala się, z wykorzystaniem mikrokomputerowego systemu monitorowania, najczęściej występujące, typowe błędy działania zwrotnicy. Wreszcie, do identyfikacji krzywych błędów stosuje się schemat kwartylowy oraz metodę derivative dynamic time warping wykorzystującą pochodne do klasyfikacji szeregów czasowych. Eksperyment oparty na krzywych prądowych zebranych przez Guangzhou Railway Bureau w Chinach pokazuje, że metoda HFD jest wyjątkowo dokładna i skutkuje niską liczbą fałszywych i brakujących alarmów. HFD daje lepsze wyniki niż szeroko stosowane do diagnozowania błędów metody maszyny wektorów nośnych (SVM) i dynamic time warping (DTW).

WANG L, PEI Z, ZHU H, LIU B. **Optimising extended warranty policies following the two-dimensional warranty with repair time threshold**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 523–530, http://dx.doi.org/10.17531/ein.2018.4.3.

This paper considers an optimal extended warranty policies after the expiration of base two-dimensional warranty with repair time threshold. During the base two-dimensional warranty period, each failure of the equipment can be either replaced or minimally repaired depending on a pre-specified repair time threshold. After the base warranty expires, the length of an extended warranty policy is available for selection. The equipment is minimally repaired on each failure during the extended warranty. In this study, the length of the extended warranty period is optimized by minimizing the expected cost rate incurred over the whole warranty coverage, from the views of customs and manufacturers respectively. For the purpose of illustration, we present and discuss some numerical examples. The effect of repair time threshold on the optimal strategy is also investigated numerically.

WANG L, PEI Z, ZHU H, LIU B. **Optymalizacja polityki gwarancyjnej w okresie po wygaśnięciu dwuwymiarowej gwarancji z ustaloną górną granicą czasu naprawy**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 523–530, http://dx.doi.org/10.17531/ein.2018.4.3.

W niniejszej pracy rozważano optymalną politykę przedłużania gwarancji po wygaśnięciu podstawowej gwarancji dwuwymiarowej z ustaloną górną granicą czasu naprawy. W podstawowym okresie obowiązywania gwarancji dwuwymiarowej, po każdej awarii urządzenie zostaje poddane minimalnej naprawie, lub – jeśli naprawa nie może być wykonana we wcześniej ustalonym czasie naprawy – wymienione. Po wygaśnięciu gwarancji podstawowej, konieczne jest wybranie długości okresu obowiązywania gwarancji rozszerzonej. Podczas trwania okresu gwarancji przedłużonej, sprzęt naprawia się w sposób minimalny (naprawa minimalna) po każdorazowym uszkodzeniu. W niniejszej pracy, optymalizowano długość przedłużonego okresu gwarancyjnego poprzez minimalizację oczekiwanych kosztów poniesionych podczas całego okresu trwania gwarancji; optymalizację przeprowadzono z perspektywy klienta jak i producenta . Dla ilustracji, przedstawiono i omówiono wybrane przykłady numeryczne. Przeprowadzono także analizę numeryczną wpływu górnej granicy czasu naprawy na optymalną strategię gwarancyjną.

LEONARCIK R, URBANIAK M, DĘBKOWSKI R. **Method for assessing the grinding wheels operational properties**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 531–541, http://dx.doi.org/10.17531/ein.2018.4.4.

The paper presents a new, multi-criteria method which allows the numerical evaluation of the machining process in terms of efficiency, quality and costs. Three indicators were developed to assess the operational properties of grinding wheels. Their values are determined on the basis of the results of short grinding tests carried out on a special test stand. The evaluation of the proposed indicators is described. Furthermore, the application exemple of this method in determining the grinding wheel's operational properties is presented. In the research, the vitrified alumina oxide grinding wheels

LEONARCIK R, URBANIAK M, DĘBKOWSKI R. **Metoda oceny właściwości eksploatacyjnych ściernic**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 531–541, http://dx.doi.org/10.17531/ein.2018.4.4.

W artykule przedstawiono nową, wielokryterialną metodę, w której grupa trzech wskaźników eksploatacyjnych, wyznaczonych na podstawie danych procesowych z krótkotrwałego testu pracy pary ściernica-przedmiot obrabiany przeprowadzonego na specjalnym stanowisku badawczym, pozwala na liczbowe wartościowanie procesu obróbki pod względem wydajności, jakości i kosztów. Przedstawiono wyniki badań ewaluacji zaproponowanych wskaźników oraz badania aplikacyjne metody w zakresie oceny właściwości eksploatacyjnych ściernic podczas szlifowania stali. W badaniach stosowano ceramiczne ściernice elektrokorundowe oraz stale narzędziowe were used for grinding of constructional and tool steels of various hardness. The results of the experiments show that the proposed indicators are an effective tool for assessing the process and results of grinding for a specific grinding wheel and material within certain tested grinding parameters range. The study also showed that the differences in indicators' values, observed during tests of grinding specific material type using grinding wheels with different properties, are useful for optimizing the choice of tool type and machining conditions.

i konstrukcyjne o różnej twardości. Wyniki doświadczeń wykazały, że zaproponowane formuły wskaźników są skutecznym narzędziem oceny przebiegu i wyników szlifowania dla określonej pary ściernica-materiał w badanym zakresie wartości nastawnych procesu szlifowania. Badania wykazały także, że różnice wartości wskaźników występujące podczas testów szlifowania określonego rodzaju materiału ściernicami o różnej charakterystyce, są przydatne do optymalizacji wyboru rodzaju narzędzia i warunków obróbki.

ŚWIDER J, ZBILSKI A. **Power losses and their properties for low range of a robot electric motor working conditions as the part of energy effectiveness research**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 542–548, http://dx.doi.org/10.17531/ein.2018.4.5.

Power losses are one of many factors affecting the energy effectiveness of production processes, however despite this, commonly investigated ranges of power losses do not explain how they change in the stages being different from a typical driving mode. This investigation focuses on low working conditions of a robot electric motor and the properties of power losses changes while going from a driving mode into a stand-still mode of electric motor work. Apart from determined values of power maps components, this work shows how to manage with technical limitations in performing measurements of industrial robot electrical states at the industrial conditions, like high disturbances, noise and limited range of robot axis angle position.

ŚWIDER J, ZBILSKI A. **Straty mocy oraz ich własności w niskim zakresie warunków pracy silnika elektrycznego robota w kontekście badań efektywności energetycznej**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 542–548, http://dx.doi.org/10.17531/ein.2018.4.5.

Straty mocy są jednym z wielu czynników wpływających na efektywność energetyczną procesów produkcyjnych, jednak pomimo tego, najczęściej badane zakresy strat mocy nie określają sposobu ich zmian w trybach pracy odmiennych od typowej pracy napędowej. Opisane badania zostały skoncentrowane na niskim zakresie warunków pracy silnika robota przemysłowego oraz na własnościach zmian postaci strat mocy podczas przechodzenia ze stanu pracy napędowej do pracy statycznej. Oprócz wyznaczonych wartości komponentów map mocy, w pracy przedstawiono techniczne rozwiązania umożliwiające wykonywanie pomiarów stanów elektrycznych robota w warunkach przemysłowych, którymi były zniekształcenia, zakłócenia oraz ograniczony zakres pozycji kątowych badanego przegubu robota.

SKAČKAUSKAS P, SOKOLOVSKIJ E. **Modification and reliability estimation of vector based Dubins path approach for autonomous ground vehicles path re-planning**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 549–557, http://dx.doi.org/10.17531/ein.2018.4.6.

Due to global purposes to ensure growth of a competitive and sustainable transport system, also to solve traffic safety and environmental problems, various engineering solutions are being sought out. It can be assumed that autonomous vehicles are the technology, which will ensure the positive change in the transport system. Even though many studies successfully advanced toward realisation of autonomous vehicles, a significant amount of technical and policy framework problems still has to be solved. This paper addresses the problem of predefined path feasibility and proposes an effective methodology for a path to follow re-planning. The proposed methodology is composed of three parts and is based on the Dubins path approach. In order to modify the vector based Dubins path approach and to ensure the path feasibility, the optimisation problem was solved. A cost function with different inequality constraints was formulated. The performance and reliability of the proposed methodology were analysed and evaluated by carrying out an experimental research while using the autonomous test vehicle.

SKAČKAUSKAS P, SOKOLOVSKIJ E. **Poprawa i ocena niezawodności opartej na wektorowej metodzie trajektorii Dubinsa dla korekty trajektorii pojazdów autonomicznych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 549–557, http://dx.doi.org/10.17531/ein.2018.4.6.

Dla zapewnienia rozwoju konkurencyjnego i zrównoważonego systemu transportowego, oraz w celu rozwiązania problemów związanych z bezpieczeństwem ruchu i środowiskiem, poszukiwane są różne rozwiązania techniczne. Można założyć, że autonomiczne pojazdy są technologią, która zapewni pozytywną zmianę w systemie transportowym. Mimo że wiele badań z powodzeniem dotyczyło realizacji autonomicznych pojazdów, należy jeszcze rozwiązać wiele problemów technicznych i prawnych. W niniejszym dokumencie poruszono problem predefiniowanej wykonalności ścieżki i zaproponowano skuteczną metodologię dla ścieżki do śledzenia ponownego planowania. Proponowana metodologia składa się z trzech części i opiera się na metodzie Dubinsa. Aby zmodyfikować metodę trajektorii Dubinsa i zapewnić optymalną trajektorię, w publikacji rozwiązano zadanie optymalizacji. Sformułowana funkcja celu z różnymi nieliniowymi ograniczeniami. Skuteczność i niezawodność proponowanej metodologii została przeanalizowana i oceniona po przeprowadzeniu eksperymentalnych badań z wykorzystaniem autonomicznego pojazdu badawczego.

DUAN R, LIN Y, ZENG Y. **Fault diagnosis for complex systems based on reliability analysis and sensors data considering epistemic uncertainty**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 558–566, http://dx.doi.org/10.17531/ein.2018.4.7.

This paper presents an information fusion method to diagnose system fault based on dynamic fault tree (DFT) analysis and dynamic evidential network (DEN). In the proposed method, firstly, it uses a DFT to describe the dynamic fault characteristics and evaluates the failure rate of components using interval numbers to deal with the epistemic uncertainty. Secondly, qualitative analysis of a DFT is to generate the characteristic function via a traditional zero-suppressed binary decision diagram, while quantitative analysis is to calculate some importance measures by mapping a DFT into a DEN. Thirdly, these reliability results are updated according to sensors data and used to design a novel diagnostic algorithm to optimize system diagnosis. Furthermore, a diagnostic decision tree (DDT) is obtained to guide the maintenance workers to recover the system. Finally, the performance of the proposed method is evaluated by applying it to a train-ground wireless communication system. The results of simulation analysis show the feasibility and effectiveness of this methodology.

DUAN R, LIN Y, ZENG Y. **Diagnozowanie błędów w systemach złożonych na podstawie analizy niezawodności oraz danych z czujników z uwzględnieniem niepewności epistemicznej**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 558–566, http://dx.doi.org/10.17531/ein.2018.4.7.

W artykule przedstawiono metodę fuzji informacji służącą do diagnozowania błędów systemu w oparciu o analizę dynamicznego drzewa błędów (DFT) oraz dynamiczną sieć dowodową (DEN). W proponowanej metodzie, pierwszym krokiem jest wykorzystanie DFT do opisania dynamicznych charakterystyk błędów oraz ocena intensywności uszkodzeń komponentów przy użyciu liczb przedziałowych, która rozwiązuje problem niepewności epistemicznej. Krok drugi stanowi jakościowa analiza DFT, która polega na wygenerowaniu funkcji charakterystycznej za pomocą tradycyjnego binarnego diagramu decyzyjnego typu "zero-suppressed" (w którym zostały wyeliminowane wszystkie węzły, których krawędź „1" prowadzi do liścia „0"), oraz analiza ilościowa polegająca na obliczeniu pewnych miar ważności poprzez odwzorowanie DFT w DEN. W kroku trzecim, otrzymane wyniki niezawodnościowe aktualizuje się zgodnie z danymi z czujników a następnie wykorzystuje do stworzenia nowego algorytmu diagnostycznego do optymalizacji diagnostyki systemu. Powstaje diagnostyczne drzewo decyzyjne (DDT), które stanowi dla pracowników utrzymania ruchu wytyczną w procesie odzyskiwania systemu. Działanie proponowanej metody oceniano poprzez zastosowanie jej do diagnostyki systemu łączności radiowej pociąg–ziemia. Wyniki analizy symulacyjnej wskazują na możliwość praktycznego wykorzystania i skuteczność omawianej metodologii.

SANTANA JMM, SANTIAGO RLV, MOURA MC, LINS ID. **Extended warranty of medical equipment subject to imperfect repairs: an approach based on generalized renewal process and Stackelberg game**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 567–578, http://dx.doi.org/10.17531/ein.2018.4.8.

Due to its advanced technology, maintenance services of healthcare equipment have been commonly executed by the original equipment manufacturer (OEM), which can be characterized as a monopolist. In this context, hospitals require high availability of their equipment at a reasonable servicing cost, whereas OEM aims to maximize its profit by selling extended warranty (EW) services for multiple consumers. The issue of drawing a maintenance contract between OEM and hospitals has already been treated by adopting a Stackelberg's game. However, the "as good as new" and "as bad as old"

SANTANA JMM, SANTIAGO RLV, MOURA MC, LINS ID. **Rozszerzona gwarancja na sprzęt medyczny podlegający niepełnym naprawom: podejście oparte na uogólnionym procesie odnowy i modelu Stackelberga**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 567–578, http://dx.doi.org/10.17531/ein.2018.4.8.

Ze względu na zaawansowanie technologiczne sprzętu medycznego, jego obsługą serwisową zazwyczaj zajmuje się producent sprzętu oryginalnego (OEM), co czyni go monopolistą w tym zakresie. Podczas gdy szpitalom zależy na wysokiej gotowości sprzętu przy rozsądnych kosztach obsługi, OEM dąży do maksymalizacji zysku poprzez sprzedaż rozszerzonej gwarancji na usługi serwisowe wielu klientom. Istnieją już badania, w których kwestię zawierania umowy o świadczenie usług serwisowych między OEM a szpitalami analizowano z zastosowaniem modelu Stackelberga. Jednak zwykle badania te zakładają, assumptions are usually considered, which are rather difficult to observe in practice, especially for healthcare institutions and their technology-intensive equipment. Thus, we here adopt generalized renewal processes (GRP) for modelling imperfect repairs, and we develop a discrete event simulation method for finding the best strategies of each player: OEM sets the prices for EW and on-demand maintenance that optimize its profit, while hospitals choose which option they should hire. We also present an application example with real data gathered from an angiography device, which is used for mapping blood vessels and diagnosing heart diseases.

że stan po naprawie może być albo "jak fabrycznie nowy" albo"jak przed uszkodzeniem", co rzadko spotyka się w praktyce, zwłaszcza w przypadku placówek służby zdrowia i ich zaawansowanego technologicznie sprzętu. W związku z tym, w przedstawionej pracy, przyjęto uogólniony proces odnowy (GRP) do modelowania niepełnych napraw oraz opracowano metodę symulacji zdarzeń dyskretnych w celu znalezienia najlepszych strategii dla każdego gracza: OEM ustala ceny rozszerzonej gwarancji oraz konserwacji na żądanie, tak by zoptymalizować swój zysk; szpital natomiast ustala, którą opcję powinien wybrać. W pracy przedstawiono również przykład zastosowania omawianego podejścia z wykorzystaniem rzeczywistych danych zebranych z angiografu, który służy do obrazowania naczyń krwionośnych i diagnozowania chorób serca.

ZHANG Y, MA Y, OUYANG L, LIU L. **A novel reliability model for multi-component systems subject to multiple dependent competing risks with degradation rate acceleration**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 579–589, http://dx.doi.org/10.17531/ein.2018.4.9.

The purpose of this paper is to establish a new reliability model of the system subject to multiple dependent competing risks. For a system subject to multiple dependent competing risks, the total degradation consists of natural degradation amount and sudden degradation increments (SDIs) caused by random shocks arriving at the system. Most researchers on this topic only focus on the SDIs. However, the impact of random shocks on degradation rate is ignored. In this paper, a novel reliability model considering degradation rate acceleration (DRA) caused by random shocks is proposed, in which the degradation model is based on the degradation path. The dependence relationship between multiple degradation processes is dealt with by copula method, and the arrival time of shocks is assumed to follow a non-homogeneous Poisson process (NHPP). Finally, the effectiveness of the proposed reliability model is demonstrated by an example of a series system. Moreover, the effect of model parameters is evaluated through sensitivity analysis.

ZHANG Y, MA Y, OUYANG L, LIU L. **Nowatorski model niezawodności dla systemów wieloelementowych narażonych na liczne zależne ryzyka konkurujące uwzględniający przyspieszenie tempa degradacji**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 579–589, http://dx.doi.org/10.17531/ein.2018.4.9.

Celem niniejszej pracy było stworzenie nowego modelu niezawodności systemu narażonego na liczne zależne ryzyka konkurujące. W przypadku systemu eksponowanego na wiele zależnych ryzyk konkurujących, na wartość całkowitą degradacji składa się wartość degradacji naturalnej oraz wartość nagłych przyrostów degradacji (sudden degradation increments, SDI) powodowanych przez losowe zaburzenia systemu. Większość badaczy tej tematyki koncentruje się wyłącznie na SDI, ignorując tym samym wpływ zaburzeń losowych na tempo degradacji. W niniejszym artykule zaproponowano nowy model niezawodności uwzględniający przyspieszenie tempa degradacji powodowane zaburzeniami losowymi, w którym model degradacji opiera się na krzywej degradacji. Zależność między mnogimi procesami degradacji rozpatrywano za pomocą metody funkcji kopuły przy założeniu, że czas wystąpienia zaburzenia odpowiada niejednorodnemu procesowi Poissona. Skuteczność proponowanego modelu niezawodności zademonstrowano na przykładzie systemu szeregowego. Ponadto, wykorzystano analizę czułości do oceny wpływu parametrów modelu na niezawodność systemu.

TANG D, SHENG W, YU J. **Dynamic condition-based maintenance policy for degrading systems described by a random-coefficient autoregressive model: A comparative study**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 590–601, http://dx.doi.org/10.17531/ein.2018.4.10.

In this paper, we optimize a dynamic condition-based maintenance policy for a slowly degrading system subject to soft failure and condition monitoring at equidistant, discrete time epochs. A random-coefficient autoregressive model with time effect is developed to describe the system degradation. The system age, previous state observations, and the item-to-item variability of the degradation are jointly combined in the proposed degradation model. Stochastic behavior for both the age-dependent and the state-dependent term are considered, and a Bayesian approach for periodically updating the estimates of the stochastic coefficients is developed to combine information from a degradation database with real-time condition-monitoring information. Based on this degradation model, the dynamic maintenance policy is formulated and solved in a semi-Markov decision process framework. Incorporated with the same semi-Markov decision process framework is a novel approach for mean residual life estimation, which enables simultaneous residual life estimation with the optimization procedure. The effectiveness of using the proposed random-coefficient autoregressive model with time effect rather than the existing fixed-coefficient ones to describe system degradation is demonstrated through a comparative study based on a real degradation dataset. The advantages of using a dynamic maintenance policy are also revealed.

TANG D, SHENG W, YU J. **Dynamiczna strategia utrzymania ruchu na podstawie stanu technicznego dla ulegających degradacji systemów opisanych modelem autoregresyjnym z parametrami losowymi – studium porównawcze**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 590–601, http://dx.doi.org/10.17531/ein.2018.4.10.

W prezentowanej pracy dokonano optymalizacji dynamicznej, uwzględniającej stan techniczny obiektu strategii utrzymania ruchu dla wolno ulegającego degradacji systemu monitorowanego w równoodległych dyskretnych chwilach czasu (epokach) pod względem uszkodzeń parametrycznych oraz stanu technicznego. Do opisu degradacji systemu opracowano model autoregresyjny z parametrami losowymi uwzględniający wpływ czasu. Proponowany model degradacji bierze pod uwagę zarówno wiek systemu jak i wcześniejsze obserwacje stanu oraz zmienność degradacji pomiędzy obiektami. Rozważano zachowanie stochastyczne zarówno składnika zależnego od wieku jak i składnika zależnego od stanu; opracowano bayesowską metodę okresowej aktualizacji oszacowań współczynników stochastycznych, która pozwala łączyć informacje z bazy danych o degradacji z informacjami z monitorowania stanu w czasie rzeczywistym. W oparciu o otrzymany model degradacji, sformułowano dynamiczną politykę utrzymania ruchu; problem optymalizacji tej polityki rozwiązywano w ramach procesu decyzyjnego semi-Markowa. Do procesu decyzyjnego włączono nowatorską metodę obliczania trwałości resztkowej, co umożliwiło ocenę trwałości resztkowej jednocześnie z przeprowadzeniem procedury optymalizacyjnej. Skuteczność wykorzystania proponowanego modelu autoregresyjnego do opisu degradacji systemu porównywano ze skutecznością dotychczasowych modeli z parametrami stałymi w badaniu opartym na rzeczywistym zbiorze danych o degradacji. Wskazano również zalety stosowania proponowanej dynamicznej strategii utrzymania ruchu.

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, http://dx.doi.org/10.17531/ein.2018.4.11.

Maintaining high efficiency of using the fleet of public mass transport vehicles puts many challenges ahead of the operator. Among them, when planning periodic operational activities, the operator should take into account the assessment of possible unexpected vehicle failures and the costs of their removal under the so-called corrective maintenance. Due to the random nature of vehicle breakdowns, knowledge about stochastic processes is necessary to maintain their efficient and safe operation. The research problem formulated in the title meets these needs. Therefore, the costs of corrective maintenance of vehicles are modelled, i.e. the costs that are not included in the scheduled maintenance costs and are not related to preventive maintenance. The costs of corrective maintenance and the costs of replacement of damaged parts are unexpectedly created at random moments of operating means of transport, usually between scheduled maintenance. Due to the variety of failure processes of individual parts of the vehicle, the methods and applications of stochastic modelling for simple failures modelled by the Poisson process are presented in this paper. The basis for the application of the presented methods is to identify those parts of the vehicle that are damaged in accordance with this process. The assessment of parameters of failure

ANDRZEJCZAK K, MŁYŃCZAK M, SELECH J. **Poissonowskie strumienie uszkodzeń w prognozowaniu kosztów obsług korekcyjnych floty pojazdów**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 602–609, http://dx.doi.org/10.17531/ein.2018.4.11.

Utrzymanie wysokiej efektywności użytkowania floty pojazdów publicznego transportu masowego stawia przed operatorem wiele wyzwań. Wśród nich planując okresowe działania eksploatacyjne operator powinien uwzględniać oceny możliwości pojawiania się niespodziewanych awarii pojazdów oraz kosztów ich usuwania w ramach tak zwanych obsług korekcyjnych. Ze względu na losowy charakter awarii pojazdów dla utrzymania efektywnej i bezpiecznej ich eksploatacji niezbędna okazuje się wiedza dotycząca procesów stochastycznych. Sformułowany w tytule problem badawczy wychodzi naprzeciw tym potrzebom. Modelowane są więc koszty obsług korekcyjnych pojazdów, tj. koszty, które nie są uwzględniane w ich przeglądach planowych i nie obciążają prewencyjnych napraw. Koszty robót korekcyjnych i koszty wymiany uszkodzonych części powstają niespodziewanie w losowych chwilach użytkowania środków transportu, zwykle między okresowymi przeglądami. W niniejszej pracy, ze względu na różnorodność procesów uszkadzania poszczególnych części pojazdu, przedstawiono metody i zastosowania stochastycznego modelowania dla prostych strumieni uszkodzeń, których modelem jest proces Poissona. Podstawą zastosowania przedstawionych metod jest zidentyfikowanie tych części pojazdu, które uszkadzają się zgodnie z tym procesem. Oceny parametrów processes of individual vehicle parts is made on the basis of the operational database of a homogeneous fleet of vehicles operated for 5 years. The operational database is dynamically updated with new events. On the basis of actual data on corrective maintenance of a distinguished group of damaged parts of vehicles, the possibilities and limitations of practical applications of the Poisson process to assess the risk of incurring costs in the further process of fleet operation were shown.

LI J, DUAN R. **Dynamic diagnostic strategy based on reliability analysis and distance-based VIKOR with heterogeneous information**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 610–620, http://dx.doi.org/10.17531/ein.2018.4.12.

This paper presents a dynamic diagnostic strategy based on reliability analysis and distance-based VIKOR with heterogeneous information. Specifically, the proposed method uses a dynamic fault tree (DFT) to describe the dynamic fault characteristics and evaluates the failure rate of components using interval numbers to deal with the epistemic uncertainty. Furthermore, DFT is mapped into a dynamic evidential network (DEN) to calculate some reliability parameters and these parameters together with test cost constitute a decision matrix. In addition, a dynamic diagnostic strategy is developed based on an improved VIKOR algorithm and the previous diagnosis result. This diagnosis algorithm determines the weights of attributes based on the Entropy concept to avoid experts' subjectivity and obtains the optimal ranking directly on the original heterogeneous information without a transformation process, which can improve diagnosis efficiency and reduce information loss. Finally, the performance of the proposed method is evaluated by applying it to a train-ground wireless communication system. The results of simulation analysis show the feasibility and effectiveness of this methodology.

KALUĐER S, FEKETE K, JOZSA L, KLAIĆ Z. **Fault diagnosis and identification in the distribution network using the fuzzy expert system**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 621–629, http://dx.doi.org/10.17531/ein.2018.4.13.

In this paper, a fuzzy expert off-line system has been developed for fault diagnosis in the distribution network based on the structural and functional operation of the relay and circuit breakers. Functional operations (correct operation, false operation and failure to operate) of the relays and circuit breakers are described by fuzzy logic. Input data for the proposed fuzzy expert fault diagnosis system (FDS) are status and time stamps of the alarms, associated with relays and circuit breakers. The diagnostic system from a huge number of alarms sets, logically organizes and quantifies the diagnosis. FDS can diagnose correct operation, false operation and failure to operate of the relays and circuit breakers. Also, it can identify and quantify fault location based on the Hamacher's operator of a fuzzy union. The additional contribution of this paper is in modeling unknown information using linear fuzzy membership function. Statuses of certain components may be unknown due to telemetry failures or are simply unavailable to the operator and proposed FDS can make diagnosis in such a situation. Developed fuzzy expert FDS is tested on the two examples of faults in real life distribution network.

FIEDKIEWICZ Ł, PIELECHA I. **Assessment of possible use of the ionization signal for the combustion process diagnostics in a spark ignition combustion engine powered by natural gas**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 630–637, http://dx.doi.org/10.17531/ein.2018.4.14.

The ionization signal, which is a result the presence of ions and electrons in the cylinder space of the internal combustion engine, is affected by many factors, including: temperature, pressure, fuel mixture composition, fuel type, presence of exhaust gases and others. The shape of the signal changes to a large extent from cycle to cycle, which indicates the stochastics of the combustion process. Nevertheless, its analysis provides a lot of useful information, such as the location of the maximum pressure or the maximum heat release rate. Using these signals allows supplementing the limited engine control systems of the combustion process in internal combustion engines. The paper presents a comparative analysis of the gas ionization current signal in the cylinder and the variable pressure at fixed operating points of a single-cylinder, four-stroke engine powered by natural gas. The analysis allowed to determine the relationship between the positions of the maximum thermal ionization signal value and of the maximum combustion pressure value. Additionally the relationship between the position of the maximum thermal fraction derivative and the maximum heat release rate was established.

procesów uszkodzeń poszczególnych części pojazdów dokonywane są na podstawie bazy danych eksploatacyjnych jednorodnej floty pojazdów eksploatowanych przez 5 lat. Wraz z upływem czasu baza danych eksploatacyjnych jest dynamicznie uzupełniana o nowe zdarzenia. Na podstawie danych rzeczywistych o obsługach korekcyjnych wyróżnionej grupy uszkadzających się części pojazdów pokazano możliwości i ograniczenia praktycznych zastosowań procesu Poissona do oceny ryzyka poniesienia kosztów w dalszym procesie eksploatacji floty.

LI J, DUAN R. **Dynamiczna strategia diagnostyczna z wykorzystaniem informacji heterogenicznych bazująca na analizie niezawodności i opartym na odległościach algorytmie VIKOR**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 610–620, http://dx.doi.org/10.17531/ein.2018.4.12.

W artykule przedstawiono dynamiczną strategię diagnostyczną, w której wykorzystuje się oryginalne informacje heterogeniczne. Metoda ta bazuje na analizie niezawodności i opartym na odległościach algorytmie VIKOR. Dokładniej, przedstawiona strategia polega na wykorzystaniu dynamicznego drzewa błędów (DFT) do opisu dynamicznych charakterystyk błędów oraz ocenie intensywności uszkodzeń komponentów przy użyciu liczb przedziałowych, co pozwala rozwiązać problem niepewności epistemicznej. Ponadto, w proponowanej metodzie, DFT zostaje odwzorowane w dynamiczną sieć dowodową (DEN) w celu obliczenia niektórych parametrów niezawodności, a parametry te wraz z kosztem badań diagnostycznych tworzą matrycę decyzyjną. Opracowana dynamiczna strategia diagnostyczna opiera się na udoskonalonym algorytmie diagnostycznym VIKOR oraz wynikach wcześniejszej diagnostyki. Algorytm VIKOR określa wagi atrybutów w oparciu o koncepcję Entropii, co pozwala wyeliminować subiektywność oceny eksperckiej i ustalić optymalną kolejność działań diagnostycznych bazując bezpośrednio na oryginalnych informacjach heterogenicznych bez konieczności ich transformacji, co może poprawić efektywność diagnozy i zmniejszyć utratę informacji. Działanie proponowanej metody oceniano poprzez zastosowanie jej do diagnostyki systemu łączności radiowej pociąg–ziemia. Wyniki analizy symulacyjnej wskazują na możliwość praktycznego wykorzystania i skuteczność omawianej metodologii.

KALUĐER S, FEKETE K, JOZSA L, KLAIĆ Z. **Identyfikacja i diagnoza błędów w elektroenergetycznej sieci rozdzielczej z wykorzystaniem rozmytego systemu eksperckiego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 621–629, http://dx.doi.org/10.17531/ein.2018.4.13.

W prezentowanym artykule opracowano rozmyty system ekspercki typu off-line do diagnozowania błędów w elektroenergetycznej sieci rozdzielczej. System bazuje na strukturze i działaniu przekaźnika i wyłączników automatycznych. Działanie (prawidłowe działanie, błędne działanie i brak działania) przekaźników i wyłączników opisano za pomocą logiki rozmytej. Dane wejściowe do proponowanego rozmytego eksperckiego systemu diagnostyki błędów (FDS) stanowią stany i sygnatury czasowe alarmów, związane z przekaźnikami i wyłącznikami. System diagnostyczny logicznie porządkuje i określa ilościowo diagnozę na podstawie ogromnej liczby zestawów alarmów. FDS pozwala zdiagnozować prawidłowe działanie, błędne działanie oraz awarię (brak działania) przekaźników i wyłączników. Ponadto umożliwia identyfikację i lokalizację błędów w oparciu o sumę Hamachera. W artykule dodatkowo omówiono metodę modelowania informacji nieznanych przy użyciu liniowej funkcji przynależności dla zbiorów rozmytych. Stany niektórych elementów mogą być nieznane z powodu awarii telemetrii lub mogą być po prostu niedostępne dla operatora. Proponowany FDS umożliwia postawienie diagnozy w takich sytuacjach. Opracowany rozmyty ekspercki FDS testowano na dwóch przykładach błędów powstałych w funkcjonującej sieci rozdzielczej.

FIEDKIEWICZ Ł, PIELECHA I. **Ocena możliwości wykorzystania sygnału jonizacji do diagnostyki procesu spalania w silniku spalinowym o zapłonie iskrowym zasilanym gazem ziemnym**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 630–637, http://dx.doi.org/10.17531/ein.2018.4.14.

Sygnał jonizacji wynikający z obecności jonów oraz elektronów w przestrzeni cylindra silnika spalinowego jest składową wielu czynników, między innymi: temperatury, ciśnienia, składu mieszanki, rodzaju paliwa, obecności reszty spalin oraz innych. Kształt sygnału zmienia się w znacznym stopniu z cyklu na cykl, co świadczy o stochastyce procesu spalania. Mimo tego, jego analiza dostarcza wielu przydatnych informacji, takich jak położenie maksymalnego ciśnienia czy maksymalnej szybkości wywiązywania się ciepła. Ich wykorzystanie pozwala uzupełnić ograniczone systemy kontroli procesu spalania w silnikach spalinowych. W artykule przedstawiono analizę porównawczą sygnału prądu jonizacji gazów w cylindrze oraz ciśnienia szybkozmiennego przy ustalonych punktach pracy jednocylindrowego, czterosuwowego silnika zasilanego gazem ziemnym. W wyniku analizy uzyskano zależność położenia maksymalnej wartości sygnału jonizacji termicznej od położenia maksymalnej wartości ciśnienia spalania, uzależniono również położenie maksimum pochodnej członu termicznego od położenia maksimum szybkości wywiązywania się ciepła.

BOGDEVICIUS M, ZYGIENE R. **Research of dynamic processes of the system "Vehicle – Track" using the new method of vehicle wheel with metal scale**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 638–649, http://dx.doi.org/10.17531/ein.2018.4.15.

Mathematical models of vehicle wheel with metal scales are introduced in this article. When analysing the interaction between vehicle wheel with a metal scale and rail in the system "Vehicle – Track", the changes of the kinematic and dynamic parameters of the wheel and rail contact points in time are examined, depending on the height of the 2 mm metal scale, when the length of the metal scale is 100 mm and the speed of movement is V= 40 - 100 km/h. The results obtained after the research of the system "Vehicle – Track", when the wheel has a metal scale, help to better understand and evaluate the impact of metal scale on wheel on dynamic loads of rail and vehicle and the regularities of their movement. The appearance of a metal scale on the wheel's surface causes technical and maintenances problems for the rolling stock. Railway standards limit the speed of movement that depends on a certain size of metal scale.

BOGDEVICIUS M, ZYGIENE R. **Badanie dynamicznych procesów zachodzących w układzie „pojazd-tor" z wykorzystaniem nowej metody dla kół z metalową łuską**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 638–649, http://dx.doi.org/10.17531/ein.2018.4.15.

W niniejszym artykule przedstawiono modele matematyczne koła pojazdu szynowego z powstałą w wyniku zużycia metalową łuską. Analizując oddziaływania pomiędzy kołem pojazdu z łuską a szyną w układzie "pojazd–tor", badano zmiany kinematycznych i dynamicznych parametrów punktów kontaktu koła z szyną zachodzące w czasie, w zależności od wysokości metalowej łuski (2 mm), przy długości łuski 100 mm i zakresie prędkości ruchu pojazdu V = 40–100 km/h. Wyniki uzyskane w badaniu układu "pojazd–tor" dla kół na powierzchni których powstała metalowa łuska, umożliwiają lepsze zrozumienie oraz ocenę wpływu łuski na dynamiczne obciążenia szyny i pojazdu oraz prawidłowości ruchu pojazdu. Pojawienie się metalowej łuski na powierzchni koła powoduje problemy techniczne i obsługowe w utrzymaniu ruchu taboru kolejowego. Normy kolejowe ograniczają prędkość ruchu pojazdów szynowych, uzależniając ją od rozmiaru łuski.

JASIULEWICZ-KACZMAREK M, ŻYWICA P. **The concept of maintenance sustainability performance assessment by integrating balanced scorecard with non-additive fuzzy integral**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 650–661, http://dx.doi.org/10.17531/ein.2018.4.16.

In response to the growing sustainability concerns, manufacturing companies have to formulate measures to assess sustainable manufacturing performance, aiming at integration of sustainability aspects. Although various models and methods to assess the sustainability of production processes, and point the role of maintenance have been developed in recent years, contribution of all the elements of the maintenance to the results of sustainable production has not been comprehensively considered, since mostly financial aspects were analyzed. Taking into account this research gap, the article presents the concept of a model and procedure for assessing maintenance from the perspective of sustainable manufacturing requirements. Authors integrate three sustainability dimensions (economic, social and environmental) with Kaplan and Norton's balance scorecard perspectives as a basis to develop the model of maintenance sustainability performance assessment. For the model developed, the assessment procedure based on the paradigm of aggregate assessment was designed. The Choquet integral, based on the so-called λ – measure, was implemented to aggregate the measures. Then, the results of research on determining the importance and interactions between the perspectives and criteria for assessing sustainable maintenance in enterprises representing the automotive and food industries are presented.

JASIULEWICZ-KACZMAREK M, ŻYWICA P. **Koncepcja oceny zrównoważonego utrzymania ruchu z zastosowaniem zrównoważonej karty wyników i nie-addytywnej całki rozmytej**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 650–661, http://dx.doi.org/10.17531/ein.2018.4.16.

W odpowiedzi na wyzwania zrównoważonego rozwoju (SD), przedsiębiorstwa produkcyjne włączają ekonomiczne, środowiskowe i społeczne wymagania SD do swoich praktyk produkcyjnych i formułują miary do oceny skuteczności podjętych działań. Pomimo, iż w ostatnich latach opracowano wiele modeli i metod oceny zrównoważonej produkcji i wskazywano w nich na rolę utrzymania ruchu, to jednak poza aspektem finansowym nie rozważano w sposób kompleksowy wszystkich elementów wkładu utrzymania ruchu w wyniki zrównoważonej produkcji. Biorąc pod uwagę tę lukę badawczą, w artykule przedstawiono koncepcję modelu i procedury oceny utrzymania ruchu z perspektywy wymagań zrównoważonej produkcji. Autorzy integrują trzy wymiary zrównoważonego rozwoju (ekonomiczny, społeczny i środowiskowy) z perspektywami zrównoważonej karty wyników Kaplana i Nortona, jako podstawę do skonstruowania modelu oceny wyników zrównoważonego utrzymania ruchu. Dla tak opracowanego modelu zaprojektowano opartą na paradygmacie oceny agregatowej procedurę oceniania. Do agregacji składników oceny zastosowano całkę Choqueta, opartą na tzw. mierze λ. Następnie przedstawiono wyniki badań pilotażowych dotyczących określenia ważności i interakcji między perspektywami i kryteriami oceny zrównoważonego utrzymania ruchu w przedsiębiorstwach branży motoryzacyjnej i spożywczej.

WANG C, XU J, WANG H, ZHANG Z. **A criticality importance-based spare ordering policy for multi-component degraded systems**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 662–670, http://dx.doi.org/10.17531/ein.2018.4.17.

With the increasing complexity and variety of production systems, more attention is being paid to preventive replacement on multi-component systems. Each component is non-identical and has its own degradation process. In this paper, we propose a criticality importance-based spare ordering policy for a complex system, which consists of multiple series-parallel degrading components. Replacement action is triggered whenever the system reliability drops below a lower threshold and spares for replacement are available. Our policy mainly consists of two steps: (1) determine which components to be replaced; (2) determine when to order spares for components selected. In step 1, when the replacement action is triggered, we select components that most need to be replaced within the system in accordance with the optimum ranking of components until the system meets an upper reliability threshold. In step 2, a spare ordering policy for components selected is made and the optimal spare ordering time is obtained by minimizing the expected replacement cost during the once replacement cycle. Finally, a numerical example is given to illustrate the proposed multi-spare ordering policy. Moreover, the proposed policy is of significance for safety-critical systems such as substation automation system, bridge system, nuclear power plants and aerospace equipment.

WANG C, XU J, WANG H, ZHANG Z. **Oparta na kryterium krytyczności polityka zamawiania części zamiennych do zdegradowanych systemów wieloelementowych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 662–670, http://dx.doi.org/10.17531/ein.2018.4.17.

Wraz ze wzrostem złożoności i różnorodności systemów produkcyjnych, coraz większą uwagę zwraca się na wymianę zapobiegawczą w systemach wieloelementowych. Każdy element takiego systemu jest nieidentyczny z pozostałymi elementami i charakteryzuje się własnym procesem degradacji. W niniejszym artykule proponujemy strategię zamawiania elementów zamiennych dla systemu złożonego składającego się z wielu ulegających degradacji komponentów tworzących strukturę szeregowo-równoległą. Omawiana strategia wymiany opiera się na kryterium krytyczności elementów. Akcja wymiany uruchamiana jest za każdym razem, gdy niezawodność systemu spada poniżej dolnego progu i dostępne są części zamienne. Na proponowaną strategię składają się zasadniczo dwa etapy: (1) określenie elementów wymagających wymiany oraz (2) określenie terminu zamówienia części zamiennych do wybranych elementów. W 1. etapie, po uruchomieniu akcji wymiany, wybiera się komponenty systemu, które najpilniej wymagają wymiany, kierując się optymalnym rankingiem komponentów, do momentu aż system osiągnie górny próg niezawodności. W 2. etapie, opracowuje się politykę zamawiania części zamiennych dla wybranych komponentów oraz określa się optymalny czas zamawiania części zamiennych poprzez minimalizację oczekiwanego kosztu wymiany podczas jednego cyklu wymiany. W artykule przedstawiono przykład numeryczny, który ilustruje proponowaną strategię jednoczesnego zamawiania wielu części zamiennych. Proponowana strategia może znaleźć zastosowanie w systemach o kluczowym znaczeniu dla bezpieczeństwa, takich jak systemy automatyki podstacji, systemy mostowe, elektrownie jądrowe i sprzęt lotniczy.

IDZIASZEK Z. **Method of analysis of productivity with an innovative model of the working capability of the object in the body ($\mathbb{C}$) for the new resource allocation on inherent and non-inherent**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 671–681, http://dx.doi.org/10.17531/ein.2018.4.18.

The aim of the article is to develop new methods of analysis, estimation and optimal selection of quantitative resources (inherent and non-inherent) in the planning of the product effect for specific environmental conditions. The required iterative approach in the construction of the mathematical model and analysis of its possible practical applications and search for how to figure those opportunities. As the testing method has been applied method intuitive, allowing you to use the experience of expert ana-

IDZIASZEK Z. **Metoda analizy produktywności z innowacyjnym modelem potencjału roboczego obiektu w ciele $\mathbb{C}$ dla nowego podziału zasobów na inherentne i nieinherentne**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 671–681, http://dx.doi.org/10.17531/ein.2018.4.18.

Celem artykułu jest opracowanie nowej metody analizy, szacowania i optymalnego doboru ilościowego zasobów (inherentnych i nieinherentnych) w planowaniu efektu produktowego w określonych warunkach środowiskowych. Realizacja celu wymagała iteracyjnego podejścia przy budowie modelu matematycznego i analizie możliwych jego zastosowań praktycznych oraz poszukiwaniu sposobu ilustracji tych możliwości. Jako metoda badawcza została zastosowana metoda intuicyjna, pozwalająca wykorzystać doświadczenie eksperckie z realizowanych analiz możliwości pełnego wykorzystania trwałości obiektów

lysis from ongoing opportunities to make full use of the sustainability properties and customize to their processes. The results were presented in the form of mathematical models in the collection of complex numbers and graphically on the plane of complex numbers. Method to estimate changes inherent and non-inherent resources objects (machines, systems, organizations) on their productivity (P_o). The method uses the original, innovative, model potential workspace object (P_rO) in the form of a complex binding numerically inherent (Z_iO) and non-inherent ($Z_{ni}O$) resources objects. Evaluation of value Po it was proposed with the P_rO. The values of the Z_iO and $Z_{ni}O$ was adopted as two independent resources constituting the whole of resources in the required in the production (or in the service). Method evaluation Po illustrates for the resources object described model $R_o = |P_rO = f(Z_pO, Z_oO)|$, where Z_iO is a work resource (Z_pO), $Z_{ni}O$ is extracted from the operation of the resource service (Z_oO), and the generating capacity of the object P_o is described using a pointer named R object (R_o). Illustrated in the complex plane analysis results and the results obtained from the calculation P_rO and R_o, for contract values of the Z_pO and Z_oO, indicate the application capabilities developed method. Method allows a very clear description of the productivity changes objects (or processes, or production organization), in the context of the selection of manufacturing resource structure, through the separation of the factors causing these changes. Method can be adapted for optimal production costs (or services) through design changes object and/or design changes of the process exploitation. Developed the method brings new opportunities for theoretical and application in relation technical and economic sciences.

i dostosowywania do tego ich procesów eksploatacji. Wyniki zostały zaprezentowane w postaci modeli matematycznych w zbiorze liczb zespolonych i graficznie na płaszczyźnie liczb zespolonych. Metoda umożliwia szacowanie zmian inherentnych i nieinherentnych zasobów obiektów (maszyn, systemów, organizacji) na ich produktywność (P_o). W metodzie wykorzystano autorski, innowacyjny, model potencjału roboczego obiektu (P_rO) w postaci liczby zespolonej wiążącej liczbowo inherentne (Z_iO) i nieinherentne ($Z_{ni}O$) zasoby obiektu. Wyznaczanie wartości P_o zaproponowano z modułu P_rO. Wartości Z_iO i $Z_{ni}O$ przyjęto jako dwa niezależne od siebie zasoby stanowiące całość zasobów w realizacji danej produkcji lub usługi. Metodę oceny P_o zilustrowano dla zasobów obiektu opisanych modelem $R_o = |P_rO = f(Z_pO, Z_oO)|$, gdzie Z_iO to zasób pracy obiektu (Z_pO), $Z_{ni}O$ to wyodrębniony z eksploatacji zasób obsług (Z_oO), a zdolności wytwórcze obiektu P_o opisano za pomocą wskaźnika nazwanego resursem obiektu (R_o). Zilustrowane na płaszczyźnie zespolonej wyniki analiz i uzyskane wyniki z obliczeń P_rO i R_o, dla umownych wartości Z_pO i Z_oO, wskazują na duże możliwości aplikacyjne opracowanej metody. Metoda umożliwia bardzo czytelny opis zmian produktywności obiektów\procesów\organizacji, w kontekście doboru struktury zasobów wytwórczych, poprzez rozdzielenie czynników powodujących te zmiany. Metodę można adaptować na potrzeby optymalizacji kosztów produkcji\usług poprzez zmiany projektowe obiektu technicznego i\lub zmiany projektowe procesu jego eksploatacji. Opracowana metoda wnosi nowe możliwości teoretyczne oraz aplikacyjne w powiązaniu nauk technicznych i ekonomicznych

Hryciów Z, Krasoń W, Wysocki J. **The experimental tests on the friction coefficient between the leaves of the multi-leaf spring considering a condition of the friction surfaces**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 682–688, http://dx.doi.org/10.17531/ein.2018.4.19.

In this study are presented the results of the simulation tests of the friction pairs occurring between the spring leaves while considering a condition of the mating surfaces and an impact of the velocity of their mutual dislocation on the values of the friction coefficients. It has been proposed a methodology in respect of a determination of the coefficients of the static and kinetic friction. Two kinds of the specimens have been prepared for the tests, which have been cut out from a spring leaf of the prototype spring – they have created the model friction pairs. The condition of the specimen surface and their selected mechanical properties have been evaluated. During the experimental tests have been considered: four sliding velocities, four variants of the surface conditions and two values of the normal load. The tests of the friction pairs have been performed at the laboratory stand for measuring the friction force. The results of the tests have been presented in a form of the time courses of friction force, graphs and tabular summaries of the friction coefficients. It has been conducted a comparative analysis of the results in order to determine an influence of the test results on the values of the determined friction coefficients. The proposed research conditions are approximate to the typical operating conditions of the road vehicles.

Hryciów Z, Krasoń W, Wysocki J. **Badania eksperymentalne współczynnika tarcia pomiędzy piórami resoru wielopiórowego z uwzględnieniem stanu powierzchni ciernych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 682–688, http://dx.doi.org/10.17531/ein.2018.4.19.

W pracy przedstawiono wyniki badań symulacyjnych węzłów tarcia występujących pomiędzy piórami resoru, z uwzględnieniem stanu powierzchni współpracujących oraz wpływu prędkości ich wzajemnego przemieszczania, na wartości współczynników tarcia. Zaproponowano metodykę wyznaczania współczynników statycznych i kinetycznych tarcia. Do badań przygotowano dwa rodzaje próbek, które wycięto z pióra resoru prototypowego - tworzyły one modelowe pary cierne. Oceniono stan powierzchni próbek i wybrane właściwości mechaniczne. W badaniach eksperymentalnych uwzględniono: cztery prędkości poślizgu, cztery warianty stanu powierzchni oraz dwie wartości obciążenia normalnego. Badania par ciernych wykonano na stanowisku laboratoryjnym do pomiaru siły tarcia. Wyniki badań przedstawiono w postaci przebiegów czasowych siły tarcia, wykresów i zestawień tabelarycznych współczynników tarcia. Wykonano analizę porównawczą wyników w celu określenia wpływu warunków badań, na wartości wyznaczonych współczynników tarcia. Zaproponowane warunki badań są zbliżone do typowych warunków eksploatacyjnych pojazdów drogowych.

Gonera J, Napiórkowski J. **Model for forecasting the geometry of the floor panel of a passenger car during its operation**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 689–695, http://dx.doi.org/10.17531/ein.2018.4.20.

A number of vehicle users pay attention to the impact of changes in the car body geometry during long-term use on the safety level. However, this issue has not been properly dealt with in research studies. The aim of this study was to identify changes in the floor panel, to develop a model to forecast the geometry during the car use and to identify the points which undergo the maximum displacement. The paper presents the effect of the car mileage on the floor panel condition, taking into account variable environmental factors. In the course of the study, the position of points fixing the front suspension, front bench and rear suspension was determined, as was the position of points situated on parts of the load bearing structure of the car body. The results were used to develop a model for forecasting changes of the floor panel geometry during car use. The probability of changes in the floor panel geometry was found to increase with the mileage. The probability of reaching the maximum permissible geometric changes (3 mm) in a floor panel is accurately described by the probabilistic model in the form of the Rayleigh distribution. Diverse models of the floor panel geometry changes were obtained depending on the environmental conditions and type of the base points under analysis.

Gonera J, Napiórkowski J. **Model prognozowania stanu geometrii płyty podłogowej samochodu osobowego w toku eksploatacji**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 689–695, http://dx.doi.org/10.17531/ein.2018.4.20.

Wielu użytkowników samochodów osobowych zwraca uwagę na istotność wpływu na poziom bezpieczeństwa zmian geometrii nadwozia pojazdów podczas ich wieloletniej eksploatacji. Jednak dotychczas zagadnienie to nie znalazło odpowiedniego odzwierciedlenia w literaturze. Celem pracy była identyfikacja zmian geometrii płyty podłogowej, opracowanie modelu prognozującego stan geometrii w toku eksploatacji i zidentyfikowanie punktów ulegającym największym przemieszczeniom. W pracy przedstawiono wpływ przebiegu pojazdu na stan geometrii płyty podłogowej z uwzględnieniem zróżnicowanych warunków środowiskowych. Podczas badań określano położenie punktów mocujących zawieszenie przednie, przednią ławę i zawieszenie tylne oraz położenie punktów znajdujących się na elementach struktury nośnej nadwozia. Na podstawie uzyskanych wyników opracowano model prognozowania zmian geometrii płyty podłogowej w toku eksploatacji. Stwierdzono, że prawdopodobieństwo zmian geometrii płyty podłogowej podczas eksploatacji rośnie w czasie, wraz ze wzrostem przebiegu. Prawdopodobieństwo osiągnięcia stanu dopuszczalnego (3 mm) zmian geometrycznych na płycie podłogowej dobrze opisuje model probabilistyczny w postaci rozkładu Rayleigha. Uzyskano zróżnicowane modele zmiany geometrii płyty podłogowej w zależności od warunków środowiskowych oraz rodzaju analizowanych punktów bazowych.

Article citation info:
KLUZ R, KUBIT A, SĘP J, TRZEPIECINSKI T. Effect of temperature variation on repeatability positioning of a robot when assembling parts with cylindrical surfaces. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 503–513, http://dx.doi.org/10.17531/ein.2018.4.1.

Rafał KLUZ
Andrzej KUBIT
Jarosław SĘP
Tomasz TRZEPIECINSKI

EFFECT OF TEMPERATURE VARIATION ON REPEATABILITY POSITIONING OF A ROBOT WHEN ASSEMBLING PARTS WITH CYLINDRICAL SURFACES

WPŁYW ZMIAN TEMPERATURY NA POWTARZALNOŚĆ POZYCJONOWANIA ROBOTA PRZY MONTAŻU CZĘŚCI O POWIERZCHNIACH CYLINDRYCZNYCH*

In this paper, the effect of the errors induced by temperature changes on the repeatability positioning error of an industrial robot is analysed. It has been shown that after the stabilization of the thermal conditions, these errors can be identified with the systematic errors. It has also been shown that if the ambient temperature cannot be sufficiently stabilized, the temperature errors can be described using a normal or uniform probability distribution. Depending on the choice of a point in the robot's workspace and temperature fluctuations, these errors can comprise a small share of the total error of the robot. Then the total repeatability positioning error can be approximated with sufficient accuracy by a normal probability distribution or it can comprise the dominant component of this error. In this case, the total error can be approximated using a flat normal distribution. It has been shown that, depending on the choice of location in the workspace in which the assembly operation is carried out, it is possible to obtain both different probabilities of assembling the parts correctly and a different effect of errors caused by slight temperature changes on the value of those probabilities. The results found indicate the potential possibility of increasing the reliability of the process by proposing the selection of the location in the robot workspace which has the lowest sensitivity to errors ascribed to changes in temperature.

Keywords: *repeatability positioning, robot's workspace, temperature, probability of parts joining.*

W niniejszej pracy przeanalizowano wpływ zmiany temperatury na błąd powtarzalności pozycjonowania robota przemysłowego. Wykazano, że po ustabilizowaniu się warunków termicznych błędy te można sklasyfikować jako błędy systematyczne. Wykazano również, że jeżeli w trakcie eksploatacji zrobotyzowanego stanowiska montażowego temperatura otoczenia nie może być wystarczająco ustabilizowana, błędy temperatury można opisać za pomocą jednostajnego rozkładu prawdopodobieństwa i w ten sposób uwzględnić w strukturze całkowitego błędu powtarzalności pozycjonowania. Błędy te na ogół stanowią niewielki udział w całkowitym błędzie robota, wówczas całkowity błąd powtarzalności pozycjonowania robota z dostateczną dokładnością można aproksymować normalnym rozkładem prawdopodobieństwa. W przeciwnym przypadku błąd ten może być przybliżony przy użyciu rozkładu płasko-normalnego. Wykazano że w zależności od wyboru miejsca realizacji zabiegu montażowego w przestrzeni stanowiska można uzyskać zarówno odmienne wartości prawdopodobieństwa połączenia części jak i różny wpływ błędów wywołanych niewielkimi zmianami temperatury na wartość tego prawdopodobieństwa. Uzyskane wyniki badań wskazują na potencjalne możliwości zwiększenia niezawodności procesu poprzez wybór miejsca w przestrzeni roboczej stanowiska charakteryzującego się najniższą wrażliwością na błędy spowodowane zmianami temperatury.

Słowa kluczowe: *powtarzalność pozycjonowania, przestrzeń robocza robota, temperatura, prawdopodobieństwo połączenia części.*

1. Introduction

An important issue in the field of robotic assembly workstations is the problem of ensuring the required probability of the parts being joined, and thus ensuring the required capability of the assembly process. The usability of the elements to be joined into the assembly units is a certain feature depending on the construction of the element, the method of joining and the construction of the assembly station. This feature can be referred to as mountability. The basic condition for achieving high reliability of the assembly work station is the fulfilment of the mounting condition for all parts that are being connected

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

[14]. In fact, due to the occurrence of accidental and systematic errors, these conditions cannot always be fully fulfilled. Therefore, the ratio of the number of connections assembled to the number of all connections can be taken as the probability of joining parts. From the point of view of work station reliability, it is advantageous to ensure a high probability of joining the parts, because this results in a reduction in the cost of its operation due to a lowering of the downtime [31]. Industrial robots, which are the main pieces of equipment on the stands, are delivered to the user with a very small amount of information regarding their accuracy. Usually the technical documentation only gives information on positioning repeatability.

In general, position error is the result of inaccuracies in the whole robotic system, typically categorized as [2]:

- controller errors due to the resolution of the axis encoder devices,
- algorithmic interpolation errors that take place throughout the movement of the robotic arm,
- kinematic errors, which mainly derive from inaccuracies of the kinematic robot model,
- dynamic errors related to the servo systems, friction, and inertia whilst moving,
- mechanical errors owing to manufacturing imprecision, joint wear, bearing wear, and temperature and flexibility deviations,
- errors due to flexibility of links.

The robots are usually constructed of beam-like links with open kinematic chains. As the number of robot links increases, the structure of the robot is more susceptible to generating errors related to positioning accuracy due to the inaccuracy of the kinematic system, inertia and change in the temperature of the environment. In conditions of unstable operating temperatures, the length of the robot links has a significant impact on the accuracy of the robot. A minimisation of the thermal effect can be obtained by using materials which have low thermal expansion in the robot's construction or by implementing an empirical model of error correction based on the signals from several temperature sensors placed inside the robot arm [30]. To avoid the problem of thermal expansion of the robot links, manufacturers use thermally stable materials, such as fibre-reinforced plastics or use isolated heat sources [24]. The process of manufacturing robot links and kinematic pairs introduces some differences in dimensions. In practice, the actual physical zero position and the physical zero position read by the robot controller is affected by errors.

In many cases, a significant proportion of the factors affecting robot error is subjected to constant change, sometimes accidental, which leads to differences between the mathematical models and real characteristics [12, 15]. Works can be found in the literature which are devoted to reducing robot errors by calibration, using laser and vision sensors placed at the end of the grip of the robot [7, 13, 32], selecting the optimal location in the robot workspace [17] and choosing the proper direction to get to the nominal position [16].

The robot calibration procedure consists of four stages [4]: modelling, measurement, identification and compensation. Slamani et al. [28] performed the positioning analysis of an ABB IRB1600 robot using the FARO laser tracker. The results presented by the authors showed that calibration allows one to reduce the robot error by a factor of three. Zhenhua et al. [33] present an attempt to calibrate a 6-DOF robot, using the MDH model (Modified Denavit-Hartenberg model). The measurements were carried out using a Leica AT901 B tracker. The maximum positioning deviation before calibration was about 2500 μm, while after calibration it was reduced to below 1000 μm. Płaczek and Piszczek [23] indicated that laser trackers (Faro Vantage tracker) form an effective method of determining the accuracy and repeatability of the mapping of the trajectory by the robot being examined. The idea of identifying accuracy errors and their calibration using CCD cameras is presented by Abderrahim et al. [1]. The robot positioning deviation measured before this calibration was 3250 μm, but after calibration it was reduced to 290 μm.

Positioning repeatability is a measure of the robot's ability to return to the same position [18], while accuracy is defined as the robot's ability to move precisely to the desired position in three-dimensional space [18]. Procedures for assessing repeatability and accuracy are set out in international standard EN ISO 9283:2003 [10].

Although many researchers have investigated methods for compensating for the geometrical errors of robots, the error related to thermal deformation has not been discussed in detail in the literature, as was noted by Eastwood and Webb in 2009 [8], and Li and Zhao in 2016 [20]. The precision of robots and machine tools is constantly growing, which requires the taking into account of an increasing number of factors affecting their accuracy. In real production conditions, it is difficult to ensure stable environmental conditions during the operation of a robotic assembly station. Small changes in ambient conditions, in particular temperature changes, are often accidental. Therefore, compensation for their impact requires the use of vision or laser measurement systems, which increases the cost of the station [30]. This article analyses the impact of temperature-induced errors on the structure of the total repeatability positioning error of an industrial robot. Based on the experiments conducted, a methodology of error summation was proposed, which was used to determine the effect of temperature changes on the probability of joining machine parts with cylindrical surfaces and the capability of the process.

2. Kinematic error of robot

During assembly processes, the robot's gripper at any moment should occupy a precise position in space set by programmed joint coordinate values q_i. Any characteristic position of the M point of the gripper (Fig. 1) can be determined, in an accepted stationary coordinate arrangement, by a certain function of the joint coordinates [6]:

$$\begin{aligned} x &= x(q_1, q_2, ..., q_n), \\ y &= y(q_1, q_2, ..., q_n), \\ z &= z(q_1, q_2, ..., q_n) \end{aligned} \tag{1}$$

In reality, the values of the joint coordinates have certain errors Δq_i (i = 1, 2, …, n), which result in deviation of positioning of the piece from the programmed one (e.g., [27]). The measure of the position dispersion or the measure of the real orientation, obtained by the n-fold repetition of motion in the same direction as the position of the set task, is referred to as the repeatability positioning [3].

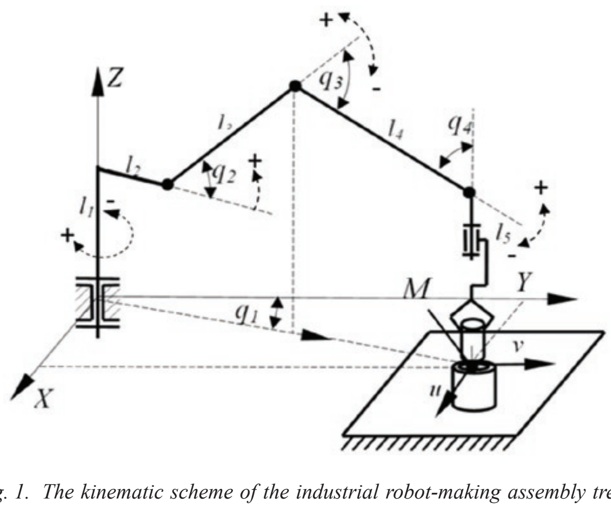


Fig. 1. The kinematic scheme of the industrial robot-making assembly treatment [17]

If we assume that the errors Δq_i of variable stochastic independence q_i relative to their nominal values have a certain given normal distribution and that they are statistically independent, then the repeatability positioning will be a 2-D variable norm, which is a deviation vector of the actual position from the nominal position of the determined parameters in the following manner [3]:

$$\sigma_{xk} = \sqrt{\left(\frac{\partial X}{\partial q_1}\right)^2 \sigma_{q_1}^2 + \left(\frac{\partial X}{\partial q_2}\right)^2 \sigma_{q_2}^2 + \ldots + \left(\frac{\partial X}{\partial q_n}\right)^2 \sigma_{q_n}^2} \quad (2)$$

$$\sigma_{yk} = \sqrt{\left(\frac{\partial Y}{\partial q_1}\right)^2 \sigma_{q_1}^2 + \left(\frac{\partial Y}{\partial q_2}\right)^2 \sigma_{q_2}^2 + \ldots + \left(\frac{\partial Y}{\partial q_n}\right)^2 \sigma_{q_n}^2} \quad (3)$$

$$\sigma_{zk} = \sqrt{\left(\frac{\partial Z}{\partial q_1}\right)^2 \sigma_{q_1}^2 + \left(\frac{\partial Z}{\partial q_2}\right)^2 \sigma_{q_2}^2 + \ldots + \left(\frac{\partial Z}{\partial q_n}\right)^2 \sigma_{q_n}^2} \quad (4)$$

The method and the results of the analysis of the Mitsubishi RV-M2 industrial robot were presented by Kluz and Trzepieciński [17]. The conducted experiments showed that the repeatability positioning error of the robot on the *X*, *Y*, and *Z* axes of the Cartesian coordinate system can be described using a random variable subject to a normal probability distribution with the expected value of 0 and standard deviation σ_k ($N(0, \sigma_k)$).

To determine the repeatability positioning error of the assembly robot, a special measurement stand was used, on which the position of the measuring block, mounted in the pneumatic gripper of the robot, is measured by the measuring head equipped with six inductive sensors. These sensors take readings on three surfaces of a hexahedron, perpendicular in regard to each other (Fig. 2).

This setup makes possible the unequivocal determination of displacement of the center of the test block, and in connection the angular and linear errors of the robot. The experiments were carried out by inductive displacement sensors, of GT61 type, from TESSA Company, with a measuring range of ± 5 mm, hysteresis error of 0.2 µm and coefficient of linear expansion of 0.09 µm/°C.

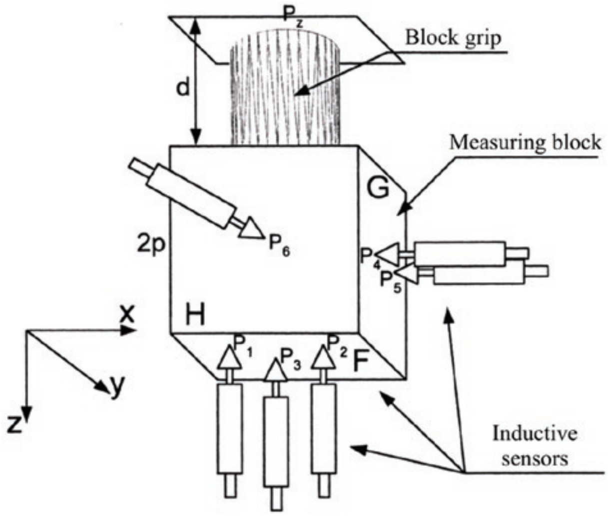
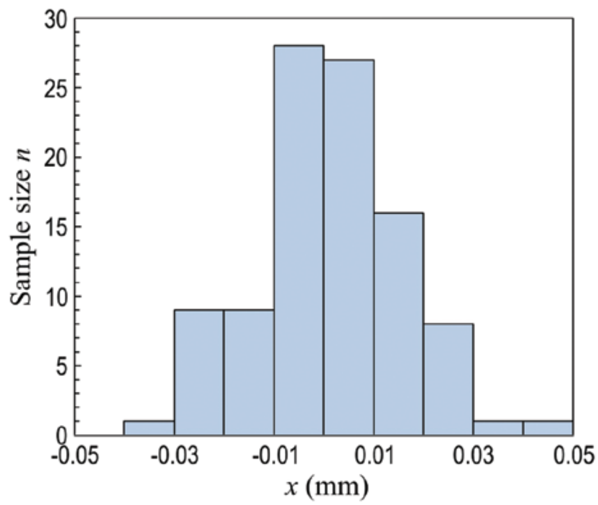


Fig. 2. System of inductive sensors of measuring head: P1-P6 – measuring points; F, G, H – measuring surfaces

The results of calculations and measurements of the errors for two sample points in the robot's workspace are presented in Table 1 and Fig. 3a,b. The statistical tests conducted at the significance level of $\alpha = 0.05$ showed that the error in positioning repeatability can be described with a 2-D random variable subject to a Gaussian distribution.

Repeatability positioning tests were carried out in laboratory conditions that ensure the required ambient conditions both by the manufacturer and the EN ISO 9283: 2003 standard. In order not to collide the measuring box with the measuring head and sensors, the robot moved between the points in a straight line - linear interpolation. During the research, it was noticed that the change of the effector's speed from 10.7% to 65.1% of the maximum speed does not significantly affect the positioning error. The increase of the effector's speed from 65.1% to 100%, i.e. to the maximum speed, resulted in a significant increase in error (at about 30%), therefore the tests were carried out at maximum speed and maximum load (in accordance with the requirements of EN ISO 9283: 2003), which for the Mitsubishi RV robot -M2 is 2 kg. Such an approach also provides the possibility to compare the results obtained with the maximum error value given in the robot's instructions (± 0.1 mm).

a)



b)

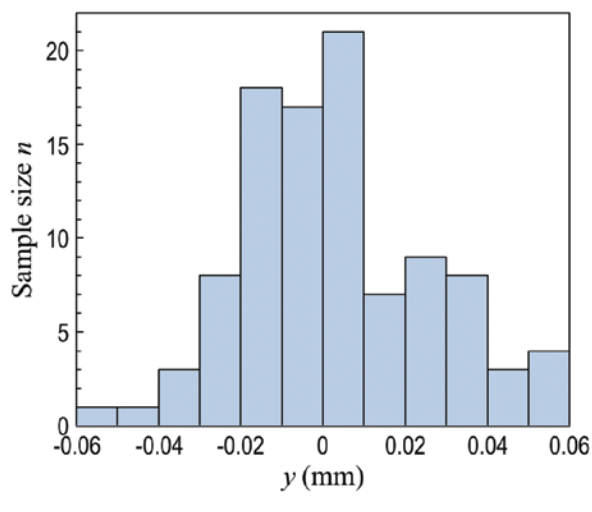


Fig. 3. Histograms of linear errors of the Mitsubishi RV-M2 robot at the point of the workspace described by a set of generalized coordinates given in Table 1 for the x-axis (a) and y-axis (b) directions

3. Robot error induced by temperature change

According to the requirements of the EN ISO 9283:2003 standard [10] on performance criteria and related test methods for robots used for manipulation operations in industrial processes, the investigations have to be carried out under stable temperature conditions. Under real

Table 1. The values of the random variable parameters of the Mitsubishi RV-M2 robot error

Joint coordinate (rad)	The experimentally determined values of parameters of normal distribution of probability density of the robot's error				
	Random variable	Minimum (mm)	Maximum (mm)	Standard deviation (mm)	Skewness
$q_1 = 0.5235$ $q_2 = 0.8726$ $q_3 = -1.3962$ $q_4 = -1.0471$	Distribution parameters evaluated theoretically: $\sigma_{xk} = 0.016$ mm, $\sigma_{yk} = 0.017$ mm, $\sigma_{zk} = 0.017$ mm				
	x	-0.032	0.043	0.014	0.067
	y	-0.054	0.057	0.018	0.287
	z	-0.051	0.050	0.017	0.075
$q_1 = 1.3963$ $q_2 = 0.3839$ $q_3 = -1.2217$ $q_4 = -0.7330$	Distribution parameters evaluated theoretically: $\sigma_{xk} = 0.021$ mm, $\sigma_{yk} = 0.016$ mm, $\sigma_{zk} = 0.016$ mm				
	x	-0.046	0.068	0.023	0.225
	y	-0.048	0.027	0.016	-0.406
	z	-0.053	0.059	0.018	0.045

production conditions, the ambient temperature can change, which has a direct impact on the repeatability positioning of the robot.

In order to determine the robot's error resulting from ambient temperature changes in the established place of the workspace, the schema of the robot's kinematic structure should be modified and some assumptions should be made. The angular dimensions of the robot set point (the values of joint coordinate q_i) should be taken as constant, while the linear dimensions, which in the robot RV-M2, for example, cannot change their values, should be modelled as kinematic reciprocating pairs that are positioned in the position l_i with the deviation Δl_i. The extension of the robot's arm unit with the length of l_i caused by temperature changes should be adopted as the deviation Δl_i. Therefore, the deviation Δl_i takes the form of a function dependent on the temperature (T), the linear expansion coefficient (χ), and the length of the robot's units (l_i).

$$\Delta l_i = f(\chi_i, l_i, \Delta T) \tag{5}$$

If the setting of each kinematic pair is burdened with a certain error Δl_i then the actual position of the robot end tip will be shifted in relation to the desired nominal position of the vector p. The coordinates of this vector end can be written as:

$$\Delta x = \sum_{i=1}^{n} \frac{\partial X}{\partial l_i} \cdot \Delta l_i \tag{6}$$

$$\Delta y = \sum_{i=1}^{n} \frac{\partial Y}{\partial l_i} \cdot \Delta l_i \tag{7}$$

$$\Delta z = \sum_{i=1}^{n} \frac{\partial Z}{\partial l_i} \cdot \Delta l_i \tag{8}$$

For the Mitsubishi RV-M2 industrial robot that is the object of the investigations, these equations (6–8) take the following form:

$$\begin{aligned}\Delta x &= \sin(q_1)\Delta l_2 + \sin(q_1)\cos(q_2)\Delta l_3 \\ &+ \sin(q_1)\cos(q_2+q_3)\Delta l_4 + \sin(q_1)\cos(q_2+q_3+q_4)\Delta l_5\end{aligned} \tag{9}$$

$$\begin{aligned}\Delta y &= \cos(q_1)\Delta l_2 + \cos(q_1)\cos(q_2)\Delta l_3 \\ &+ \cos(q_1)\cos(q_2+q_3)\Delta l_4 + \cos(q_1)\cos(q_2+q_3+q_4)\Delta l_5\end{aligned} \tag{10}$$

$$\Delta z = \Delta l_1 + \sin(q_2)\Delta l_3 + \sin(q_2+q_3)\Delta l_4 + \sin(q_2+q_3+q_4)\Delta l_5 \tag{11}$$

Knowing the value of the deviations Δl_i, it is possible to determine the value of the robot error caused by temperature changes. Determination of deviations Δl_i based on direct measurement requires the use of specialized measuring equipment. Furthermore, gear clearances that cause errors in the setting of joint coordinates Δq_i can partially compensate the temperature-induced increase in the length of each robot unit. The use of the indirect method based on knowledge of the kinematic structure of the robot is preferred in this case. For this purpose, measurements of the robot's error at five different points of its workspace were made. The increment in the temperature ΔT is 3°C with no change in the joint coordinates q_i. Then the values of the deviations Δl_i are determined by solving the system of equations:

$$\begin{cases} \Delta x_k = \sum_{i=1}^{n} \frac{\partial X}{\partial l_i} \cdot \Delta l_i \\ \Delta z = \sum_{i=1}^{n} \frac{\partial Z}{\partial l_i} \cdot \Delta l_i \end{cases} \tag{12}$$

where $k = 1, 2, \ldots, n-1$; n is the number of robot links.

It was found that the average deviations Δl_i induced by temperature changes are equal to $\Delta l_1 = 0.014$ mm, $\Delta l_2 = 0.005$ mm, $\Delta l_3 = 0.009$ mm, $\Delta l_4 = 0.007$ mm, and $\Delta l_5 = 0.008$ mm. To determine whether the temperature-induced errors affect the nature of the random variable distribution of the robot's error caused by setting errors of joint coordinates, investigations consisting of linear displacement of the measuring block (Fig. 4) to the desired position in the robot's workspace at two different ambient temperatures, 20 and 23°C, were conducted. Next, the arithmetic mean from the sample, which is a consistent and unbiased estimator of the expected value μ, is determined.

Because the investigations were carried out at points of the workspace for which standard deviations of the robot's error were known, for statistical verification of the investigations the parametric test of significance of average value is used. For each sample the hypothesis about the mean value H: $\mu = \mu_0$ is adopted. This hypothesis states that the average value of the analysed characteristic of the population is equal to the value of μ_0 determined during the static measurement of the error at increased ambient temperature (Table 2), assuming that the analysed population characteristic has a distribution $N(\mu, \sigma)$, while the alternative hypothesis is K: $\mu \neq \mu_0$. To verify this hypothesis, the test statistic U is used, which is defined by the formula:

$$U = \frac{\overline{X} - \mu_o}{\sigma}\sqrt{n} \qquad (13)$$

which, assuming the truth of the hypothesis H: $\mu = \mu_0$ (for the number of results n >30 [25]) is a standardized normal random variable with distribution $N(0,1)$. The sample size was n = 100.

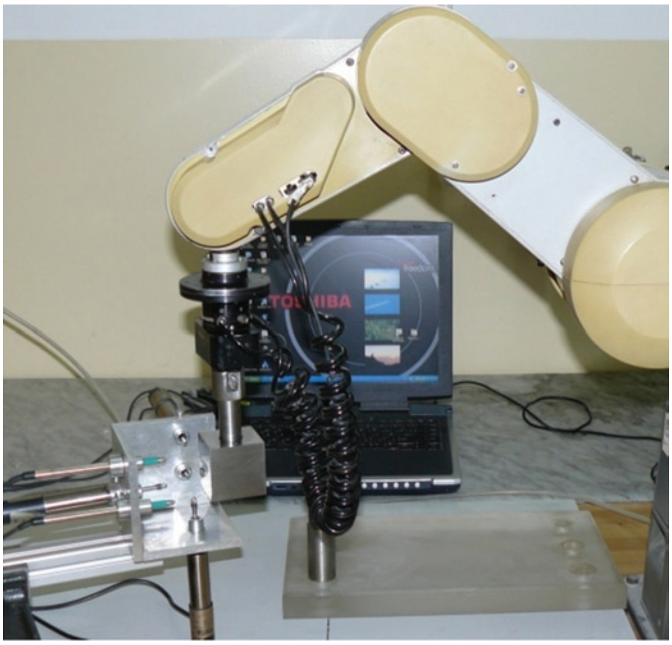


Fig. 4. The view of a measuring head equipped with a measuring block and sensors

In the case in which the measurements were carried out after stabilization of environmental heat conditions, at the level of significance α = 0.05, the value of the statistic U (Eq. 13) did not belong to a critical set $\left(-\infty, -u\left(1-\frac{1}{2}\alpha\right)\right\rangle \cup \left\langle u\left(1-\frac{1}{2}\alpha\right), +\infty\right)$, where $u\left(1-\frac{1}{2}\alpha\right)$ is the quantile of the order $1-\frac{1}{2}\alpha$ of the distribution $N(1,0)$. So, there was no reason to reject the hypothesis H. This means that errors due to temperature changes do not affect the form of the distribution describing the robot's errors but only cause an increase or a decrease in the average value of the obtained results. The results of measurements of the gripper displacement due to temperature change and the mean values of the sample in the two exemplary points in the workspace of the assembly stand are shown in Table 2.

It should be noted that the measurements were carried out after stabilizing the ambient thermal conditions, so that all the arms obtained an equal temperature. In the absence of temperature stabilization, it is very difficult to determine the values of the robot's errors, because the lengths of the robot arms do not increase proportionally (the values of statistics U (Eq. 13) belong to the critical set). The robot end tip before the heat stabilization of mechanisms can change its position in relation to the nominal position; that is, it can be displaced within a certain area. To determine this area, the origin of the local reference system is adopted in the nominal position. To investigate how the deviation of the position of the robot end tip from its ideal position changes, it was assumed that the deviation Δl_r in the setting of one of kinematic pairs will change, and the deviation of the setting accuracy of the other kinematic pairs will retain a permanent value. To do this in Eq. 14, which specifies the components of the deviation vector of the working tip from an ideal position, Δl_r should be taken as a variable parameter.

$$\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} \frac{\partial Y}{\partial l_1} \cdots \frac{\partial Y}{\partial l_r} \cdots \frac{\partial Y}{\partial l_i} \\ \frac{\partial Z}{\partial l_1} \cdots \frac{\partial Z}{\partial l_r} \cdots \frac{\partial Z}{\partial l_i} \end{bmatrix} \cdot \begin{bmatrix} \Delta l_1(\chi_1, l_1, \Delta T) \\ \Delta l_r(\chi_r, l_r, \Delta T) \\ \Delta l_i(\chi_i, l_i, \Delta T) \end{bmatrix} \qquad (14)$$

The complete matrix of total differentials in Eq. 14 can be treated as a Jacobian matrix of coefficients of sensitivity of the robot to the change in the length of the kinematic chain due to the temperature change. This task entails finding the equation of the family of parallel lines in Cartesian coordinates:

$$\frac{\partial Y}{\partial l_r} z = \frac{\partial Z}{\partial l_r} y + \sum_{i=1}^{n} \begin{vmatrix} \frac{\partial Y}{\partial l_r} & \frac{\partial Y}{\partial l_i} \\ \frac{\partial Z}{\partial l_r} & \frac{\partial Z}{\partial l_i} \end{vmatrix} \Delta l_i(\chi_i, l_i, \Delta T) \qquad (15)$$

Choosing the appropriate values of the extreme deviations Δl_i caused by temperature changes, we can find two lines from the family of parallel lines defined by Eq. 15. In this way, the polygon of the robot's positioning accuracy taking into account changes in temperature-induced linear dimensions of the robot can be created. Inside the polygon are all the possible vectors of position deviation of the robot's end tip from the desired nominal position. Knowing the most extreme position of a polygon, the largest displacement of the end tip can be found. Figures 5a and 5b show a tolerant polygon of the robot's error caused by temperature changes at two different points of the robot's workspace. An analysis of the figures shows that the maximum error values depend not only on temperature change but also on the choice of the points in the robot's workspace. This makes it possible to reduce the robot's error by choosing (i) the place in its workspace characterized by the lowest error value or (ii) the place with the least sensitivity to the change of the length of the robot's arms due to temperature changes.

Table 2. Values of the Mitsubishi RV-M2 robot error induced by changes in ambient temperature ΔT = 3 °C

Joint coordinate, (rad)	Δx (mm)	Δy (mm)	Δz (mm)	$\overline{\Delta x}$ (mm)	$\overline{\Delta y}$ (mm)	$\overline{\Delta z}$ (mm)
q_1 = 0.5235 q_2 = 0.8726 q_3 = -1.3965 q_4 = -1.0471	0.008	0.014	0.009	0.007	0.016	0.011
q_1 = 1.3963 q_2 = 0.3839 q_3 = -1.2217 q_4 = -0.7330	0.017	0.003	0.004	0.020	0.003	0.006

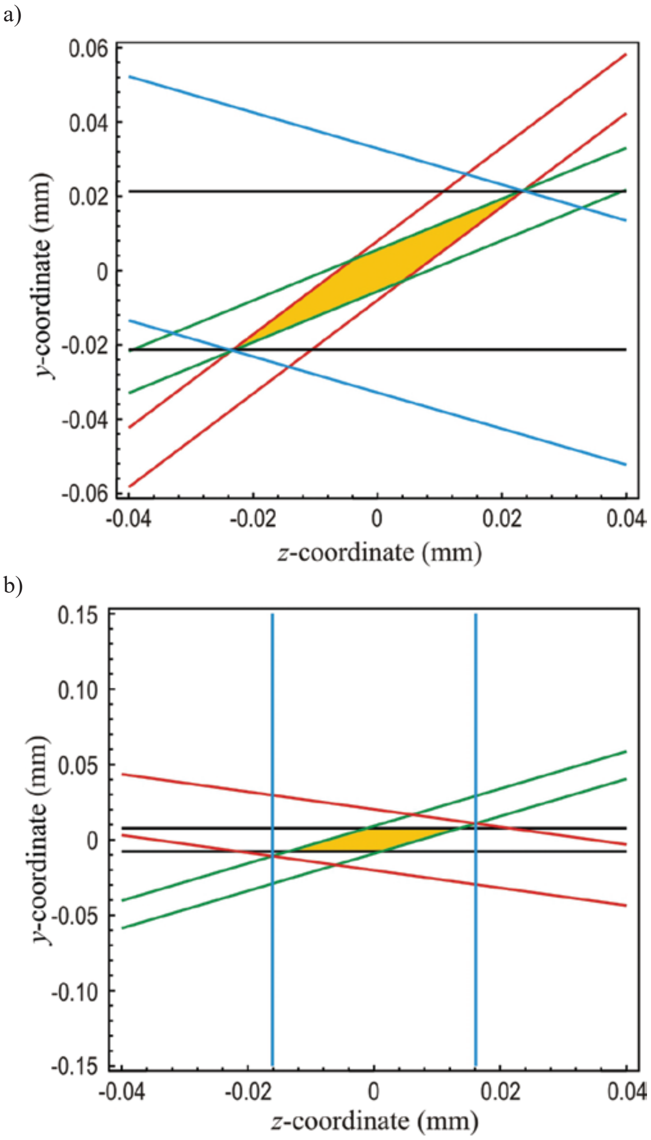


Fig. 5. Polygon of Mitsubishi RV-M2 robot error caused by temperature changes at the point defined by joint coordinates: (a) $q_1 = 0.5235$ rad, $q_2 = 0.6981$ rad, $q_3 = -0.3490$ rad, $q_4 = -1.0471$ rad; (b) $q_1 = 0.5235$ rad, $q_2 = 0.8726$ rad, $q_3 = -1.3962$ rad, $q_4 = -1.0471$ rad

4. Randomization of temperature-induced error of robot

The systematic error resulting from the temperature change can be compensated based on the determined correction [e.g., 4, 9, 21, 29]. However, this requires constant monitoring of the temperature value. If the robot operates in conditions where it is not possible to stabilize the ambient temperature, the correction is unknown. The correction is associated with the error value that can be determined on the basis of the expected range in which the correction is contained. It is connected with the assumption that the systematic error may have a value in the range of $\mu = (0 \pm \delta)$. In this way, the systematic temperature-induced error may be theoretically randomized.

Practical randomization depends on assuring the error dispersion of temperature-induced conditions during the tests, so that in the following measurements the systematic error takes random values from the range $(0 \pm \delta)$ or extreme values. In this case, however, it requires knowledge of the form of the distribution describing the variation of systematic errors. If the temperature in the robotized assembly stand stabilizes at a certain level and the lowest operating temperature occurs as rarely as the highest temperature, the randomization of systematic error can be carried out based on the normal distribution, which greatly simplifies the process of assessing the robot's accuracy.

If, however, there is an equal probability of occurrence of both the lowest and the highest temperature, the uniform distribution should be used. Such a situation may take place when total heat stabilization of the manipulator does not occur. Convolution of the normal distribution (repeatability) with density function $f_{Yk}(y)$ and uniform distribution (temperature error) with density function $f_{Yt}(y)$ exhibits a flat normal distribution. The probability density function of this distribution is described by the formula:

$$PDF\left(\eta_y\right) = f_{Y_k+Y_t}\left(y\right) = \int_{-\infty}^{\infty} f_{Y_t}\left(y\right) \cdot f_{Y_k}\left(\varsigma - y\right) \tag{16}$$

Density functions of this distribution are generally characterized by a constant value in the vicinity of the expected value and in slopes described by a Gaussian function (Fig. 6). The range of the stability of the density function depends on the parameter r of the distribution, which determines the ratio of the standard deviation σ_t of its rectangular component to the standard deviation σ_k of its normal component [11, 19]:

$$r = \frac{\sigma_t}{\sigma_k} \tag{17}$$

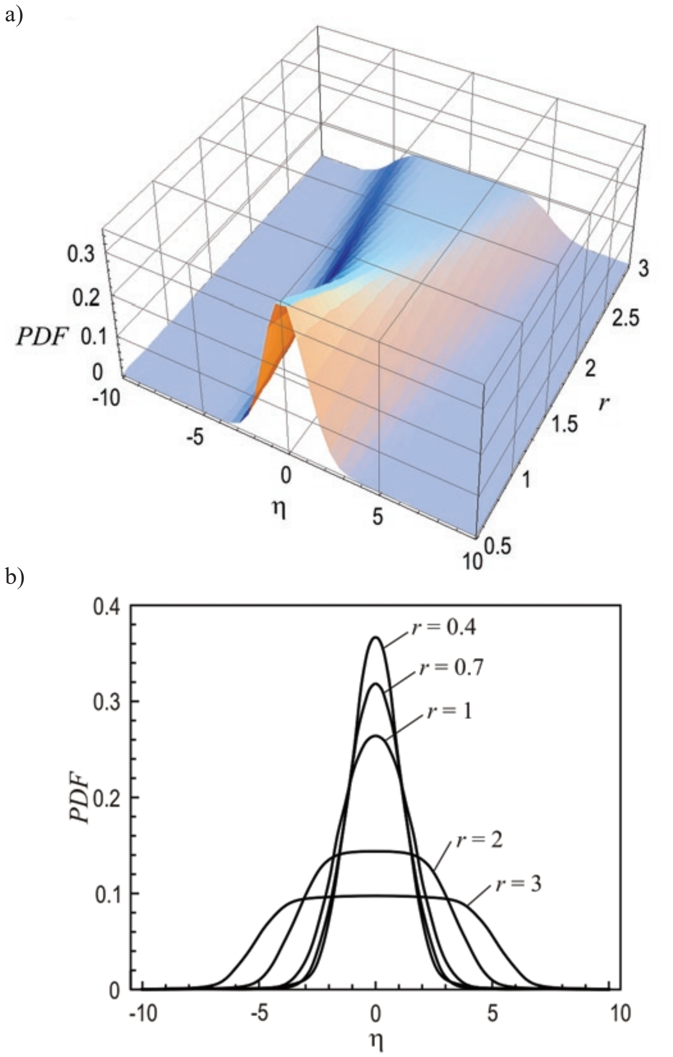


Fig. 6. Distribution of the function of the density of a flat normal distribution depending on the coefficient r (a) and the effect of the r parameter value on the shape of the PDF function (b)

$$PDF\left(\eta_y\right)=\frac{1}{2\sqrt{6\pi}r}\int\limits_{\eta-\sqrt{3}*r}^{\eta+\sqrt{3}*r}\exp\left[-\frac{\varsigma^2}{2}\right]d\varsigma \tag{18}$$

If the standard deviation σ_t of the robot temperature error is less than or equal to the standard deviation σ_k of the kinematic error then the shape of the plane-normal distribution is close to a Gaussian distribution (Fig. 6). Thus, it can be assumed that the total error of positioning repeatability can be approximated in this case in the form of a normal probability density distribution.

For the verification of the abovementioned assumptions, investigations of the repeatability positioning of the robot at the working point defined by the joint coordinates given in row 1 of Table 1 (σ_y = 0.018 mm) have been carried out. The error values of the Mitsubishi RV-M2 error caused by changes in the ambient temperature at the considered point after the stabilization of the thermal conditions were $\Delta y = \pm\, 0.016$ mm. Due to the fact that the purpose of the study was to analyse the influence of temperature errors of the robot on the value of the total error of repeatability positioning, the investigations were conducted in a wide range of temperature variation, which assures a significant effect of these errors. During investigations, changes of the ambient temperature were applied in the range of ± 3°C without waiting for stabilization of the link temperature. It was also assumed that there is an identical probability of occurrence of the temperature from the considered variation range (uniform distribution). It should be stressed that a similar share of temperature errors can occur with smaller temperature changes, but they occur elsewhere in the workspace or in the case of robots with a greater length of links [2, 7]. It was assumed that the robot error caused by temperature changes is subject to a uniform probability distribution. On this basis, the variance of the randomized distribution of the random variable is determined:

$$\sigma_t^2=\frac{\left(\Delta y_{t\max}-\Delta y_{t\min}\right)^2}{12}=8.1\cdot 10^{-5}\,\mathrm{mm} \tag{19}$$

Then the variance of the resultant distribution:

$$\sigma_{\eta y}{}^2=\sigma_{yk}^2+\sigma_{yt}^2= \tag{20}$$

is determined.

The results were statistically analysed using the Shapiro-Wilk test [25, 26] to verify the normality of the distribution of a random variable. During the investigations the following hypothesis was formulated: the null hypothesis H_0 that the distribution of the analysed characteristic is normal. For $\alpha = 0.05$ and $n = 100$, the tabulated critical value W(α, n) = W (0.05, 100) = 0.964 was less than the calculated value, which meant that there was no basis to reject the hypothesis of the normality of the distribution of the obtained data. A histograms and a graphs of the normal distribution of the obtained results are shown in Figs. 7a and 8a.

An analysis of the results shows that when the standard deviation of a random variable induced by temperature changes (uniform distribution) is equal to 50% of the standard deviation of the robot's kinematic error then the total error of the robot with sufficient accuracy can be approximated by the normal distribution of the random variable. Because the sample size used in the experiment $n = 100$ could be too small to confirm the correctness of the assumptions employed, simulation investigations were carried out.

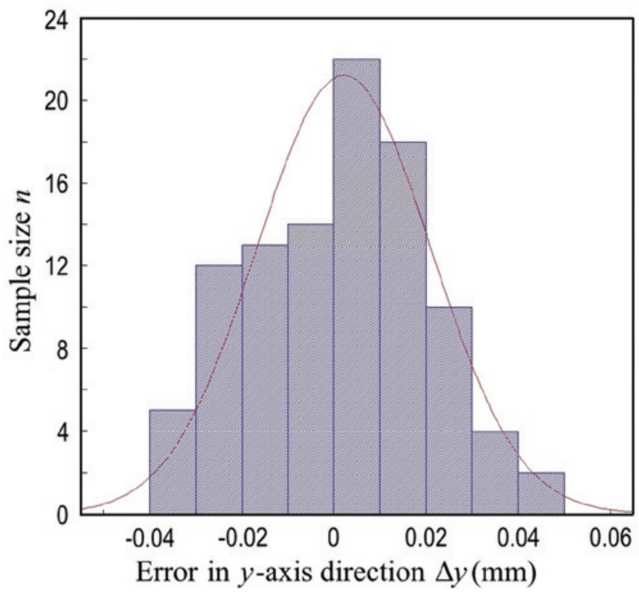
During the research, 5000 pseudo-random numbers subject to a normal distribution with parameters derived based on the both the measurements (Table 1) and the uniform distribution simulated during the experiment (Fig. 7b) were generated.

Then, the results of a sum of random variables' distributions were investigated to find their consistency with the normal distribution. To carry out this analysis, the Shapiro-Wilk test is also used. The results showed that there were no reasons to reject the hypothesis about the consistency of the results with the normal distribution (Fig. 8b). The results of the simulations confirm, therefore, the results of experimental investigations.

5. The probability of joining parts

An important issue in the operation of a robotic assembly station is the problem of ensuring the required probability of joining the parts involved. The tasks related to the robotisation of assembly can be greatly facilitated by decomposing joints according to the shape of the surface of the assembled parts. From this point of view, the assembly of typical joints can be examined as a typical series of joining parts with flat, cylindrical, conical, spherical, threaded and other sur-

a)



b)

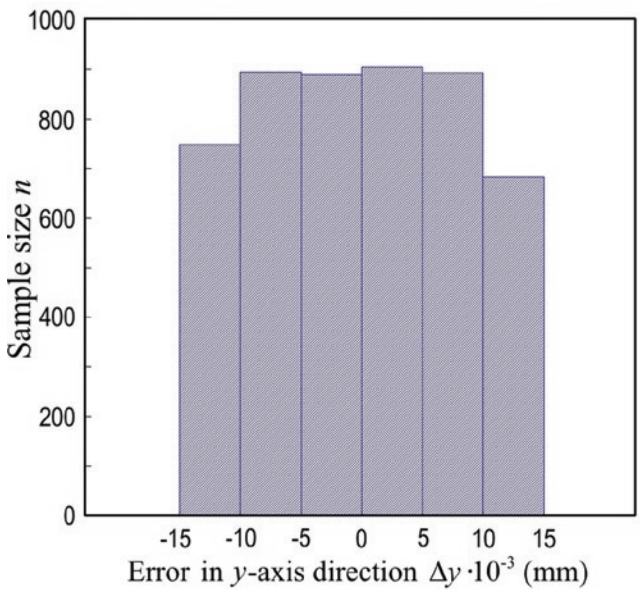


Fig. 7. (a) histogram of the results of the measurement repeatability positioning error of the robot, taking into consideration the temperature errors; and (b) histogram of a random variable of the robot error induced by temperature changes

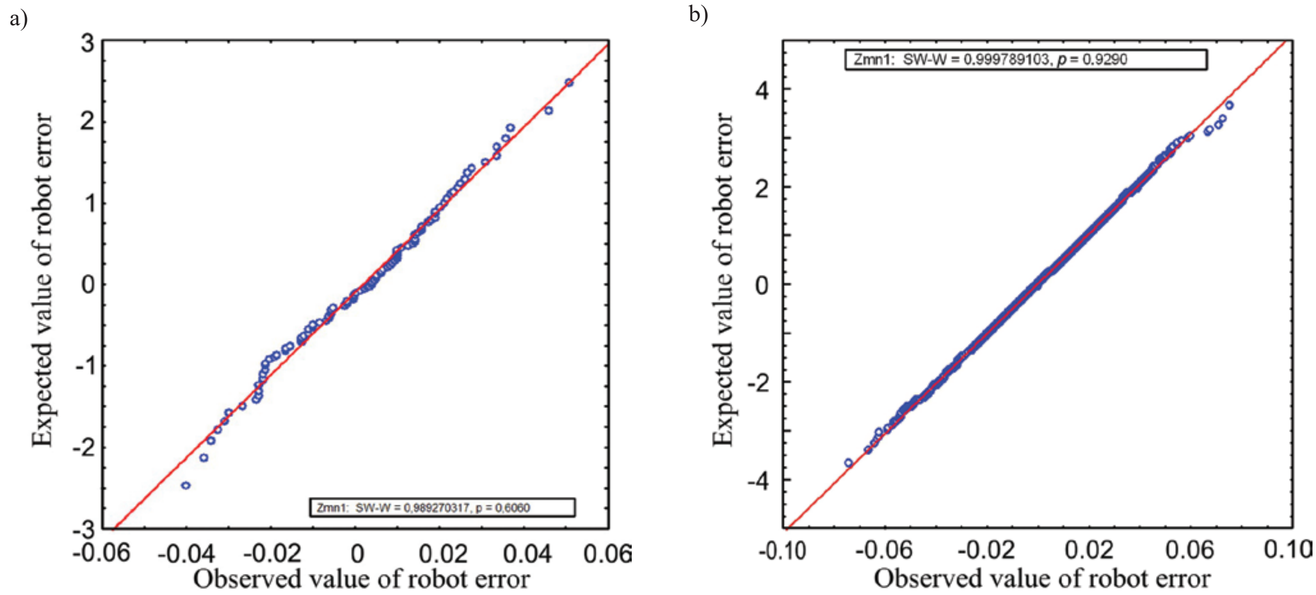


Fig. 8. Diagram of compatibility of the results of measurement (a) and simulation (b) of the robot error with a normal distribution

faces. Among them, joints with cylindrical surfaces constitute about 40% of the total number of connections [5]. Since, in the majority of cases, robots are used to carry out the process of assembly of cylindrical parts with guaranteed clearance [22], the next section of the work is limited to joining parts with cylindrical surfaces. The repeatability positioning error of the robot causing the relative displacement of the axes of the joined parts is a two-dimensional random variable $X = [x, y]^T$ subjected to the normal probability distribution with the covariance matrix Λ_k and the matrix of expected values μk^T.

$$f(X) = \frac{1}{2\pi\sqrt{|\Lambda_K|}} \exp\left[-\frac{1}{2}(X-\mu_K)^T \Lambda_K{}^{-1}(X-\mu_K)\right] \quad (21)$$

The elements of the covariance matrix Λ_k correspond to the boundary standard deviations listed in Table 1. If during the operation of the robotic assembly station there are small temperature changes causing a random displacement of the error mean values, the total positioning repeatability error for r < 1 can be described as a two-dimensional normal random variable with the covariance matrix Λ_η and the matrix of expected values $\mu\eta^T$ (Fig. 9).

The probability of joining cylindrical parts is the probability of an event occurring that the distance between their axes reaches an assumed value of $0.5L$ (η_1, η_2), i.e. the probability of the event that a random variable describing the distance between the axes of the joined parts will be inside a hypothetical cylinder with the centre located in the nominal point N and a diameter corresponding to the clearance of the parts to be joined L. In order to determine the value of the prob-

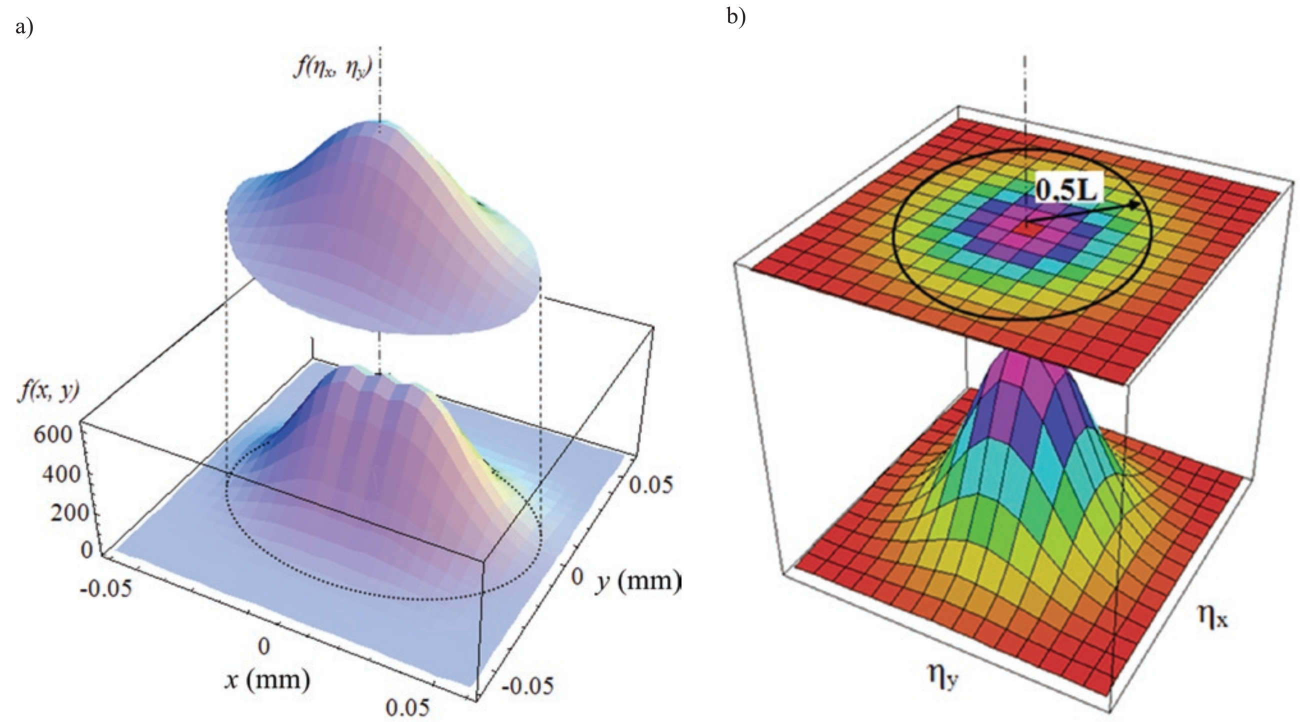


Fig. 9. The method for determining the probability of joining cylindrical parts: a) random variable of the displacement error of the part axes $f(\eta_x, \eta_y)$, b) the area of integration of the random variable f $f(\eta_x, \eta_y)$

ability, one should integrate the density function of the relative error distribution of the displacement of the part axis in the area of O:{ $\eta_x^2 + \eta_y^2 \le (0{,}5\,L)^2$} as follows:

$$P = \iint_{\eta_x^2+\eta_y^2 \le (0.5L)^2} \frac{1}{2\pi\sqrt{|\Lambda_\eta|}} \exp\left[-\frac{1}{2}\left(\eta-\mu_\eta\right)^T \Lambda_\eta^{-1}\left(\eta-\mu_\eta\right)\right] d\eta_x d\eta_y \quad (22)$$

where: $\Lambda\eta$ - covariance matrix of a random variable of error of relative displacement of a part axis, $\mu\eta^T$ - matrix of expected values of a random variable of relative error of the displacement of part axes.

The occurrence of temperature changes during the assembly process leads to an increase in robot error, which results in a reduction of the probability of joining parts and a reduction of the reliability of the whole process. Depending on the place where the joining is made, a different probability value can be obtained in the work space of the station (Fig. 10). In the case of an assembly operation at a point characterised by the error value given in Table 1 (row 1), a change in temperature causes a reduction in the probability and therefore also reduces the qualitative capability C_p of the process by 17.8% (from 1.33 to 1.1). Ensuring the required level of quality capability of the process (C_p = 1.33) requires an increase in the joining clearance by 17.21%.

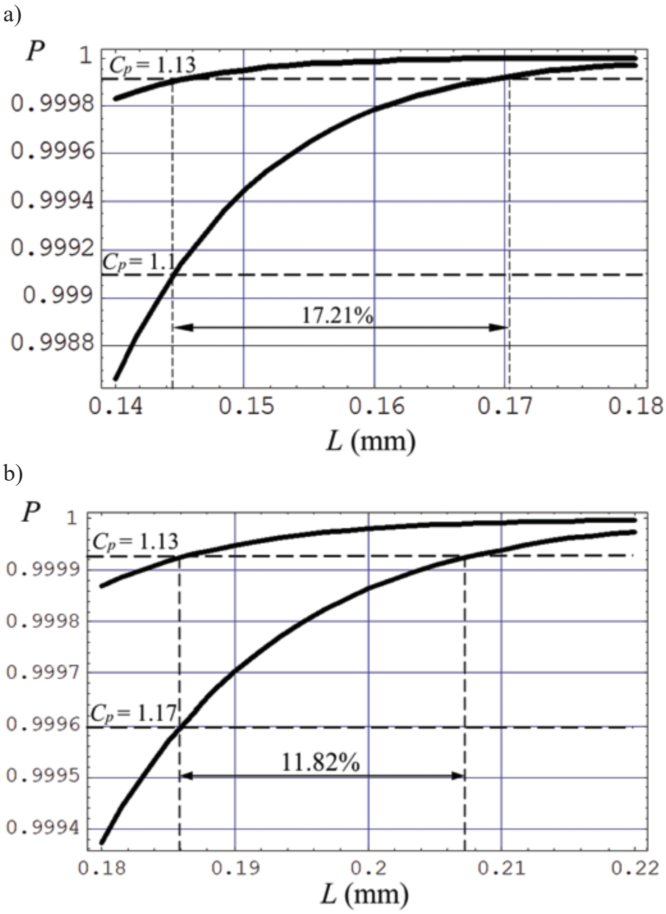


Fig. 10. The influence of temperature changes on the probability of joining parts: a) at the point specified in row 1 of Table 1, b) at the point specified in row 2 of Table 1

At the point in the robot workspace corresponding to the parameters listed in row 2 of Table 1 for a joint clearance of L = 0.145 mm, a 24.81% lower quality capability of the process (C_p = 1) can be obtained. This is due to the fact that at this point the robot exhibits a much higher error value of positioning repeatability. A change in the temperature at this point also causes a decrease in the quality capability index of the process from C_p = 1.33 to C_p = 1.17 and thus by 12%. It is possible to ensure the required probability of joining the part (corresponding to a process quality capability index at a level of C_p = 1.33) by increasing the joint clearance by 12%. Thus, depending on the choice of location in the robot workspace where the assembly process is carried out, one can firstly obtain a different level of quality of the capability index of the process, and secondly a different impact of the errors related to the inability to stabilise the ambient temperature on the probability of joining parts.

6. Summary and conclusions

Industrial robots are successfully used in many areas of manufacturing processes such as welding, drilling, and handling. However, in recent years, the interest of many researchers is focused on the possibility of implementing robots in processes that require high precision such as assembly or measurement. Ensuring high reliability of these processes requires consideration of all the factors affecting their repeatability positioning. The most important parameters influencing the repeatability positioning include the error that results from the kinematic errors of settings of programmed joint coordinates' values and the errors caused by changes in ambient temperature.

This paper shows that if the ambient temperature can change within a small range during the operation of a robotic station, it may be very difficult to correct the error associated with it. Adoption of the maximum error value seems to be unjustified, because this error can be partially compensated for by a kinematic error subject to the normal probability distribution. The research conducted has shown that in practice this error can be randomised based on the expected range of error variability in the form of a uniform probability distribution. In the situation where the ratio of the standard deviation of the temperature error σ_t to the standard deviation of the kinematic error σ_t is less than r < 1, one can make a sufficiently accurate approximation of the distribution of the total repeatability positioning error using a normal probability distribution. When the value of the coefficient r is greater than 1 this error is subjected to a flat-normal distribution.

The analysis of the results in this study has shown that performing assembly in conditions of unstable ambient temperature results in a lower probability of joining parts and process quality capability. Ensuring the required level of assembly of parts requires an increase in the joint clearance from 12 to 17% depending on the choice of location in the workspace. Otherwise, the quality capacity of the process may be reduced, and thus the reliability of the workstation may be reduced. The results obtained indicate that the required reliability of the process can be obtained by selecting an appropriate location for the joining process characterised by the smallest error value and the smallest sensitivity to changes in the environmental conditions.

References

1. Abderrahim M, Khamis A, Garrido S, Moreno L. Accuracy and Calibration Issues of Industrial Manipulators. [in:] Hua L K. Industrial Robotics- Programming, Simulation and Applications. ARS/plV: Germany, 2006, https://doi.org/10.5772/4895.
2. Angelidis A, Vosniakos G C. Prediction and compensation of relative position error along industrial robot end-effector paths. International Journal of Precision Engineering and Manufacturing 2014; 15(1): 63-73, https://doi.org/10.1007/s12541-013-0306-5.
3. Brethé J F, Vasselin E, Lefebvre D, Dakyo B. Modeling of repeatability phenomena using the stochastic ellipsoid approach. Robotica 2006; 24(4): 477-490, https://doi.org/10.1017/S0263574705002481.
4. Chen H, Fuhlbrigge T, Choi S, Wang J, Li X. Practical industrial robot zero offset calibration, Proceedings of the 4th IEEE Conference on Automation Science and Engineering; 2008 Aug 23-26; Washington; 2008: 516-521.
5. Cho N, Tu J F. Quantitative circularity tolerance analysis and design for 2D precision assemblies. International Journal of Machine Tools and Manufacture 2002: 42(13): 1391-1401, https://doi.org/10.1016/S0890-6955(02)00080-9.
6. Craig J J. Introduction to robotics: Mechanics and control. Massachusetts: Addison-Wesley, 1989.
7. Du G., Zhang P. Online robot calibration based on vision measurement. Robotics and Computer-Integrated Manufacturing 2013; 29 (1): 484-492, https://doi.org/10.1016/j.rcim.2013.05.003.
8. Eastwood S, Webb P. Compensation of thermal deformation of a hybrid parallel kinematic machine. Robotics and Computer-Integrated Manufacturing 2009; 25(1): 81-90, https://doi.org/10.1016/j.rcim.2007.10.001.
9. Elatta A, Gen L P, Zhi F L, Daoyuan Y, Fei L. An overview of robot calibration. Information Technology Journal 2004; 3(1): 74-68, https://doi.org/10.3923/itj.2004.74.78.
10. EN ISO 9283:2003, Manipulating industrial robots - Performance criteria and related test methods. Geneva: International Standards Organization, 2003.
11. Fotowicz P. Systematic effect as a part of the coverage material. Metrology and Measurement Systems 2010; 17(3): 439-446, https://doi.org/10.2478/v10178-010-0037-1.
12. Gong C, Yuan J, Ni J. Nongeometric error identification and compensation for robotic system by inverse calibration. International Journal of Machine Tools Manufacturing 2000; 40(14): 2119-2137, https://doi.org/10.1016/S0890-6955(00)00023-7.
13. Gürsel A, Bijan S. A systematic technique to estimate positioning errors for robot accuracy improvement using laser interferometry based sensing. Mechanism and Machine Theory 2005; 40 (8): 879-906, https://doi.org/10.1016/j.mechmachtheory.2004.12.012.
14. Hameed Farhan U, Tolouei-Rad M, Osseiran A. Assembly modelling approach for special purpose machines. Assembly Automation 2018; 38 (2):158-172, https://doi.org/10.1108/AA-12-2016-170.
15. Heisel U, Richter F, Wurst K H. Thermal behaviour of industrial robots and possibilities for error compensation. CIRP Annals 1997; 46(1): 283-286, https://doi.org/10.1016/S0007-8506(07)60826-9.
16. Józwik J, Ostrowski D, Jarosz P, Mika D. Industrial robot repeatability testing with high speed camera Phantom. v2511. Advances in Science and Technology. Research Journal, 2016; 10(32): 86-96.
17. Kluz R, Trzepieciński T. The repeatability positioning analysis of the industrial robot arm. Assembly Automation 2014; 34(3): 285-295, https://doi.org/10.1108/AA-07-2013-070.
18. Kluz R, Trzepieciński T. Analysis of optimal orientation of robot gripper for an improved capability assembly process. Robotics and Autonomous Systems 2015; 74: 253-266, https://doi.org/10.1016/j.robot.2015.08.007.
19. Kluz R. Effect of temperature-induced errors on accuracy of robotised assembly stand with Mitsubishi RV-2 robot. Technology and Automation Assembly 2004; 2: 19-24.
20. Li R, Zhao Y. Dynamic error compensation for industrial robot based on thermal effect model. Measurement 2016; 88: 113-120, https://doi.org/10.1016/j.measurement.2016.02.038.
21. Nguyen H N, Zhou J, Kang H J. A new full pose measurement method for robot calibration. Sensors 2013; 13(7): 9132-9147, https://doi.org/10.3390/s130709132.
22. Pitchandi N, Perumaal S S, Irulappan M. Insertion force analysis of compliantly supported peg-in-hole assembly. Assembly Automation 2017; 37 (3): 285-295, https://doi.org/10.1108/AA-12-2016-167.
23. Płaczek M, Piszczek Ł. Testing of an industrial robot's accuracy and repeatability in off and online environment. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (3): 455-464, https://doi.org/10.17531/ein.2018.3.15.
24. Ramesh R, Mannan M A, Poo A N. Error compensation in machine tools - a review: Part II: Thermal errors. International Journal of Machine Tools Manufacturing 2000; 40(9): 1257-1284, https://doi.org/10.1016/S0890-6955(00)00010-9.
25. Ruppert D. Statistics and finance. An introduction. New York: Springer, 2004.
26. Shapiro S S, Wilk M B. An analysis of variance test for normality. Biometrika 1965; 52(3/4): 591-611, https://doi.org/10.2307/2333709.
27. Shiakolas P S, Conrad K L, Yih T C. On the accuracy, repeatability and degree of influence of kinematics parameters for industrial robots. International Journal of Modelling and Simulation 2002; 22(4): 1-10, https://doi.org/10.1080/02286203.2002.11442246.
28. Slamani M, Nubiola A, Bonev I. Assessment of the positioning performance of an industrial robot. Industrial robot: An International Journal, 2012; 39(1): 57-68, https://doi.org/10.1108/01439911211192501.
29. Wu Y, Klimchik A, Caro S, Furet B, Pashkevich A. Geometric calibration of industrial robots using enhanced partial pose measurements and design of experiments. Robotics and Computer-Integrated Manufacturing 2015; 35: 151-168, https://doi.org/10.1016/j.rcim.2015.03.007.
30. Yin S, Guo Y, Ren Y, Zhu J, Yang S, Ye S. Real-time thermal error compensation method for robotic visual inspection system. International Journal of Advanced Manufacturing Technology 2014; 75(5-8): 933-946, https://doi.org/10.1007/s00170-014-6196-6.
31. Yu H, Zhang G, Ran Y, Li M, Wang Y. A comprehensive and practical reliability allocation method considering failure effects and reliability costs. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20(2): 244-251, https://doi.org/10.17531/ein.2018.2.09.
32. Yu Z, Shi Q, Wang H, Yu N, Huang Q, Fukuda T. How to achieve precise operation of a robotic manipulator on a macro to micro/nano scale. Assembly Automation 2017; 37 (2):186-199, https://doi.org/10.1108/AA-02-2017-017.
33. Zhenhua W, Hui X, Guodong C, Rongchuan S, Sun L. A distance error based industrial robot kinematic calibration method. Industrial Robot: An International Journal, 2014; 41(5): 439-446, https://doi.org/10.1108/IR-04-2014-0319.

Rafał KLUZ
Department of Manufacturing and Production Engineering
Rzeszow University of Technology
al. Powstańców Warszawy 8, 35-959 Rzeszów, Polska

Andrzej KUBIT
Department of Manufacturing and Production Engineering
Rzeszow University of Technology
al. Powstańców Warszawy 8, 35-959 Rzeszów, Polska

Jarosław SĘP
Department of Manufacturing and Production Engineering
Rzeszow University of Technology
al. Powstańców Warszawy 8, 35-959 Rzeszów, Polska

Tomasz TRZEPIECINSKI
Department of Materials Forming and Processing
Rzeszow University of Technology
al. Powstańców Warszawy 8, 35-959 Rzeszów, Polska

E-mails: rkktmiop@prz.edu.pl, akubit@prz.edu.pl, jsztmiop@prz.edu.pl, tomtrz@prz.edu.pl

Article citation info:
OU D, TANG M, XUE R, YAO H. Hybrid fault diagnosis of railway switches based on the segmentation of monitoring curves. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 514–522, http://dx.doi.org/10.17531/ein.2018.4.2.

Dongxiu OU
Maojie TANG
Rui XUE
Hongjing YAO

HYBRID FAULT DIAGNOSIS OF RAILWAY SWITCHES BASED ON THE SEGMENTATION OF MONITORING CURVES

HYBRYDOWA DIAGNOSTYKA USZKODZEŃ ZWROTNIC KOLEJOWYCH W OPARCIU O SEGMENTACJĘ KRZYWYCH PRĄDOWYCH

Switches are one of the most important pieces of infrastructure in railway signal systems, and they significantly influence the efficiency and safety of train operation. Currently, the identification of switch failures mainly depends on the experience of railway staff and the use of simple thresholding methods. However, these basic methods are highly inaccurate and frequently result in false and missing alarms. This paper aims to develop a hybrid fault diagnosis (HFD) method for railway switches. The method is an intelligent diagnosis method that uses massive current curves collected by microcomputer monitoring systems. We first divide the switch operation current curves into three segments based on the three mechanical processes that occur during switch operation. Then, a standard curve is selected from the fault-free curves, and common typical faults are ascertained through a microcomputer monitoring system. Finally, derivative dynamic time warping and a quartile scheme are employed to identify fault curves. An experiment based on current curves collected from the Guangzhou Railway Bureau in China demonstrates that the HFD method is extremely accurate and has low false and missing alarm rates. HFD performs better than the studied support vector machine (SVM) and dynamic time warping (DTW) methods, which are widely used for fault diagnosis.

Keywords: *switch system, fault detection and diagnosis, intelligent method.*

Zwrotnice stanowią jeden z najważniejszych elementów infrastruktury systemów sygnalizacji kolejowej i mają znaczący wpływ na wydajność i bezpieczeństwo eksploatacji pociągów. Obecnie, identyfikacja awarii zwrotnic zależy głównie od doświadczenia personelu kolejowego i opiera się na stosowaniu prostych metod progowania. Jednakże te elementarne metody są wysoce niedokładne i często skutkują fałszywymi alarmami lub brakiem alarmu. Niniejszy artykuł ma na celu opracowanie hybrydowej metody diagnostyki błędów (HFD) dla zwrotnic kolejowych. Metoda ta jest inteligentną metodą diagnostyczną, która wykorzystuje wykresy przebiegu prądowego zebrane przez mikrokomputerowe systemy monitorowania. Najpierw krzywe prądowe działania zwrotnicy dzieli się na trzy segmenty w oparciu o trzy procesy mechaniczne, które zachodzą podczas jej działania. Następnie, spośród krzywych opisujących działanie bezusterkowe, wybiera się przebieg standardowy, a w dalszej kolejności ustala się, z wykorzystaniem mikrokomputerowego systemu monitorowania, najczęściej występujące, typowe błędy działania zwrotnicy. Wreszcie, do identyfikacji krzywych błędów stosuje się schemat kwartylowy oraz metodę derivative dynamic time warping wykorzystującą pochodne do klasyfikacji szeregów czasowych. Eksperyment oparty na krzywych prądowych zebranych przez Guangzhou Railway Bureau w Chinach pokazuje, że metoda HFD jest wyjątkowo dokładna i skutkuje niską liczbą fałszywych i brakujących alarmów. HFD daje lepsze wyniki niż szeroko stosowane do diagnozowania błędów metody maszyny wektorów nośnych (SVM) i dynamic time warping (DTW).

Słowa kluczowe: *układ zwrotnicowy, wykrywanie i diagnozowanie usterek, metoda inteligentna.*

1. Introduction

With the rapid development of high-speed rail (HSR) around the world, the current world speed record for a commercial train has reached to 574.8 km/h. Increased train speed is convenient but also causes safety and reliability problems. Track circuits, railway annunciators and switches are generally the three main components that contribute to the operational safety of HSR. Of these three components, switches (Fig. 1), which connect equipment that supports train transit from one track to another, are mainly responsible for the efficiency and safety of HSR. However, switch failures have recently caused several major railway accidents [28]. According to a statistical report by the Jinan Railway Bureau in 2015-2016, 191 switch faults accounted for approximately 60% of signal faults. Thus, early diagnosis of issues with switch systems is critical for the operational safety of HSR.

To achieve the safe operation of HSR, microcomputer monitoring systems (MMSs) have been widely introduced to timely monitor switch states in China [27]. MMSs collect switch operation current and power curves that allow maintenance staff to identify the state of switches and make diagnoses based on their experience. However, a lack of experience can lead to missing or false alarms, both of which pose serious security risks. Furthermore, the number of switch operation curves is relatively large, and many financial and human resources are involved in such work.

Domestic and foreign experts have conducted several studies on fault diagnosis. Early attempts employed simple thresholding methods [16, 21] to detect faults, but frequent false and missing alarms limit the extensive application of these methods. A more recent study [6] summarized three primary approaches in the literature for switch diagnostics: feature, model and empirical methods.

Fig. 1. Railway track switch

For feature-based methods, special features that can be rapidly identified are extracted. Data collection, feature extraction, and feature selection form three subsections of this model. Marquez et al. [17] used data from tests conducted on a commonly found point mechanism and discussed the benefits of adopting a Kalman filter for preprocessing data collected during tests. Eker et al. [6] proposed a support vector machine (SVM) operated through principle component analysis (PCA) for dimensionality reduction to diagnose faults in switches. Six different features were selected, and four remained following a T-test. Asada et al. [2] developed a new approach to fault detection and diagnosis that involved utilizing parameters collected from low-cost and accessible sensors; they focused on fault detection and diagnosis for ‘overdriving’ and ‘underdriving.’ Lee et al. [13] introduced a data mining solution that employs audio data to detect and diagnose switch faults. Zhou et al. [27] proposed an improved SVM that accommodates fault detection, and the authors optimized the geometric parameter feature extraction method developed by He [10].

In model-based methods, a model is defined to characterize a system. Deviation from the model is defined as a failure and is identified as the difference between the model outcome and actual data. Eker et al. [7] presented a simple state-based prognostic (SSBP) method for fault detection and forecasting in electromechanical systems. Ardakani et al. [1] established a strategy and technical architecture for the prognostic and health management (PHM) of electromechanical point machines. Zhang et al. [23, 24] proposed a switch fault detection algorithm based on a probabilistic neural network and back propagation neural network. Letot et al. [14] proposed a model for degradation trend assessment and a methodology that updates degradation paths and reliability data to accurately estimate the remaining useful life. Wang et al. [22] proposed a failure prediction model based on a Bayesian network to evaluate the effects of weather patterns on railway switches.

In empirically based methods, a fault-free sample is used as a reference signal, and failures are identified based on the resemblance of a given signal to a reference signal. Atamuradov et al. [3] introduced an expert system based on an economic analysis method that identifies the best maintenance policy for a failure mode and/or system component. Zhao and Lu [26] presented a switch fault diagnosis method based on gray correlation analysis. The authors related the basis of the variations in the power curve to the typical faults of a switch machine. Kim et al. [12] proposed a diagnosis method that involves applying DTW to manage variations in the duration of railway point machine use; this model manages only phase-shifted shape faults, and the parameter δ of DTW chosen by maintenance staff serves as a threshold.

However, the abovementioned methods do not adequately address the problem at hand. For example, the Kalman filter method can achieve success only for a portion of a dataset (reverse to normal). SVM-based methods are sensitive to feature selection, and few authors can explain how to select such features. Artificial neural networks are not suitable for this task, as lacking a sufficient number of fault samples can result in underfitting. In addition, an expert system functions according to large amounts of priori knowledge, thereby requiring a considerable amount of manpower from experienced railway staff. Although DTW performs effectively for shape faults, it cannot detect faults over shorter or longer durations. To overcome these limitations, this paper trains HFD using a small sample dataset, i.e., with a small amount of priori knowledge. In addition, HFD is used to detect and diagnose eleven typical faults summarized by the maintenance staff of the Guangzhou HSR. Moreover, HFD identifies fault curves automatically from a computer and can reduce the quantities of manpower and resources required.

The remainder of this paper is organized as follows. Section 2 introduces switch operation current curves and explains why these curves must be divided into three segments before fault diagnosis. The mathematical principles and calculation processes of HFD are explained in Section 3. Section 4 presents a numerical experiment using real switch operation current curves for fault diagnosis, followed by a discussion and concluding remarks in Section 5.

2. Analysis of switch operation current curves

2.1. Basic analysis of current curves

Although MMSs can collect current and power curves, only current curves have been widely used for fault diagnosis because current values provide an enormous amount of information regarding switches, such as their electrical and mechanical characteristics [25]. Therefore, experienced maintenance staff can identify switch faults by observing various characteristics of current curves. Current curves can be divided into the following three segments based on three mechanical processes: the start stage, action stage and release stage. For example, Fig. 2 shows the fault-free curves of a railway switch. The start stage $(0 - T_1)$ exhibits a peak current when the machine begins to operate; the action stage $(T_1 - T_2)$ is relatively smooth, and it corresponds to the working process of the switch; and the release stage $(T_2 - T_3)$, which is typically called the “small step”, indicates that the switch has finished switching and has connected the relevant circuit.

2.2. Fault types and segmented current curves

Through long-term observation and analysis, the maintenance staff of the Guangzhou HSR summarized the fault current curves for the track. Eleven types of faults occurred on the track: abnormal fluctuation, poor contact in the action circuit, abnormal impedance in the action circuit, start failure, conversion failure, release failure, open start-up circuit, electric relay 2DQJ switch failure, blocking in the gap, machine idling, and overlong release time of the starting relay. These faults, referred to as M1-M11, are described in Table 1.

In Table 1, the faults include shape and duration faults. The data associated with these fault modes are different from fault-free data in shape or duration. The fault stage indicates the stage in which a

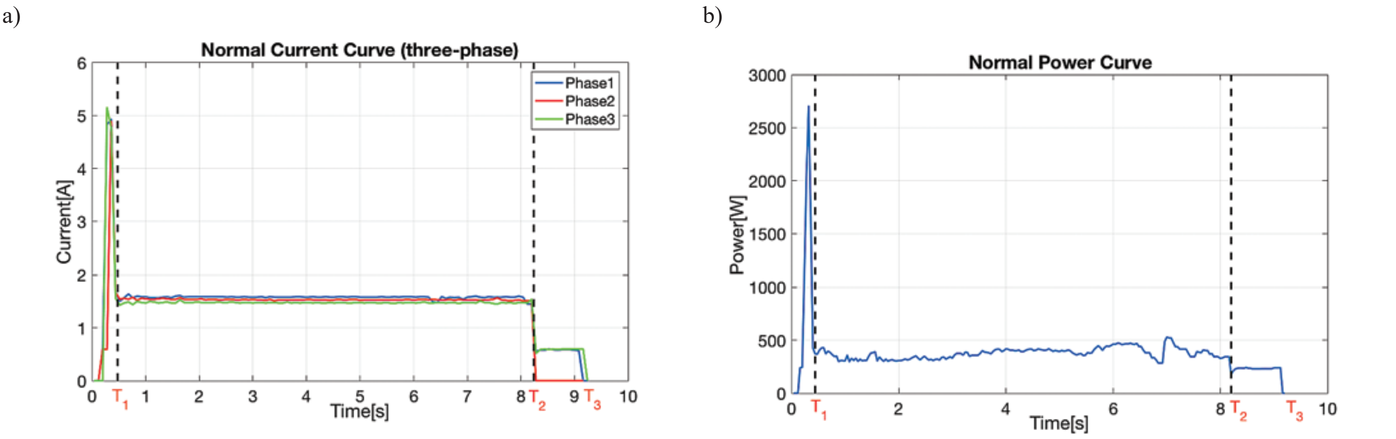


Fig. 2. Switch operation curves

fault occurs. Therefore, the maintenance staff can make rapid fault-solving decisions when the anomalous stage is known.

Currently, the segmenting methods mainly depend on two fixed points to divide current curves into three stages. However, the two fixed points may not apply to all switches. Fig. 3 shows the cumulative switch current curves of Station #1 (Fig. 3a) and Station #2 (Fig. 3b) for January (taking single-phase current data as an example). In Fig. 3, the duration of the current data is approximately 5.5 s for Station #1, and it is 9 s for Station #2. The durations are typically different at all stations, which can be referred to as a "different durations" problem. Therefore, only two adaptive points can divide all current curves into three stages with high accuracy rather than using two fixed points.

3. Model and algorithm for railway switch hybrid fault diagnosis

The proposed HFD method involves the following three steps: fault-free dataset selection, standard curve selection and fault detection and diagnosis. The first step involves dividing samples (current curves) into three segments and constructing a fault-free dataset; the second step involves selecting the best sample, referred to as the "standard curve," from the fault-free dataset; and the third step involves comparing test samples with the standard curve and other fault types for fault detection and diagnosis. The details of HFD are presented below.

3.1. Fault-free dataset selection

3.1.1. Curve segmentation

In this section, an adaptive mean-shift (AMS) algorithm is used for segmentation [5, 8]. This algorithm iterates by pointing in the direction of the maximum increase in density and involves the following six steps.

- Step 1: Collect a current curve from MMSs, and start with an input $X = [x_1, x_2 \ldots x_n]$.
- Step 2: Choose an arbitrary point as the initial center y_0 from X, a bandwidth h and a kernel function $K(x)$. In AMS, the bandwidth equals σ_X (the standard deviation of X), and the

Table 1. Fault types and corresponding attributes

Fault Types	Corresponding Curve Characteristics	Fault Modes	Abnormal Stages	Symbols
Abnormal fluctuation	Abnormal fluctuations in the action current	Shape Fault	$T_1 - T_2$	M_1
Poor contact in the action circuit	Abrupt change in the action current	Shape Fault	$T_1 - T_2$	M_2
Abnormal impedance in the action circuit	Conversion current that exceeds the limit	Shape Fault	$T_1 - T_2$	M_3
Start failure	Small step in the action stage	Shape Fault	$T_1 - T_2$	M_4
Conversion failure	Rising current in the action stage	Shape Fault	$T_1 - T_2$	M_5
Release failure	Two peaks exist in the action stage	Shape Fault	$T_1 - T_2$	M_6
Open start-up circuit	Zero value curve	Shape Fault	$T_1 - T_2$	M_7
Electric relay 2DQJ switch failure	A "small steps" curve	Shape Fault	$T_2 - T_3$	M_8
Blocking in the gap	Missing "small steps"	Shape Fault	$T_2 - T_3$	M_9
Machine idling	Overly long conversion time	Duration Fault	$T_1 - T_2$	M_{10}
Overlong release time of the starting relay	Overly long "small steps"	Duration Fault	$T_2 - T_3$	M_{11}

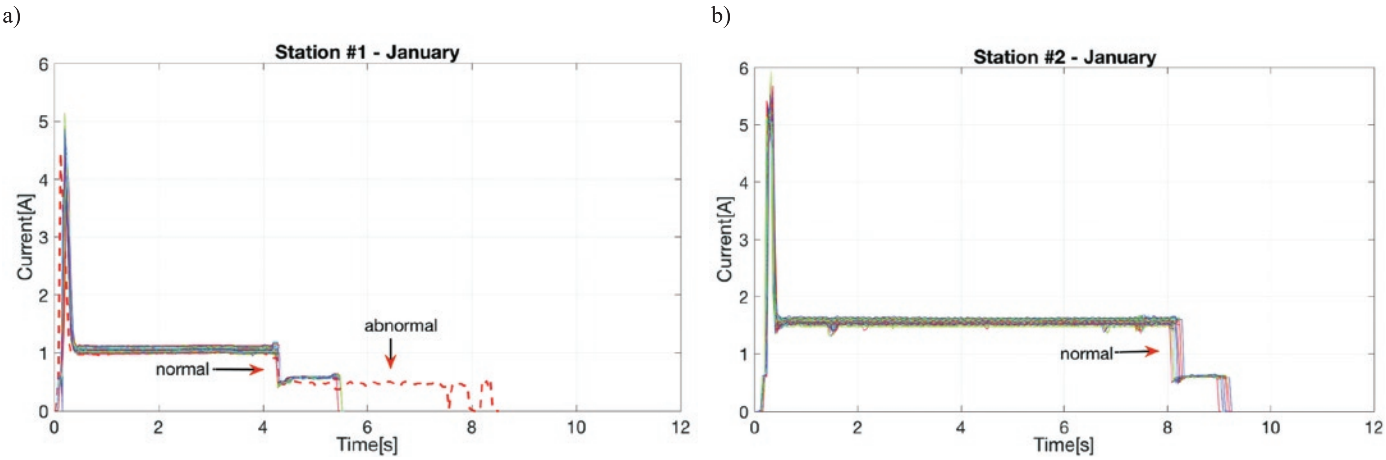


Fig. 3. Cumulative Current Curves

spherical normal kernel [8] function $\tilde{K}(x)$ is coordinated with the bandwidth. The multivariate kernel density estimate $f(x)$ obtained from $\tilde{K}(x)$ and σ_X is:

$$f(x)=\frac{1}{n\sigma_X}\sum_{i=1}^{n}\tilde{K}\left(\frac{x-x_i}{\sigma_X}\right). \tag{1}$$

For radially symmetric kernels, the profile of the kernel $k(x)$ is determined to satisfy:

$$\tilde{K}(x)=c_k k\left(x^2\right), \tag{2}$$

where c_k is a normalization constant that ensures that $K(x)$ satisfies:

$$\int_R \tilde{K}(x)=1 \tag{3}$$

- Step 3: Calculate the gradient of the density estimate as follows:

$$\nabla f(x)=\frac{2c_k}{n\sigma_X^3}\sum_{i=1}^{n}(x_i-x)g\left(\frac{y_t-x_i}{\sigma_X}^2\right)$$
$$=\frac{2c_k}{n\sigma_X^3}\sum_{i=1}^{n}g\left(\frac{y_t-x_i}{\sigma_X}^2\right)\left(\frac{\sum_{i=1}^{n}x_i g\left(\frac{y_t-x_i}{\sigma_X}^2\right)}{\sum_{i=1}^{n}g\left(\frac{y_t-x_i}{\sigma_X}^2\right)}-x\right) \tag{4}$$

where $g(s)$ is equal to $-k'(s)$ and y_t is the center of the current iteration (t starts at index 0). The first term is proportional to the density estimate at x computed from kernel $G(x)=c_k g\left(x^2\right)$, and the second term is the mean-shift.

$$m_{\sigma_X}(x)=\frac{\sum_{i=1}^{n}x_i g\left(\frac{y_t-x_i}{\sigma_X}^2\right)}{\sum_{i=1}^{n}g\left(\frac{y_t-x_i}{\sigma_X}^2\right)}-x \tag{5}$$

- Step 4: Iterate the mean-shift procedure until convergence is achieved, including the successive computation of the mean-shift vector $m_{\sigma_X}\left(x^t\right)$ and the translation of the center $y_{t+1}=y_t+m_{\sigma_X}\left(x^t\right)$. This iteration is guaranteed to converge to a point where the gradient of the density function is zero [4].
- Step 5: Divide the points in X that satisfy Equation (6) into one cluster and remove them from X .

$$\left|x_i-y_t\right|\le\sigma_X \quad 1\le i<n \tag{6}$$

- Step 6: Return to Step 2 until there are no points in X .

AMS can divide input X into several clusters. The cluster with the largest number of elements is defined as the action cluster. Furthermore, X can be grouped into three segments based on the two elements with the minimum subscript i and maximum subscript j of the action cluster. The segmentation result is shown in Fig. 4.

Due to the electromechanical properties of railway switches [20], the action cluster always corresponds to the action stage; therefore, the three parts of X correspond to the three stages of switch operation.

3.1.2. Fault-free dataset extraction

In this section, the K-means method is used to obtain a fault-free dataset. In the "different durations" problem, several features

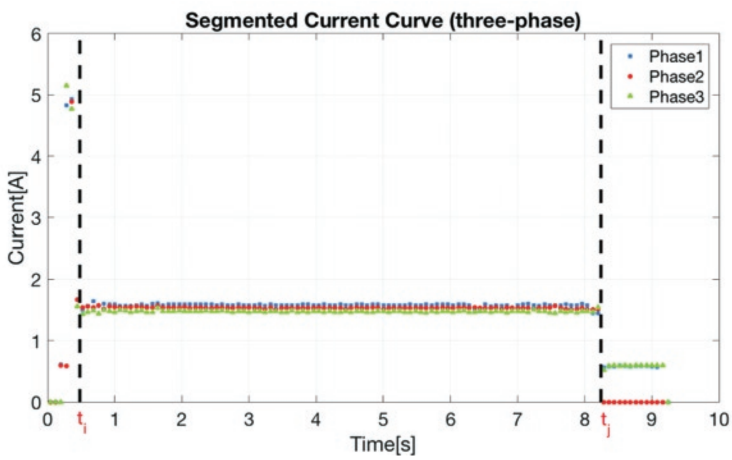


Fig. 4. Segmented Current Curve

are extracted based on previous research [27, 6, 15] to unite dimensions, as shown in Table 2. In the start stage, no fault type exists, and relatively few features have been chosen. In the action and release stages, duration and shape faults both exist; thus, the time span has been selected for duration faults, and other metrics are associated with shape faults.

Table 2. Features of Different Stages

Stages	No.	Features
Start stage	1	Time span
	2	Maximum value
	3	Mean current value
	4	Median current value
Action stage	5	Time span
	6	Max current value
	7	Minimum current value
	8	Mean
	9	Median
	10	Standard deviation
	11	Peak factor
	12	Fluctuation factor
Slow release stage	13	Time span
	14	Max current value
	15	Minimum current value
	16	Mean
	17	Median
	18	Standard deviation
	19	Peak factor
	20	Fluctuation factor

The K-means method has been widely used in clustering for simplicity, and the algorithmic details have been summarized in previous research [9]. The inputs of this method consist of two parameters: the feature matrix $\bar{M}$ and number of clusters K.

$$Idx^K = K\_means\left(\bar{M}, K\right) \tag{7}$$

where $\mathrm{Idx}^{\mathrm{K}}$ is an array and the superscript of Idx represents the number of clusters in the array. The feature matrix $\bar{M}$ is defined by the twenty features shown in Table 2 (e.g., m sequences $\left(n_1, n_2, \ldots, n_m\right)$ can generate a feature matrix with m rows and twenty columns). The number of clusters K is determined by assuming that more than half of the samples are fault-free for regular switches. The optimal K^* can be determined from the following optimization problem:

$$\max K$$

$$\begin{cases} \sum_{i=1}^{m} 1\left(Idx_i^{K+1} = mode\left(Idx^{K+1}\right)\right) < m/2 \\ \sum_{i=1}^{m} 1\left(Idx_i^{K} = mode\left(Idx^{K}\right)\right) \geq m/2 \end{cases} \tag{8}$$

where:

$1 =$ indicator function

$\mathrm{mode}(\mathrm{x}) =$ value that appears most often in array x

$\mathrm{m} =$ number of samples

The above integer programming problem can be solved by the enumeration method. As a result, the fault-free dataset N^* with K^* satisfies:

$$N^* = \left\{ n_j^* \in \left\{n_1, n_2, \ldots, n_m\right\} \mid n_j^* * 1\left(Idx_j^{K^*} = mode\left(Idx^{K^*} \right)\right) = n_j^* \right\}_{j=1}^{m} . \tag{9}$$

3.2. Standard curve selection

3.2.1. Derivative dynamic time warping

Derivative dynamic time warping (DDTW) is a modified DTW method [11]. The approach involves obtaining similarities between two arbitrary trajectories, and it achieves better alignment by “warping” the time axis of one sequence or both sequences. The algorithm details can be summarized as follows.

Assume two arbitrary switch current sequences X^1 and X^2 of lengths n_1 and n_2, respectively, where:

$$X^1 = \left\{x_1^1, x_2^1, \ldots x_i^1 \ldots, x_{n_1}^1\right\} \tag{10}$$

$$X^2 = \left\{x_1^2, x_2^2, \ldots x_j^2 \ldots, x_{n_2}^2\right\} . \tag{11}$$

To align the two sequences, an n-by-m matrix is considered, where the $\left(\mathrm{i}^{\mathrm{th}}, \mathrm{j}^{\mathrm{th}}\right)$ element represents the distance $\mathrm{d}\left(\mathrm{x}_\mathrm{i}^1, \mathrm{x}_\mathrm{j}^2\right)$ between points x_i^1 and x_j^2. With DDTW, the distance measure $\mathrm{d}\left(\mathrm{x}_\mathrm{i}^1, \mathrm{x}_\mathrm{j}^2\right)$ is the square of the difference of the estimated derivatives [18] of x_i^1 and x_j^2. Each matrix element (i, j) corresponds to the alignment between points x_i^1 and x_j^2. Therefore, a warping path W is used to define the mapping between X^1 and X^2. The l^{th} element of W is defined as $\mathrm{w}_\mathrm{l} = (\mathrm{i}, \mathrm{j})_\mathrm{l}$; thus, we obtain the following relation:

$$W = \left\{w_1, w_2, \ldots, w_l, \ldots, w_L\right\}, \; max\left(n_1, n_2\right) \leq L < n_1 + n_2 - 1 . \tag{12}$$

Previous research [18] on DTW has demonstrated that W can be efficiently found by dynamic programming. To formulate a dynamic programming problem, a distance measure must be used between two elements. In this paper, the 1-norm ($\left\|\cdot\right\|$) is chosen as the distance function δ:

$$\delta\left(i, j\right) = x_i^1, x_{j1}^2 \tag{13}$$

After a distance measure is defined, the DTW problem can be formally defined as a minimization over potential warping paths based on the cumulative distance of each path, where δ is a distance measure between two elements. As a result, the similarity between two sequences is defined by Equation (14).

$$DDTW\left(X^1, X^2\right) = min_W\left[\sum_{l=1}^{L}\delta\left(w_l\right)\right] \tag{14}$$

3.2.2. Standard curve selection

In this section, the "best" option from N^* is selected as the standard curve. For the fault-free dataset N^* with l cardinality, an l-rank square matrix D can be constructed for which the $\left(i^{th}, j^{th}\right)$ element represents the similarity between the i^{th} sequence and j^{th} sequence in N^* (the "similarity" is typically defined as $D_{ij} = DDTW\left(n_i^*, n_j^*\right)$).

In this paper, the s^{th} sequence is defined as the standard curve if its index satisfies:

$$s = arg_i min\left(max_j\left(D_{ij}\right)\right). \tag{15}$$

3.3. Fault detection and diagnosis

3.3.1. Duration fault detection and diagnosis

In this section, an arbitrary sequence can be detected using a quartile scheme to determine whether a duration fault has occurred. The three steps of the quartile scheme are as follows:

- Step 1: Assume that a dataset with m samples (current curves) has been segmented into three stages (set the start stage F^{sta} as an example):

$$F^{sta} = \left\{F_1^{sta}, F_2^{sta} \ldots, F_m^{sta}\right\} \tag{16}$$

where F_i^{sta} is the start stage of the i^{th} sample. In addition, an array C^{sta} is set for when the j^{th} element c_j^{sta} equals the cardinality of F_j^{sta}.

- Step 2: Calculate the interquartile range of C^{sta} as:

$$IQR^{sta} = Q_3^{sta} - Q_1^{sta} \tag{17}$$

where Q_1^{sta} and Q_3^{sta} are the first and third quartiles of the start stage, respectively.

- Step 3: Define a decision function $P(i)$.

$$P(i) = 1\left(c_i^{sta}\left\langle IQR^{sta} - 1.5Q_1^{sta} \vee c_i^{sta}\right\rangle IQR^{sta} + 1.5Q_3^{sta}\right) \quad 1 \le i \le m \tag{18}$$

where:

$\vee$ = logical OR

As a result, the i^{th} sample can be identified as a duration fault when $P(i)$ equals one.

3.3.2. Shape fault detection and diagnosis

In this section, three steps are used to diagnose an arbitrary sequence F.

- Step 1: Divide F into three segments with the curve segmentation method:

$$F = \left\{F^{sta}, F^{act}, F^{rel}\right\} \tag{19}$$

- Step 2: Define a diagnosis dataset M that includes the standard current curves of three stages and their corresponding shape faults, as follows:

$$M = \left\{\left[S_1\right], \begin{bmatrix} S_2 \\ M_{1(2)} \\ M_{2(2)} \\ M_{3(2)} \\ M_{4(2)} \\ M_{5(2)} \\ M_{6(2)} \\ M_{7(2)} \end{bmatrix}, \begin{bmatrix} S_3 \\ M_{8(3)} \\ M_{9(3)} \end{bmatrix}\right\} \tag{20}$$

where $M_{i(j)}$ denotes j^{th} stage data of the i^{th} fault (S_j is the standard curve of the j^{th} stage). In the first stage (start stage), there is no fault type, which means that only S_1 exists in the first column of M. $M_1 - M_7$ occur in the second stage (action stage); thus, $M_{1(2)} - M_{7(2)}$ and S_2 are grouped together in the second column of M. Furthermore, M_8 and M_9 occur in the third stage (release stage); therefore, the third column of M consists of $M_{8(3)}$, $M_{9(3)}$ and S_3.

- Step 3: DDTW is employed to calculate the similarities between one stage in F and the corresponding stage in M. Each stage of F can be evaluated with Equation (21) and diagnosed with Table 3.

Table 3. Diagnostic Results for Shape Faults

Equation (20) Outputs / Diagnostic Results	$Label^{sta}$	$Label^{act}$	$Label^{rel}$
Fault-free	1	1	1
M1	1	2	1
M2	1	3	1
M3	1	4	1
M4	1	5	1
M5	1	6	1
M6	1	7	1
M7	1	8	1
M8	1	1	2
M9	1	1	3

$$\begin{cases} Label^{sta} = arg_i\, min\left(DDTW\left(F_{sta}, M_{i,1}\right)\right) \\ Label^{act} = arg_i\, min\left(DDTW\left(F_{act}, M_{i,2}\right)\right) \\ Label^{rel} = arg_i\, min\left(DDTW\left(F_{rel}, M_{i,3}\right)\right) \end{cases} \tag{21}$$

where $\mathrm{Label}^{\mathrm{sta}}$, $\mathrm{Label}^{\mathrm{act}}$ and $\mathrm{Label}^{\mathrm{rel}}$ respectively denote the classification results of the three stages.

4. Experiment and results

In this study, 1,964 fault-free curves and 115 fault curves were collected from the Guangzhou-Shaoguan Railway in China. The dataset was randomly split into two subsets (training and testing sets) that account for 70% and 30% of the entire dataset. For HFD, all training data are used to generate the standard curve. Then, 70% fault curves of the training set and the standard curve are combined to form the diagnosis dataset. The diagnostic results of 10 current curves are shown in Table 4.

In Table 4, M_{10} and M_{11} are determined by the quartile scheme, and the other faults are determined by DDTW. Test samples can be classified only as M_{10} and M_{11} when the corresponding decision function equals one. Without considering duration faults (M_{10} and M_{11}), the minimum of each row is found, which indicates that the i^{th} test sample is highly similar to the reference template; therefore, the samples can be classified in the same class.

Additionally, the DTW method [8] with the quartile scheme and the SVM method based on twenty features (Table 2) are compared with HFD. For the SVM, a Gaussian kernel is used as the kernel function, and the penalty factor and kernel parameter are determined by a 10-fold cross-validation method [19]. A quantitative comparison of the three methods is provided in Table 5. Two indicators, the false alarm rate (FAR) and missing alarm rate (MAR), are introduced in the table. FAR denotes the probability of classifying the fault-free data as faulty, and MAR denotes the probability of classifying fault data as fault free.

The following conclusions can be drawn from Table 5 regarding the experimental results.

- The HFD method is the best of the three methods due to its high accuracy, low FAR and low MAR.
- Compared to HFD, the DTW method exhibits classification results and cannot be used for fault diagnosis because of its high MAR. HFD performs better than DTW for two reasons. First, drawbacks such as “singularities” [26] prevent DTW from producing the best warping results. Second, DTW is focused on current values, but HFD focuses on both current values and data fluctuations.
- Compared to HFD, the SVM method offers a generally acceptable level of classification quality, but it still makes incorrect classifications and generates a relatively high MAR, which prevents the application of the SVM method in practical applications. As shown in Table 5 the HFD method performs better than the SVM method because HFD makes full use of all available information, whereas SVM disregards certain information when applying the feature extraction method.

5. Conclusions

In this paper, an intelligent fault diagnosis method is proposed based on the segmentation of railway switches. Through previous analysis, this paper illustrates how to divide current curves based on three mechanical processes for all railway switches and how to determine the similarities between them.

The experimental results show that the HFD method can detect faults with 99.43% accuracy and can diagnose faults with 98.67% accuracy. This approach is superior to the other two methods introduced above. Furthermore, the lower FAR and MAR of the HFD method demonstrate that HFD is the most robust tool for fault detection and diagnosis.

Future work will strive to integrate power curves with the proposed HFD method to achieve more accurate results. Furthermore, undefined switch faults will be examined for broader applicability and operability of the method. The final future objective is to more intelligently detect railway switch faults and eventually improve the safety and efficiency levels for passenger and cargo transport.

Acknowledgement

This work was supported by the National Key R&D Program of China (2016YFB1200401).

Table 4. Distance between the Test Samples and Reference Faults

i	S	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	Predicted Label	Actual Label
1	0.31	0.42	49.92	2.26	223.38	8.74	172.04	181.12	1.79	17.51	0	0	S	S
2	3.21	1.04	2.57	11.30	49.43	37.62	33.28	20.94	31.24	16.25	0	0	M_1	M_1
3	4.48	1.49	2.34	3.91	29.21	0.59	43.15	31.23	1.67	17.21	0	0	M_5	M_5
4	0.41	0.60	1.04	2.26	31.24	8.76	49.43	37.61	1.84	17.22	0	0	S	S
5	2.31	1.41	3.10	2.28	32.10	6.91	21.13	0.83	1.91	16.83	0	0	M_7	M_7
6	0.36	0.49	0.53	2.27	29.89	8.73	51.62	39.69	1.75	17.24	0	0	S	S
7	0.36	0.66	0.48	2.25	28.30	8.79	51.21	40.04	1.72	17.23	0	0	S	S
8	5.40	0.93	1.02	2.21	19.93	6.97	50.83	30.21	1.73	17.28	1	0	M_{10}	M_{10}
9	4.31	2.10	0.43	2.31	20.08	9.31	53.21	32.25	1.79	17.30	0	0	M_2	M_2
10	0.40	0.64	0.37	2.28	24.78	8.76	52.67	40.76	1.77	17.23	0	0	S	S

Table 5. Fault Detection and Diagnosis Results

DTW	M1%	M2%	M3%	M4%	M5%	M6%	M7%	M8%	M9%	M10%	M11%	S%
M1%	97	0	0	0	3	0	0	0	0	0	0	0
M2%	0	100	0	0	0	0	0	0	0	0	0	0
M3%	0	0	64	0	0	0	0	0	0	0	0	36
M4%	0	3	0	94	0	0	0	0	0	0	0	42
M5%	0	0	0	0	58	0	0	0	0	0	0	0
M6%	0	0	0	0	0	100	0	0	0	0	0	0
M7%	0	0	0	0	0	0	100	0	0	0	0	0
M8%	0	0	0	0	0	0	0	97	3	0	0	0
M9%	0	0	0	0	0	0	0	0	100	0	0	0
M10%	0	0	0	0	0	0	0	0	0	100	0	0
M11%	0	0	0	0	0	0	0	0	0	0	100	0
S%	0	0	0	0	0	0	0	0	0	0	0	100
Detection Accuracy(%)	84.08											
FAR	0.02											
MAR	0.44											
Diagnosis Accuracy(%)	68.32											

(a) DTW

SVM	M1%	M2%	M3%	M4%	M5%	M6%	M7%	M8%	M9%	M10%	M11%	S%
M1%	100	0	0	0	0	0	0	0	0	0	0	0
M2%	0	0	0	0	0	0	0	0	0	0	0	100
M3%	0	0	91	0	0	0	0	0	0	0	0	9
M4%	0	0	3	97	0	0	0	0	0	0	0	0
M5%	0	0	0	0	0	0	0	0	0	0	0	100
M6%	0	0	0	0	0	100	0	0	0	0	0	0
M7%	0	0	0	0	0	0	100	0	0	0	0	0
M8%	0	0	0	0	0	0	0	100	0	0	0	0
M9%	0	0	0	0	0	0	0	0	100	0	0	0
M10%	0	0	0	0	0	0	0	0	0	100	0	0
M11%	0	0	0	0	0	0	0	0	0	0	100	0
S%	0	0	0	0	0	0	0	0	0	0	0	100
Detection Accuracy(%)	90.29											
FAR	0.013											
MAR	0.182											
Diagnosis Accuracy(%)	82.11											

(b) SVM

HFD	M1%	M2%	M3%	M4%	M5%	M6%	M7%	M8%	M9%	M10%	M11%	S%
M1%	100	0	0	0	0	0	0	0	0	0	0	0
M2%	0	100	0	0	0	0	0	0	0	0	0	0
M3%	0	0	100	0	0	0	0	0	0	0	0	0
M4%	0	3	0	94	0	0	0	0	0	0	0	3
M5%	0	0	0	0	100	0	0	0	0	0	0	0
M6%	0	0	0	0	0	100	0	0	0	0	0	0
M7%	0	0	0	0	0	0	100	0	0	0	0	0
M8%	0	0	0	0	0	0	0	100	0	0	0	0
M9%	0	0	0	0	0	0	0	0	100	0	0	0
M10%	0	0	0	0	0	0	0	0	0	100	0	0
M11%	0	0	0	0	0	0	0	0	0	0	100	0
S%	0	0	0	0	0	0	0	0	0	0	0	100
Detection Accuracy(%)	99.43											
FAR	0.013											
MAR	0.013											
Diagnosis Accuracy(%)	98.67											

(c) HFD

References

1. Ardakani H D, Lucas C, Siegel D, Chang S, Dersin P, Bonnet B, Lee J. PHM for railway system-a case study on the health assessment of the point machines. IEEE Conference on Prognostics and Health Management 2012: 1-5, https://doi.org/10.1109/ICPHM.2012.6299533.
2. Asada T, Roberts C, Koseki T. An algorithm for improved performance of railway condition monitoring equipment: alternating-current point machine case study. Transportation Research Part C: Emerging Technologies 2013; 30: 81-92, https://doi.org/10.1016/j.trc.2013.01.008.
3. Atamuradov V, Camci F, Baskan S, Sevkli M. Failure diagnostics for railway point machines using expert systems. IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives 2009: 1-5, https://doi.org/10.1109/DEMPED.2009.5292755.
4. Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002; 24: 603-619, https://doi.org/10.1109/34.1000236.
5. Comaniciu D, Ramesh V, Meer P. The variable bandwidth mean shift and data-driven scale selection. Proceedings 8th IEEE International Conference on Computer Vision ICCV 2001: 438-445, https://doi.org/10.1109/ICCV.2001.937550.
6. Eker O, Camci F, Kumar U. SVM based diagnostics on railway turnouts. International Journal of Performability Engineering 2012; 8: 289-298.
7. Eker O F, Camci F, Guclu A, Yilboga H, Sevkli M, Baskan S. A simple state-based prognostic model for railway turnout systems. IEEE Transactions on Industrial Electronics 2011; 58: 1718-1726, https://doi.org/10.1109/TIE.2010.2051399.
8. Fasshauer G E. Positive definite kernels: past, present and future. Dolomite Research Notes on Approximation 2011; 4: 21-63.
9. Hartigan J A, Wong M A. Algorithm AS 136: A K-means clustering algorithm. Journal of the Royal Statistical Society Series C 1979; 28: 100-108, https://doi.org/10.2307/2346830.
10. He Y. Research on fault diagnosis method of high-speed railway turnouts. Master's Thesis. Beijing: Beijing Jiaotong University, 2014.
11. Keogh E J, Pazzani M J. Derivative dynamic time warping. Proceedings of the 2001 SIAM International Conference on Data Mining 2001: 1-11, https://doi.org/10.1137/1.9781611972719.1.
12. Kim H, Sa J, Chung Y, Park D, Yoon S. Fault diagnosis of railway point machines using dynamic time warping. Electronics Letters 2016; 52: 818-819, https://doi.org/10.1049/el.2016.0206.
13. Lee J, Choi H, Park D, Chung Y, Kim H-Y, Yoon S. Fault detection and diagnosis of railway point machines by sound Analysis. Sensors 2016; 16: 549, https://doi.org/10.3390/s16040549.
14. Letot C, Dersin P, Pugnaloni M, Dehombreux P, Fleurquin G, Douziech C, La-Cascia P. A data driven degradation-based model for the maintenance of turnouts: a case study. IFAC-PapersOnLine 2015; 48: 958-963, https://doi.org/10.1016/j.ifacol.2015.09.650.
15. Lu Q. Method of turnout fault diagnosis based on grey correlation analysis. MSc Thesis. Beijing, China: Beijing Jiaotong University, 2015.
16. Márquez F P G, Roberts C, Tobias A M. Railway point mechanisms: condition monitoring and fault detection. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 2010; 224: 35-44, https://doi.org/10.1243/09544097JRRT289.
17. Marquez F P G, Schmid F, Collado J C. A reliability centered approach to remote condition monitoring. A railway points case study. Reliability Engineering & System Safety 2003; 80: 33-40, https://doi.org/10.1016/S0951-8320(02)00166-7.
18. Mills T C. Time series techniques for economists. Cambridge, MA: Cambridge University Press, 2012.
19. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. Scikit-learn: machine learning in Python. Journal of Machine Learning Research 2011; 12: 2825-2830.
20. Ren Z, Sun S, Zhai W. Study on lateral dynamic characteristics of vehicle/turnout system. Vehicle System Dynamics 2005; 43: 285-303, https://doi.org/10.1080/00423110500083262.
21. Shaw D C. A universal approach to points condition monitoring. 4th IET International Conference on Railway Condition Monitoring 2008: 1-6, https://doi.org/10.1049/ic:20080315.
22. Wang G, Xu T, Tang T, Yuan T, Wang H. A Bayesian network model for prediction of weather-related failures in railway turnout systems. Expert Systems with Applications 2017; 69: 247-256, https://doi.org/10.1016/j.eswa.2016.10.011.
23. Zhang K. The railway turnout fault diagnosis algorithm based on BP neural network. IEEE International Conference on Control Science and Systems Engineering 2014: 135-138, https://doi.org/10.1109/CCSSE.2014.7224524.
24. Zhang K, Du K, Ju Y. Algorithm of railway turnout fault detection based on PNN neural network. 7th International Symposium on Computational Intelligence and Design 2014: 544-547, https://doi.org/10.1109/ISCID.2014.140.
25. Zhao H, Chen H, Dong W, Sun X, Ji Y. Fault diagnosis of rail turnout system based on case-based reasoning with compound distance methods. Chinese Control and Decision Conference (CCDC) 2017: 4205-4210, https://doi.org/10.1109/CCDC.2017.7979237.
26. Zhao L, Lu Q. Method of turnout fault diagnosis based on grey correlation analysis. Journal of the China Railway Society 2014; 36: 69-74.
27. Zhou F, Xia L, Dong W, Sun X, Yan X, Zhao Q. Fault diagnosis of high-speed railway turnout based on support vector machine. IEEE International Conference on Industrial Technology (ICIT) 2016: 1539-1544, https://doi.org/10.1109/ICIT.2016.7474989.
28. Zhou J-L, Lei Y. Paths between latent and active errors: analysis of 407 railway accidents/incidents' causes in China. Safety Science 2017, https://doi.org/10.1016/j.ssci.2017.12.027.

Dongxiu OU
Maojie TANG
Rui XUE
The Key Laboratory of Road and Traffic Engineering, Ministry of Education
China School of Transportation Engineering, Tongji University
Shanghai, 201804, P.R. China

Hongjing YAO
Jinan Railway Bureau, Jinan, Shangdong, China, 250000, P.R. China

Emails: ou.dongxiu@tongji.edu.cn, 1632486@tongji.edu.cn, xuerui@tongji.edu.cn, 13954123465@139.com

Article citation info:
WANG L, PEI Z, ZHU H, LIU B. Optimising extended warranty policies following the two-dimensional warranty with repair time threshold. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 523–530, http://dx.doi.org/10.17531/ein.2018.4.3.

Liying WANG
Zhaona PEI
Huihui ZHU
Baoyou LIU

OPTIMISING EXTENDED WARRANTY POLICIES FOLLOWING THE TWO-DIMENSIONAL WARRANTY WITH REPAIR TIME THRESHOLD

OPTYMALIZACJA POLITYKI GWARANCYJNEJ W OKRESIE PO WYGAŚNIĘCIU DWUWYMIAROWEJ GWARANCJI Z USTALONĄ GÓRNĄ GRANICĄ CZASU NAPRAWY

This paper considers an optimal extended warranty policies after the expiration of base two-dimensional warranty with repair time threshold. During the base two-dimensional warranty period, each failure of the equipment can be either replaced or minimally repaired depending on a pre-specified repair time threshold. After the base warranty expires, the length of an extended warranty policy is available for selection. The equipment is minimally repaired on each failure during the extended warranty. In this study, the length of the extended warranty period is optimized by minimizing the expected cost rate incurred over the whole warranty coverage, from the views of customs and manufacturers respectively. For the purpose of illustration, we present and discuss some numerical examples. The effect of repair time threshold on the optimal strategy is also investigated numerically.

Keywords: *extended warranty, repair time threshold, two-dimensional warranty, renewing warranty.*

W niniejszej pracy rozważano optymalną politykę przedłużania gwarancji po wygaśnięciu podstawowej gwarancji dwuwymiarowej z ustaloną górną granicą czasu naprawy. W podstawowym okresie obowiązywania gwarancji dwuwymiarowej, po każdej awarii urządzenie zostaje poddane minimalnej naprawie, lub – jeśli naprawa nie może być wykonana we wcześniej ustalonym czasie naprawy – wymienione. Po wygaśnięciu gwarancji podstawowej, konieczne jest wybranie długości okresu obowiązywania gwarancji rozszerzonej. Podczas trwania okresu gwarancji przedłużonej, sprzęt naprawia się w sposób minimalny (naprawa minimalna) po każdorazowym uszkodzeniu. W niniejszej pracy, optymalizowano długość przedłużonego okresu gwarancyjnego poprzez minimalizację oczekiwanych kosztów poniesionych podczas całego okresu trwania gwarancji; optymalizację przeprowadzono z perspektywy klienta jak i producenta . Dla ilustracji, przedstawiono i omówiono wybrane przykłady numeryczne. Przeprowadzono także analizę numeryczną wpływu górnej granicy czasu naprawy na optymalną strategię gwarancyjną.

Słowa kluczowe: *przedłużona gwarancja, górna granica czasu naprawy, gwarancja dwuwymiarowa, odnowienie gwarancji.*

1. Introduction

Along with increasing the warranty period for complex systems, reducing the warranty servicing costs has become an issue of great importance to the manufacturers. One possible way to reduce the expected warranty servicing cost is by making sound decision on the product warranty and maintenance strategies [21]. A great number of warranty and maintenance models have been proposed in the literature. Blischke and Murthy [1, 2, 3] reviewed a variety of warranty policies and their cost models among them. Shafiee and Chukova [21] reviewed the literature published between 2001 and 2011. This paper is the first identifiable academic literature review to deal with warranty and maintenance and suggested a possible classification of the mathematical models employed in this area of research.

In recent years, warranties and maintenance models for various products such as operating in discrete time [7], products with different failure rate [5], products with competing failures [22], products from heterogeneous population [11] are discussed. Optimizations of warranty and maintenance policies for various system such as parallel systems [13], systems with multiple-failure-mode [14], systems with reliability thresholds [15] and so on are considered. Several two-phase warranty policies(Jung et al. [10], Wang and Su [26], Su and Wang [23]) are proposed.

An increasing number of literatures focus on two-dimensional warranties and maintenance models recently(see Tong et al. [24], Huang et al. [8], Wang et al. [27], Cheng et al.[6], Wang and Su [26], Su and Wang [23], Peng et al. [20], Huang et al. [9], Wang et al. [25]). The traditional two-dimensional warranty policies are characterized by a region on the plane, where the axes represent the age and the usage of a system. These models under the traditional two-dimensional warranty policies have widely usefulness in various situations. It may be difficult for many products to obtain their usage information, but their failure and repair time information can be obtained conveniently by manufacturers. Furthermore many states in USA have acted a regulation named "lemon laws" to protect the customers who purchased defective vehicles, either new or used. The lemon law [4,12] regulates that, when the manufacturer or its authorized service agent cannot conform a vehicle to its warranty either by repairing or correcting any nonconformity within a repair time threshold or after a reasonable number of attempts, the manufacturer must replace the vehicle. The policy of warranty services for repair, replacement and refund in China has similar terms. Based on the aforementioned background, another two-dimensional warranty policy with repair threshold that is different from the classical one is introduced by Park et al. [16]. Under the two-dimensional warranty policy both failure time and repair time are considered simultaneously to determine the warranty action upon the equipment's failure. Manufacturers specify a repair time threshold in advance to set the limit on the repair time for the failed equipment. In the case that repair service can't be completed within the repair time threshold, the failed equipment is to be replaced by a new one carrying with the same original warranty terms instead of continuing the repair works. Otherwise, only minimal repair is conducted. Such a warranty policy is referred to as ''two-dimensional warranty policy with repair threshold'' throughout this paper. They obtained the optimal maintenance policy following the expiration of the warranty period by minimizing the expected cost rate per unit time during the life cycle of the system, from the customer's perspective.

The two-dimensional warranty policy with repair time threshold receives a great attention from many researchers due to its practical applications in automobile industry. From the view of manufacturers, Park et al. [17] determined the optimal warranty period to minimize the expected cost rate during the warranty. Park and Pham [18] proposed cost models for age replacement and block replacement policies under the two-dimensional warranty. Park et al. [19] proposed a periodic preventive maintenance policy after the expiration of the two-dimensional warranty.

Despite the fact manufacturers offer various types of warranty policies at the sale of an equipment, it is necessary for customers to purchase extended warranty services for safeguarding equipment maintenance. Furthermore the performing of extended warranty services is convenient and free from worry for customers. So many customers chose to purchase an extended warranty, instead of planning the maintenance activities after the expiration of a base warranty themselves. In this case the length of the extended warranty period is important from a customer's perspective. A long extended warranty period may be more costly, but it can be more cost-effective in the long run as any failures during the extended warranty will be served by the warranty provider. On the other hand, if an extended warranty service is scheduled for a period of time that is excessively long, the risk of potential costs for the manufacturer may increase. Therefore, establishing an optimal extended warranty policy would be crucial, particularly from the cost perspective.

In this study, we assume the pro-rata renewal two-dimensional warranty with repair time threshold is adopted during the base warranty. We also assume that the length of the extended warranty policies is available for selection. The length of the extended warranty period is optimized to minimize the expected cost per unit time over the whole warranty coverage.

The remainder of this paper is organized as follows. In Section 2, model formulation is given. Cost structures, from the view of customers and manufacturers respectively, are presented. In Section 3, the explicit expressions for the expected cost rates over the whole warranty coverage, from the views of customers and manufacturers respectively, are given. The optimal solution for the extended warranty period is obtained by minimizing the expected cost rate. In Section 4 numerical examples are presented for illustrative purpose, assuming the Weibull failure and repair times. Concluding remarks are given in Section 5.

Nomenclature

pdf, *cdf*	probability density function, cumulative distribution function, respectively
i.i.d	independent, identically distributed
r. v.	random variable
ECR	expected cost rate
MRR	minimal repair-replacement
$RMRR$	renewable minimal repair-replace
T , Y	failure time and repair time, respectively
$f(t)$, $F(t)$, $\bar{F}(t)$	density function, distribution function, and reliability function of T , respectively
$g(y)$, $G(y)$, $\bar{G}(y)$	pdf, cdf and reliability function of Y , respectively
C_{U_r}	total replacement cost the customer is responsible for during the based warranty period
C_{Uf}	total failure stoppage cost the customer is responsible for during the warranty period
C_{Mr}	total replacement cost the manufacturer is responsible for during the warranty period
C_{Mm}	total minimal repair cost the manufacturer is responsible for during the warranty period
c_r	unit cost of replacement
c_{Uf_1}	unit cost of replacement stoppage during the based warranty period
c_{Uf_2}	unit cost of minimal repair stoppage during the based warranty period
c_{Uf_3}	unit cost of minimal repair stoppage during the extended warranty period
c_{Mf_2}	unit cost of minimal repair during the base warranty period
c_{Mf_3}	unit cost of minimal repair during the extended warranty period
w	the length of the original warranty period
c_e	unit cost for unit length of extended warranty period
L'	unit length of extended warranty period
r_0	repair time threshold

2. Model formulation

The warranty policy considered in this paper is a combination of base and extended warranties. The equipment is sold with an original warranty period w. The warranty term works as follows.

During the warranty period, a renewable minimal repair-replace (RMRR) warranty policy is performed. Under the RMRR policy, the manufacturer starts the minimal repair for the failed equipment immediately. The manufacturer sets the repair time threshold, r_0, in advance, which works as the time limit for the minimal repair service for the customer's satisfaction. If manufacturer successfully provides the minimal repair to fix the failed equipment within the repair time limit, the warranty would be effective only in the remaining warranty period and it would not be renewed. On the other hand the warranty policy is renewed for the replaced equipment with exactly the same warranty terms as the original one.

Once the base warranty expires, the customer purchases extended warranty service for a time period kL', where L' is a given time period and k $(k = 0,1,2,\cdots)$ can be optimized. During the extended warranty period, all failures (whose repair time can be within or not within the limit) are corrected with minimal repairs. The possible system state evolution path for such a warranty policy is shown in Fig. 1.

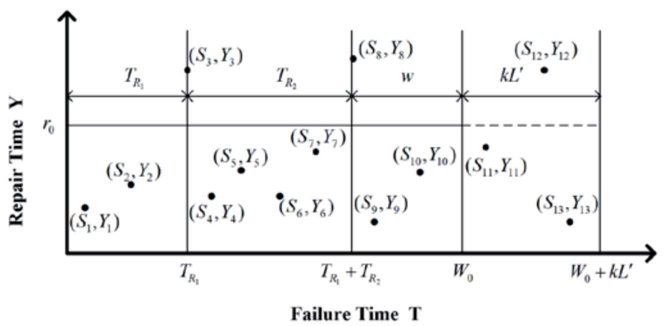


Fig. 1. A possible system state evolution path with two times of renewals

In Fig. 1, $(S_i,Y_i)(i=1,2,\cdots,14)$ denote time instants of failures and times for repairs. The repair times for the 1th and 2th failures don't exceed the threshold r_0. While the 3th one exceeds it. Furthermore the total times for the three failures don't exceed the original warranty period w. So the equipment is replaced at time S_3 and the time for the 1th renewal of the warranty terms is T_{R_1}. Similarly, the equipment is replaced at the 8th failure and the inter-arrival time between the 1th and 2th renewal is T_{R_2}. During the following w times, the 9th and the 10th failures happen and their repair times don't exceed the threshold. Hence the base warranty period ends and the base warranty period is $T_{R_1}+T_{R_2}+w$. We denoted by W_0. Over the extended warranty period, the 11th, the 12th and the 13th failures happen and their repair times can exceed or not exceed the threshold. The warranty period is defined from the purchasing point of the equipment to the end of the extended warranty time at which the customer buy kL' extended warranty period after the based warranty expires, as shown in Fig.2.

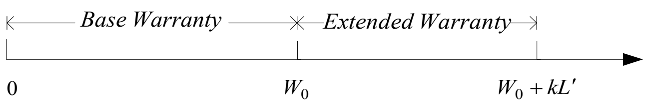


Fig. 2. The warranty span with k extended warranty periods

Length of time necessary to repair/replace the failed system is negligible and not included within the warranty period. Under the renewing PMRR, the customer is responsible for the pro-rated replacement cost. Failure cost, which incurs for each failure due to the stoppage of system operation, is charged to the customer during the based and extended warranty. Minimal repair is performed at no charge to the customer during the based and extended warranty period. We assume that all the warranty claims are valid and accepted.

From the view of customers and manufacturers respectively, the number of the extended warranty periods k is optimized by minimizing the expected cost rate (ECR) per unit time over the whole warranty coverage in this paper.

3. Formulation of the expected cost rate

3.1. Expected length of the warranty period

Let T_R be the time interval between two adjacent replacements. According assumptions in Section 2, the warranty terms would be renewed if $T_R < w$. Otherwise the warranty policy would not be renewed and the based warranty period will end. Let W_0 denote the length of the base warranty. It is clear that W_0 is a r. v. depending on the total number of warranty renewals, the inter-arrival times between two successive renewals of warranty terms and the length of original warranty period. Then, the based warranty period W_0 can be expressed as:

$$W_0 = T_{R_1} + T_{R_2} + \cdots T_{R_{N_R}} + w, \tag{1}$$

where N_R is the number of renewals during the base warranty period, $T_{R_i}(i=1,2,\cdots,N_R)$ are the time intervals between two subsequent warranty renewals and they are independent and identically distributed non-negative random variables.

After the based warranty expires, the customer purchases k units of extended warranty periods, then the warranty period $L(k)$ can be expressed as:

$$L(k) = W_0 + kL'. \tag{2}$$

From Eq. (1), Eq. (2) can be rewritten as:

$$L(k) = T_{R_1} + T_{R_2} + \cdots + T_{R_{N_R}} + w + kL'. \tag{3}$$

Let T be the r. v. denoting the failure time of the equipment having $f(t)$, $h(t)$, $F(t)$, $\bar{F}(t)$ as its probability density function, hazard function, cumulative distribution function and reliability function, respectively. Let $T_i(i=1,2,\cdots)$ denote the time intervals between the $i-1$ th and i th failures of the equipment, then $T_i(i=1,2,\cdots)$ are i.i.d. continuous with the same distribution to T and $T_i = S_i - S_{i-1}(i=1,2,\cdots,S_0=0)$, where S_i is the time that the i th failure occurs. We assume that $Y_i(i=1,2,\cdots)$ is the repair time for the i th failure. They are assumed to be i.i.d. continuous r.v.'s having $G(y)$ and $\bar{G}(y)$ as its cdf and reliability function, respectively. According to assumptions in Section 2, upon failures, the system is replaced with probability $\bar{G}(r_0)$ and corrected minimally with probability $G(r_0)$.

According to Block et al. [4], the cumulative distribution function of the time to minimal repairs is given by:

$$F_M(t) = 1 - \exp[-\int_0^t G(r_0)h(u)du]. \tag{4}$$

The cdf of the time interval between two subsequent replacements (T_R), is:

$$F_R(t) = 1 - \exp[-\int_0^t \overline{G}(r_0)h(u)du]. \tag{5}$$

Deriving on Eq. (5), the pdf of T_R , can be given by:

$$f_R(t) = \overline{G}(r_0)h(t)\exp[-\int_0^t \overline{G}(r_0)h(u)du]. \tag{6}$$

Since the warranty terms would be renewed when $T_R < w$, the time interval between two subsequent warranty renewals $T_{R_i}(i = 1,2,\cdots)$ is the truncated distribution of T_R over the interval $(0,w)$ and its pdf is $f_R(t)/F_R(w)(0 < t < w)$. Note that this expression is different from Eq.(2) of Park et al. [16].

Given $N_R = n$, the conditional expected length of based warranty period can be represented as:

$$E(W_0 \mid N_R = n) = \sum_{j=1}^{n} E(T_{R_j}) + w, \tag{7}$$

where $T_{R_i}(i = 1,2,\cdots,n)$ are i.i.d. non-negative random variables. Eq. (7) can be written as:

$$E(W_0 \mid N_R = n) = n\frac{\int_0^w tf_R(t)dt}{F_R(w)} + w. \tag{8}$$

According to assumptions in Section 2, the warranty terms would be renewed if the time interval between two adjacent replacements $T_R < w$. Thus, if $N_R = n$, then the event $T_R < w$ occurs in the first n trials and the event $T_R > w$ occurs in the following $n+1$ th trial. Since these events are independent of each other,

$$P(N_R = n) = (F_R(w))^n(1 - F_R(w)), n = 0,1,2,\cdots. \tag{9}$$

Note that this expression is also different from Eq. (4) in Park et al. [16].

By taking the expectation for the conditional expectation of Eq. (8) with respect to N_R , we obtain the following expected length of the based warranty period:

$$\begin{aligned} E(W_0) &= E(E(W_0 \mid N_R)) \\ &= \sum_{n=0}^{\infty} P(N_R = n)E(W_0 \mid N_R = n) \\ &= \sum_{n=0}^{\infty} (n\int_0^w tf_R(t)dt / F_R(w) + w)(F_R(w))^n(1 - F_R(w)) \\ &= \frac{1}{1 - F_R(w)}\int_0^w tf_R(t)dt + w. \end{aligned} \tag{10}$$

From Eq.(10), the expected length of whole warranty period is:

$$E(L(k)) = E(W_0) + kL' = \frac{1}{1 - F_R(w)}\int_0^w tf_R(t)dt + w + kL'. \tag{11}$$

3.2. Expected cost rate from the view of customers

In this section, we will derive the expected cost rate per unit time during the warranty period of the equipment from the view of customers. Let C_{Ur} denote total replacement cost that the customer is responsible for during the base warranty period. Assume that C_{Uf} represents the total failure stoppage cost during the base and extended warranty that charged to the customer. Further, let c_e be the fixed unit cost for unit length of extended warranty period. Then, the user would be charged the total amount of cost equaling $C_{Ur} + C_{Uf} + kc_e$.

Under the renewing PMRR, the customer is responsible for prorated replacement cost during the based warranty period and thus, the replacement cost can be expressed as a function of T_{R_i} 's as (see Park et al. [16]),

$$C_{Ur} = \sum_{j=1}^{N_R} c_r\frac{T_{R_j}}{w}, \tag{12}$$

where c_r is the unit cost of replacement.

Given $N_R = n$, the conditional replacement cost can be represented as:

$$E(C_{Ur} \mid N_R = n) = \frac{c_r}{w}n\frac{\int_0^w tf_R(t)dt}{F_R(w)}. \tag{13}$$

Consequently, the total expected replacement cost during the based warranty period:

$$\begin{aligned} E(C_{Ur}) &= E(E(C_{Ur} \mid N_R)) \\ &= \sum_{n=0}^{\infty} P(N_R = n)E(C_{Ur} \mid N_R = n) \\ &= \sum_{n=0}^{\infty} P(N_R = n)\sum_{j=1}^{n} c_r\frac{E(T_{R_j})}{w} \\ &= \frac{c_r}{w}\sum_{n=0}^{\infty}(F_R(w))^n(1 - F_R(w))n\frac{\int_0^w tf_R(t)dt}{F_R(w)} \\ &= \frac{c_r}{w(1 - F_R(w))}\int_0^w tf_R(t)dt. \end{aligned} \tag{14}$$

Under the renewable MRR warranty policy, the expected total number of replacements can be given by:

$$\begin{aligned} E(N_R) &= \sum_{n=0}^{\infty} nP(N_R = n) \\ &= \sum_{n=0}^{\infty} n(F_R(w))^n(1 - F_R(w)) \\ &= \frac{F_R(w)}{(1 - F_R(w))}. \end{aligned} \tag{15}$$

Conditioning on the possible time between two subsequent warranty renewals, the expected total number of minimal repairs between subsequent warranty renewals can be given by:

$$E(N_{M_r}) = \int_0^w \int_0^t G(r_0)h(u)f_R(t) / F_R(w)dudt. \tag{16}$$

From Eq. (16), the expected total number of minimal repairs during the N_R times renewals,

$$\begin{aligned} E(N_{M_R}) &= E(E(N_{M_R} \mid N_R)) \\ &= \sum_{n=0}^{\infty} P(N_R = n) E(N_{M_R} \mid N_R = n) \\ &= \sum_{n=0}^{\infty} P(N_R = n) n E(N_{M_r}) \\ &= E(N_{M_r}) \sum_{n=0}^{\infty} P(N_R = n) n \\ &= E(N_{M_r}) E(N_R) \\ &= \frac{1}{1-F_R(w)} \int_0^w \int_0^t G(r_0) h(u) f_R(t) du dt. \end{aligned} \tag{17}$$

During the last base warranty period w, minor repairs are performed with intensity function $G(r_0)h(t)(0<t<w)$, consequently, the expected total number of minor repairs is $\int_0^w G(r_0)h(t)dt$.

In sum, the expected total number of minor repairs over the based warranty period is:

$$E(N_M) = \frac{1}{(1-F_R(w))} \int_0^w \int_0^t G(r_0)h(u) f_R(t) du dt + \int_0^w G(r_0) h(u) du \tag{18}$$

During the extended warranty period kL', minimal repair service is conducted on each failure. Consequently, the expected total number of minor repairs $E(N_e) = \int_w^{w+kL'} h(u)du$.

The expected total failure cost due to the stoppage of system operation during the based and extended warranty period is:

$$E(C_{Uf}) = c_{Uf_1} E(N_R) + c_{Uf_2} E(N_M) + c_{Uf_3} E(N_e) \tag{19}$$

where c_{Uf_1} and c_{Uf_2} are unit cost of replacement and minimal repair stoppage time during the based warranty period, respectively. c_{Uf_3} $(c_{Uf_2} < c_{Uf_3} < c_{Uf_1})$ is the unit cost of minimal repair stoppage time during the extended warranty period.

From Eq.(14), Eq.(15), Eq.(17), Eq.(18), Eq.(19), the expected cost rate over the whole warranty period can be formulated as:

$$ECR_U(k) = \frac{E(C_{Ur}) + c_{Uf_1} E(N_R) + c_{Uf_2} E(N_M) + c_{Uf_3} E(N_e) + kc_e}{\frac{1}{1-F_R(w)} \int_0^w t f_R(t) dt + w + kL'}. \tag{20}$$

3.3. Expected cost rate from the view of manufacturers

Let C_{Mr}, C_{Mm} be the r.v.'s representing total replacement cost, total minimal repair cost incurred during the warranty period that the manufacturer is responsible for, respectively. Then (see Park et al. [16]):

$$C_{Mr} = \sum_{j=1}^{N_R} c_r \frac{w - T_{R_j}}{w} \tag{21}$$

According to Eq. (13) and Eq. (15):

$$E(C_{Mr}) = \frac{c_r (wF_R(w) - \int_0^w t f_R(t) dt)}{w(1-F_R(w))}. \tag{22}$$

From Eq.(15) and Eq.(16), the expected total minor repair cost over the whole warranty period can be given as:

$$\begin{aligned} E(C_{Mm}) = c_{Mf_2} \frac{1}{1-F_R(w)} \int_0^w \int_0^t G(r_0) h(u) f_R(t) du dt + \\ c_{Mf_2} \int_0^w G(r_0) h(u) du + c_{Mf_3} \int_w^{w+kL'} h(u) du. \end{aligned} \tag{23}$$

where c_{Mf_2} and c_{Mf_3} ($c_{Mf_2} < c_{Mf_3} < c_r$) are unit costs of minimal repair during the base and the extended warranty period, respectively. The first term in Eq. (23) corresponds to the expected cost incurred during the renewal based warranty period. The second term is those during the non-renewal based warranty period.

The expected cost rate over the whole warranty period can be formulated as:

$$ECR_M(k) = \frac{E(C_{Mr}) + E(C_{Mm})}{\frac{1}{1-F_R(w)} \int_0^w t f_R(t) dt + w + kL'}. \tag{24}$$

Substituting Eq's (22) and (23) into Eq.(24), we can obtain the expression of $ECR_M(k)$.

4. Numerical experiments

4.1. Optimisation of the length of the extended warranty

This section presents numerical examples to illustrate the optimal warranty policy model derived in Section 3. The failure time and the repair time of the equipment are assumed to follow Weibull distributions with the failure rate function $h(t) = (\beta/\eta)(t/\eta)^{\beta-1} (\beta, \eta > 0)$ and cdf $G(t) = 1 - e^{-(t/\lambda)^{\nu}} (\lambda, \nu > 0)$ respectively. The parameter values we set for this particular numerical example are listed in Table 1.

For this particular numerical example, we set $w = 6$ (unit: year) and $r_0 = 1/12$ (unit: year). This implies that if the failed equipment requires the repair time of more than one month, it is rather replaced by a new one instead of being repaired. In practice, the extended warranty of an equipment might has a limit number of possibility. Hence, we consider situations where $L' = 0.25$, $k = \{0, 1, \cdots, 12\}$. Under such a situation, the customer purchases k extended warranties, each with the length of 0.25 years and the longest extended warranty is 3 years.

From the view of customers, the cost rate is decided by unit cost for unit length of extended warranty period (c_e), unit cost of minimal repair stoppage time(c_{Uf_3}) and the length of the extended warranty period after the based warranty expires. So we determine optimal extended warranties for various c_e and c_{Uf_3} by using Matlab software. See Table 2 for detail. The optimal length of extended warranty period

Table 1. Parameter values assumed for our numerical example

w	L'	c_r	c_{Uf_1}	c_{Uf_2}	c_{Mf_2}	r_0	β	η	ν	λ
6	0.25	5.5	0.05	0.02	0.85	1/12	1.097	0.241	0.787	0.033

is denoted by k^* and its corresponding expected cost rate is denoted by $ECR_U(k^*)$.

Table 2 . Optimal extended warranties for various c_e and c_{Uf_3} (from the view of customers)

c_e	c_{Uf3}	k^*	$ECR_U(k^*)$
0.01	0.035	8	0.988291
	0.034	12	0.987665
	0.033	12	0.98696
	0.032	12	0.986256
	0.031	12	0.985551
0.02	0.035	3	0.988809
	0.034	7	0.988537
	0.033	11	0.988019
	0.032	12	0.987319
	0.031	12	0.986615
0.03	0.035	0	0.988945
	0.034	2	0.988895
	0.033	5	0.988707
	0.032	9	0.98832
	0.031	12	0.987679
0.04	0.035	0	0.988945
	0.034	0	0.988945
	0.033	1	0.98894
	0.032	3	0.988835
	0.031	7	0.988561

For instance, $k^* = 2$ and $ECR_U(k^*)$ =0.988876 when c_e =0.01 and c_{Uf_3} =0.035 from Table 2. This indicates that the expected cost rate is minimized when the customer buys 8 units extended warranty, and the customer's cost rate becomes 0.988291 per year over the whole warranty period of the equipment in this situation. In some cases, such as c_e =0.01 and c_{Uf_3} =0.034, c_e =0.02 and c_{Uf_3} =0.032, c_e =0.03 and c_{Uf_3} =0.031 and so on, the cost rate is minimized by buying the longest extended warranty periods. However, in some cases, such as c_e =0.01 and c_{Uf_3} =0.034, c_e =0.02 and c_{Uf_3} =0.032, c_e =0.03 and c_{Uf_3} =0.031 and so on, the cost rate is minimized when customers don't buy extended warranty services.

Similarly, we can obtain the optimal extended warranties for various c_e and c_{Mf_3}, from the view of manufacturers. See Table 3 for detail. The optimal length of extended warranty period and its corresponding expected cost rate are denoted by k^{**} and $ECR_M(k^{**})$ respectively.

Table 3. Optimal extended warranties for various c_e and c_{Mf_3} (from the view of manufacturers)

c_e	c_{Mf_3}	k^{**}	$ECR_M(k^{**})$
0.01	1	10	6.427558
	1.01	7	6.428713
	1.02	4	6.429429
	1.03	1	6.429752
	1.04	0	6.429782
	1.05	0	6.429782
0.02	1	12	6.471456
	1.01	9	6.472935
	1.02	6	6.473955
	1.03	3	6.474555
	1.04	0	6.474781
	1.05	0	6.474781
0.03	1	12	6.515258
	1.01	11	6.516934
	1.02	8	6.518271
	1.03	5	6.519163
	1.04	2	6.519652
	1.05	0	6.519779
0.04	1	12	6.559059
	1.01	12	6.560748
	1.02	10	6.562367
	1.03	7	6.563567
	1.04	4	6.564338
	1.05	1	6.564725

Table 4. Optimal extended warranties for various r_0 (from the view of customers)

r_0	k^*	$ECR_U(k^*)$
1/11.4	4	0.971906
1/11.6	5	0.977926
1/11.8	6	0.983465
1/12	7	0.988561
1/12.2	7	0.993246
1/12.4	8	0.997554
1/12.6	9	1.001518

4.2. The effect of the repair time threshold on the optimal strategy

To analysis the effect of the repair time threshold r_0 on the optimal strategy, we change it from 1/11.4 to 1/12.6 and set w=6, $L' = 0.25$, $c_r = 5.5$, $c_e = 0.04$, $c_{Uf_1} = 0.05$, $c_{Uf_2} = 0.02$, $c_{Uf_3} = 0.031$, c_{Mf_2}=0.85, c_{Mf_3}=1.02. The optimal strategies and its corresponding expected cost rates are presented in Tables 4 and 5. The tables indicate that the value of r_0 has some effect on k^*, k^{**}, $ECR_U(k^*)$ and $ECR_M(k^{**})$. As r_0 decreases, the length of the optimal extended warranty and the corresponding expected cost rate increase from both the views of customers and manufacturers. It is due to the fact that as

Table 5. Optimal extended warranties for various r_0 (from the view of manufacturers)

r_0	k^{**}	$ECR_M(k^{**})$
1/11.4	0	6.336242
1/11.6	2	6.39698
1/11.8	5	6.457649
1/12	8	6.518271
1/12.2	11	6.578851
1/12.4	12	6.639496
1/12.6	12	6.700443

r_0 decreases, the chance of the failed equipment is replaced increases and so has a higher cost.

5. Conclusions

An extended warranty decision model considering the "lemon law" is built in this paper. It is of interest and practical especially for the products including automobiles, etc. Under the two-dimensional renewal repair-replacement warranty policy, each failure of the system can be either replaced or minimally repaired depending on a repair time threshold. We obtained the optimal extended warranty period, from the views of customers and manufacturers respectively, so that the expected cost rate over the whole warranty coverage, is optimized. By assuming Weibull distributions for failure and repair times, numerical examples are discussed to demonstrate the applicability of the methodology derived in the paper. Numerical example indicate the larger the repair time thresholds are, the smaller are the expected cost rates.

The repair time the threshold in this study is a constant. However in practice it may be changed with the degradation of the equipment. For customer's satisfaction, the time limits of repair services are shorter at the early stage of the warranty period. So the extended warranty models with different repair time thresholds would be an interesting future research.

Acknowledgements

This work is supported partly by the NSF of China Grants (11471218) and universities in Hebei province science and technology research project (ZD2018073).

References

1. Blischke W and Murthy D. Product warranty management-I: A taxonomy for warranty policies. European Journal of Operational Research 1992; 62: 127–48, https://doi.org/10.1016/0377-2217(92)90242-2.
2. Blischke W . Warranty cost analysis. CRC Press, 1994.
3. Blischke W, Murthy D . Product warranty handbook. CRC Press, 1996.
4. Block H, Borges W, Savits T. Age-dependent minimal repair. Journal of Applied Probability 1985; 22 (2), 370–85, https://doi.org/10.2307/3213780.
5. California.lemon.law. http://www.lemonlawspecialists.com.
6. Chang W. Optimal single-replacement for repairable products with different failure rate under a finite planning horizon. International Journal of Systems Science 2015; 46 (6):1003-09, https://doi.org/10.1080/00207721.2013.807381.
7. Cheng Z, Yang Z, Zhao Z, Wang Y, Li Z . Preventive maintenance strategy optimizing model under two-dimensional warranty policy. Eksploatacja i Niezawodnosc-Maintenance and Reliability 2015; 17(3):365-72, https://doi.org/10.17531/ein.2015.3.6.
8. Chien Y, Zhang Z. Analysis of a hybrid warranty policy for discrete-time operating products. IIE Transactions 2015; 47(5):442-59, https://doi.org/10.1080/0740817X.2014.953645.
9. Huang Y, Gau W, Ho J. Cost analysis of two-dimensional warranty for products with periodic preventive maintenance. Reliability Engineering & System Safety 2015; 134:51-58, https://doi.org/10.1016/j.ress.2014.10.014.
10. Huang Y, Huang C, Ho J . A customized two-gimensional extended warranty with preventive maintenance. European Journal of Operational Research 2017; 257 (3): 971-78, https://doi.org/10.1016/j.ejor.2016.07.034.
11. Jung K, Park M, Dong H. Cost optimization model following extended renewing two-phase warranty. Computers & Industrial Engineering 2015; 79:188-94, https://doi.org/10.1016/j.cie.2014.10.016.
12. Lee H, Ji H, Finkelstein M. On information-based warranty policy for repairable products from heterogeneous population. European Journal of Operational Research 2016; 253(1):204-15, https://doi.org/10.1016/j.ejor.2016.02.020.
13. Lemon Law. https://en.wikipedia.org/wiki/Lemon_law.
14. Liao G . Optimal economic production quantity policy for a parallel system with repair, rework, free-repair warranty and maintenance. International Journal of Production Research 2016; 20:1-16, https://doi.org/10.1080/00207543.2016.1203074.
15. Li X, Jia Y, Wang P, Zhao J. Renewable warranty policy for multiple-failure-mode product considering different maintenance options. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2015; 17 (4): 551 - 60, https://doi.org/10.17531/ein.2015.4.10.
16. Lin Z, Huang Y, Fang C. Non-periodic preventive maintenance with reliability thresholds for complex repairable systems. Reliability Engineering & System Safety 2015; 136:145-56, https://doi.org/10.1016/j.ress.2014.12.010.
17. Park M, Jung K, Park D. Optimal post-warranty policy with repair time threshold for minimal repair. Reliability Engineering & System Safety 2013;111:147–53, https://doi.org/10.1016/j.ress.2012.10.017.
18. Park M, Jung K, Dong H. Optimal warranty policies considering repair service and replacement service under the manufacturer's perspective. Annals of Operations Research 2014; 1:1-16.
19. Park M, Pham H. Cost models for age replacement policies and block replacement policies under warranty. Applied Mathematical Modelling 2016; 40(9-10): 5689-5702, https://doi.org/10.1016/j.apm.2016.01.022.
20. Park M, Jung K, Park D. Optimization of periodic preventive maintenance policy following the expiration of two-dimensional warranty. Reliability Engineering & System Safety 2018; 170: 1–9, https://doi.org/10.1016/j.ress.2017.10.009.

21. Peng T, Song X, Liu Z. A maintenance strategy for two-dimensional extended warranty based on dynamic usage rate. International Journal of Production Research 2017; 3:1-17.
22. Shafiee M, Chukova S. Maintenance models in warranty: A literature review. European Journal of Operational Research 2013; 229:561–72, https://doi.org/10.1016/j.ejor.2013.01.017.
23. Shang L, Si S , Cai Z . Optimal maintenance–replacement policy of products with competing failures after expiry of the warranty. Computers & Industrial Engineering 2016; 98: 68-77, https://doi.org/10.1016/j.cie.2016.05.012.
24. Su C, X Wang. A two-stage preventive maintenance optimization model incorporating two-dimensional extended warranty. Reliability Engineering & System Safety 2016; 155: 169-178, https://doi.org/10.1016/j.ress.2016.07.004.
25. Tong P, Z Liu, F Men, L Cao. Designing and pricing of two-dimensional extended warranty contracts based on usage rate. International Journal of Production Research 2014; 52 (21): 6362-6380, https://doi.org/10.1080/00207543.2014.940073.
26. Wang J , Zhou Z , Peng H . Flexible decision models for a two-dimensional warranty policy with periodic preventive maintenance. Reliability Engineering & System Safety 2017; 162: 14–27, https://doi.org/10.1016/j.ress.2017.01.012.
27. Wang X, Su C. A two-dimensional preventive maintenance strategy for items sold with warranty. International Journal of Production Research 2016; 54 (19): 1-15, https://doi.org/10.1080/00207543.2016.1187314.
28. Wang Y, Liu Z , Liu Y. Optimal preventive maintenance strategy for repairable items under two-dimensional warranty. Reliability Engineering & System Safety 2015; 142 (5): 326-333, https://doi.org/10.1016/j.ress.2015.06.003.

Liying WANG
Zhaona PEI
Huihui ZHU
Baoyou LIU
Department of Mathematics & Physics
Shijiazhuang Tiedao Institute
Shijiazhuang, 050043, China

E-mails: wly_sjz@126.com; 946622468@qq.com; 2822537623@qq.com; liubyde@tom.com

Article citation info:
LEONARCIK R, URBANIAK M, DĘBKOWSKI R. Method for assessing the grinding wheels operational properties. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 531–541, http://dx.doi.org/10.17531/ein.2018.4.4.

Rafał LEONARCIK
Mirosław URBANIAK
Ryszard DĘBKOWSKI

METHOD FOR ASSESSING THE GRINDING WHEELS OPERATIONAL PROPERTIES

METODA OCENY WŁAŚCIWOŚCI EKSPLOATACYJNYCH ŚCIERNIC*

The paper presents a new, multi-criteria method which allows the numerical evaluation of the machining process in terms of efficiency, quality and costs. Three indicators were developed to assess the operational properties of grinding wheels. Their values are determined on the basis of the results of short grinding tests carried out on a special test stand. The evaluation of the proposed indicators is described. Furthermore, the application exemple of this method in determining the grinding wheel's operational properties is presented. In the research, the vitrified alumina oxide grinding wheels were used for grinding of constructional and tool steels of various hardness. The results of the experiments show that the proposed indicators are an effective tool for assessing the process and results of grinding for a specific grinding wheel and material within certain tested grinding parameters range. The study also showed that the differences in indicators' values, observed during tests of grinding specific material type using grinding wheels with different properties, are useful for optimizing the choice of tool type and machining conditions.

Keywords: *grinding, grinding wheel, operational properties, grinding wheel characteristics.*

W artykule przedstawiono nową, wielokryterialną metodę, w której grupa trzech wskaźników eksploatacyjnych, wyznaczonych na podstawie danych procesowych z krótkotrwałego testu pracy pary ściernica-przedmiot obrabiany przeprowadzonego na specjalnym stanowisku badawczym, pozwala na liczbowe wartościowanie procesu obróbki pod względem wydajności, jakości i kosztów. Przedstawiono wyniki badań ewaluacji zaproponowanych wskaźników oraz badania aplikacyjne metody w zakresie oceny właściwości eksploatacyjnych ściernic podczas szlifowania stali. W badaniach stosowano ceramiczne ściernice elektrokorundowe oraz stale narzędziowe i konstrukcyjne o różnej twardości. Wyniki doświadczeń wykazały, że zaproponowane formuły wskaźników są skutecznym narzędziem oceny przebiegu i wyników szlifowania dla określonej pary ściernica-materiał w badanym zakresie wartości nastawnych procesu szlifowania. Badania wykazały także, że różnice wartości wskaźników występujące podczas testów szlifowania określonego rodzaju materiału ściernicami o różnej charakterystyce, są przydatne do optymalizacji wyboru rodzaju narzędzia i warunków obróbki.

Słowa kluczowe: *szlifowanie, ściernica, właściwości eksploatacyjne, charakterystyka ściernicy.*

1. Introduction

Grinding is the basic process of finishing hardened objects and made of difficult-to-cut materials, which allows achieving high level of accuracy in terms of shape and dimension, low surface roughness and low tensile stress in the surface layer [1, 15]. The mentioned values of the machined surfaces are important for the durability and reliability of parts operating independently or as components of devices [35]. The compliance of the grinding results with the requirements given to the workpiece in the design phase is determined by the method of designing and performing the grinding operations. When designing the grinding process, many variables should be taken into account regarding the characteristics of the grinding wheel and its operating conditions, i.e. grinding parameters, the method of cooling the treatment zone, the frequency and way of performing the dressing treatment. Each of these factors has a significant impact on the final result of the treatment, because it affects the phenomena occurring in the contact zone of tribological pair, which is created between the grinding wheel and workpiece material. The impact of abrasive grains on a workpiece can be divided into three stages [8, 11]: elastic deformation and friction between tip and the material, material plastic deformation with the formation of flash and internal friction and chip formation. The share of each of the mentioned stages in the chip formation process depends on the properties of the workpiece, grinding parameters, friction conditions between the abrasive grains and the workpiece and the geometry of the abrasive grains [2, 11, 28]. As a result of the occurring friction, a significant part of the mechanical energy is largely transformed into heat [20, 23] causing a significant increase in the temperature of the surface layer of the object and its structural changes. Among the negative effects of the thermal impact of the grinding process following can be indicated: grinding burn on the surface layer, phase changes of the material, tempering (softening) with the possibility of repeated hardening of the surface layer, unfavourable residual tensile stresses, grinding cracks and reduced fatigue strength [4, 20, 26]. In order to eliminate these negative effects that the heat has on the workpiece and on the surface of the grinding wheel, the proper **c**ooling **l**iquid (CL), the method of its feeding to the treatment zone and the flow rate should be selected [3, 31]. The second element of the tribological pair – grinding wheel – is subject to wear due to various mechanisms, causing changes in the shape and properties of the tool during machining, therefore reducing the efficiency of the grinding operation and worsening of the quality of the workpiece. The main forms of wear of the grinding wheel surface are abrasive wear and chipping. The vertex of abrasive grains become dulled due to abrasive wear caused by friction on the surface of the workpiece, plastic flow resulting from high temperature and pressure as well as due to chemical reactions that occur at the point of contact

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

with the workpiece material. The bond bridges binding abrasive grains are also worn away. Crumbling occurring as a result of the cracking or chipping of abrasive particles and binder bridges is a consequence of impact loads, mechanical and thermal fatigue wear [10,19,22,24,27]. These forms of wear do not occur separately, the domination of one of them depends on the grinding conditions, i.e. the characteristics of the grinding wheel (its structure, type and size of abrasive material, hardness, type of binder) and grinding parameters that change the force exerted on grains. Each of these types of wear affects the final machining result. In the case of volume wear resulting from grain and adhesive chipping, there will be shape errors or dimensional deviations in the workpiece [29]. On the other hand, the creation of flat surfaces at the vertex of the grains increases the power and specific energy of grinding, thermal damage and loss of machining accuracy [19]. The assessment of such conditions is carried out most efficiently with optical measuring techniques [5, 18].

The number of active cutting edges on the surface of the grinding wheel has a significant influence on the interplay of the tribological pair of grinding wheel / workpiece. Unfortunately, due to the random distribution of abrasive grains in the grinding wheel and the assumed dressing conditions, the number of active cutting edges is not a constant value for a particular type of grinding wheel [17, 39]. The grinding forces, the surface roughness, the temperature and efficiency of the grinding process depend on the number and uniformity of the distribution of the active cutting edges [30]. Therefore, for a better control over the grinding process, a group of tools with a defined position of abrasive grains is being developed [9, 40].

As the above description shows, the effect of the grinding operation depends on many simultaneously occurring factors affecting the interaction of the workpiece - grinding wheel pair. The correct selection of grinding wheel determines the success of the grinding operation. The properties of the grinding wheel in interaction with grinding parameters and machined material have a decisive influence on the productivity and the achievable quality of the workpiece.

The evaluation of the grinding wheel operational properties can be made based on physical quantities accompanying the grinding process and the quantities describing the machining result [34]. Drawing conclusions based on the value of one quantity is not sufficiently reliable. Hence in literature, among others [16, 25], suggestions of grinding indicators can be found, which use mathematical equations to combine several characteristic quantities, thanks to which one indicator enables multilateral or thematically directed evaluation of the studied process. These indicators take into account the performance, energy and quality aspects of the grinding process and their formulas contain recognized grinding process sizes [11, 12, 14, 21, 33, 38], such as force, power, temperature and grinding performance, grinding wheel wear, vibration amplitude and acoustic emission, as well as grinding results in the form of parameters for assessing the geometric structure of the surface and the heat affected zone.

Examples of indicators based on which one can adjudicate about the course of the grinding process, including the wheel's operational properties, are represented by formulas (1) – (4). The basic grinding ratio (the so-called *G*-ratio) is given by the formula (1). Its high value indicates that the adopted grinding wheel characteristics and grinding conditions ensure high relative efficiency of grinding, as a result of which the share of tool costs in the costs of grinding operations is low.

$$G = \frac{V_m}{V_s} \tag{1}$$

where:

V_m – material removal [mm^3],
V_s – grinding wheel wear [mm^3].

The indicators K_s described in formulas (2) [16] and (3) [25] cover a wide range of grinding process values. In the numerator of both formulas there are values that are high when efficiency of the grinding process increases and wear of the grinding wheel lowers. The denominator takes on smaller values when the grinding effect is a smooth surface and the machining requires low energy consumption. When comparing grinding processes using these indicators, the one with higher value is considered to be better.

$$K_s = \frac{Z'_{max} \cdot G \cdot V'_w}{Rz \cdot W_{sp}} \tag{2}$$

$$K_s = \frac{G}{P_s \cdot Ra} \tag{3}$$

where:

Z'_{max} – maximum specific material removal rate [mm^3/(mm·s)],
V'_w – specific material removal [mm^3/mm],
Rz, Ra – roughness parameters of the machined surface [μm],
W_{sp} – specific grinding energy ($W_{sp} = P_s/Z'_{max}$ [W/mm^3/(mm/s)]),
P_s – grinding power [W].

The indicator Q'_t according to the formula (4) [6], used to assess the cutting ability of the active surface of the grinding wheel, is based on the values obtained from tests performed with a special device outside the machining zone. It determines the velocity of the linear decrement of the sample (tester) approached with a constant force to the rotating grinding wheel.

$$Q'_t = \frac{\Delta l}{\Delta t} \tag{4}$$

where:

Δl – decrement of the tester length [μm],
Δt – time of grinding the tester [s].

A similar principle of measurement was used by the inventors of the G_c index [7] to evaluate the process of grinding samples made of ceramic materials with diamond abrasive belts. During the test, the sample is approached with a constant, controlled force to the abrasive belt on a special test device. The material removal rate Z_w of the grinding carried out under these conditions is a function of the normal grinding force F_n, the speed of the tape v_s, and the properties of the sample material Φ_c and the cutting properties of the tape Φ_d. On the basis of the experimental investigations carried out, the authors [7] found that the material removal rate is proportional to the speed of the tape v_s and the normal grinding force F_n. Thus, they assumed that the G_c index characterizing the interaction of the sample-abrasive tape pair will be defined by the formula (5). Its value, depending on the pair configuration adopted in the tests: the same type of abrasive tape – various sample materials or various abrasive tapes – the same sample material, will depend respectively on the properties of the sample material or the abrasiveness of the abrasive belt.

$$G_c = \frac{Z_w}{v_s F_n} \tag{5}$$

where:

Z_w – material removal rate [mm^3/s],
v_s – peripheral speed of the abrasive belt [m/s],
F_n – normal force [N].

Based on the abovementioned examples, it can be noticed that the number of quantities on the basis of which the value of the index is determined depends on the adopted evaluation criteria and the authors' view on the way of this assessment. However, the inclusion in the indicator of very many values characterizing the grinding process is not favourable due to on their varying variability during the ongoing process. Therefore, it is intended that it contains only quantities that capture the expected grinding effect. And so, if the purpose of grinding is to obtain high quality machining, which is characterized by low roughness of the surface treated and no thermal damage to the surface layer of the workpiece, the indicator should be inversely proportional to the parameters describing these requirements (e.g. *Ra* parameter and depth of heat affected zone z_T), formula (6) [25]:

$$K_{s1} = f\left(\frac{1}{Ra, z_T}\right) \tag{6}$$

In the case if the goal is high efficiency, the indicator should take into account those process parameters, which will define the volume of the layer being removed per unit of time and those that may accompany this process and at the same time disrupt it. The first should be included in the numerator by increasing the value of the index for an efficient process and the second one in the denominator to reduce the value of the indicator, if the negative phenomena of the assessed process are at a high level. An example of such an indicator is represented by the formula (7), in which it is proposed to apply characteristics such as: maximum specific material removal rate Z'_{max}, specific material removal V'_m and vibration amplitude x_d and degree of wheel loading – A_z [25]:

$$K_{s2} = f\left(\frac{Z'_{max}, V'_m}{x_d, A_z}\right) \tag{7}$$

Evaluation of the grinding process in the economic terms should mainly take into account the speed of wear of the grinding wheel and energy consumption related to the machining process. The indicator, which accomplishes the above goal (8), contains the grinding ratio *G* in the numerator, which refers to the material removal for volume wear of the grinding wheel, and in the denominator of the dimensions that adversely affect the grinding result, but also indicate energy consumption on the machining process, i.e. temperature grinding T_{max}, components of grinding forces F_n and F_t, amplitude of vibrations x_d and degree of wheel loading A_z [25]:

$$K_{s3} = \frac{G}{f\left(T_{max}, F_n, F_t, x_d, A_z\right)} \tag{8}$$

Despite the use of several measurable quantities in the indicators, no results have yet been obtained that would confirm that it is possible to determine, course and results of the grinding process using one equation. Therefore, in order to obtain transparent information about the examined process, several indicators should be applied, properly selected to the adopted criteria. It also contributes to the lower number of factors that need to be analysed, and thus to easier implementation of the research. Determination of process values, which are inputs to indicators, requires testing with the use of control and measurement equipment which in most cases is only available in research laboratories, not in production conditions. This is a barrier to the precise and efficient design of industrial grinding operations in terms of optimization of the grinding wheel's operational properties.

The above conclusion induced authors to undertake two types of actions. The first was to develop a set of indicators that would characterize the course and results of the grinding process to the widest extent possible, which might also be useful for comparative assessment of grinding wheel's operational properties. The second action was to design the device, useful both in laboratory and industrial conditions, to perform short-term grinding tests, on the basis of which it is possible to determine the values of parameters used in the indicators.

2. Description of the method for assessing the wheel's operational properties

In order to evaluate the operational properties in terms of efficiency, quality and economics, a group of three indicators was suggested, that describe the grinding process using the tested grinding wheels.

It has been assumed that the performance indicator K_w is determined by equation (9):

$$K_w = \frac{Z'}{F_t'} \quad \left[\frac{mm^3}{N \cdot s}\right] \tag{9}$$

where:

Z' – specific material removal rate [mm^2/s],
F_t' – tangential grinding force per width unit [N/mm].

In the presented equation, the specific material removal rate is related to the tangential component of the grinding force that occurs during machining. This relation determines the amount of material grounded in the time unit per tangential component of the grinding force. A higher value of the indicator means better performance.

The K_j index, which describes the quality of the surface layer of the workpiece, was determined by the below formula (10):

$$K_j = \frac{F_n}{Ra \cdot F_t \cdot \Delta T_p} \quad \left[\frac{1}{K \cdot \mu m}\right] \tag{10}$$

where:

F_t – tangential grinding force [N],
F_n – normal grinding force [N],
ΔT_p– increase of sample's temperature during the test [K],
Ra – surface roughness of the specimen [μm].

The quotient of the grinding force components (F_n/F_t) used in the index can be treated as the inverse of the tribological contact coefficient combining the removal of the allowance and the friction occurring during the grinding process of the object with the grinding wheel. Its increasing value indicates less smoothing effect of the grinding wheel. The second factor that is important for the quality of the object's surface layer is the heat flow. This was taken into account by introducing an increase in surface temperature ΔT_p into the index. What is more, the geometric surface structure of the object was taken into account by introducing the roughness parameter *Ra*.

Index K_e (11), which task is economic description of the machining process, is equal to a well-known G-ratio grinding index expressing the relative grinding efficiency. Its value is the quotient of the material removal V_m and the grinding wheel wear V_s:

$$K_e = G \tag{11}$$

Evaluation of the grinding process based on three indicators can give a more complete view on the relations between the grinding wheel and the workpiece, enabling the analysis of grinding wheel abrasive properties in terms of efficiency, quality of the machined surface and the tool's share in the operation costs. The described indicators have been developed taking into account the possibility of their practical application, paying attention to the scope of the adopted assessment criteria and the ability to easy interpretation of the obtained results.

3. Grinding wheel testing device

One of the main design assumptions for the construction of the device was the possibility of its application in industrial conditions and the possibility of its application on various grinders. The designed device, the diagram of which is shown in Figure 1, is adapted to perform tests on surface grinders. The operating principle of device is to transmit cylindrical rotary motion to sample 1 and induce its infeed at constant speed to the active surface of the grinding wheel 9. After contact with a rotating grinding wheel, the plunge grinding of the sample takes place. The applied rotational and radial movement of the sample ensures control over the machining efficiency.

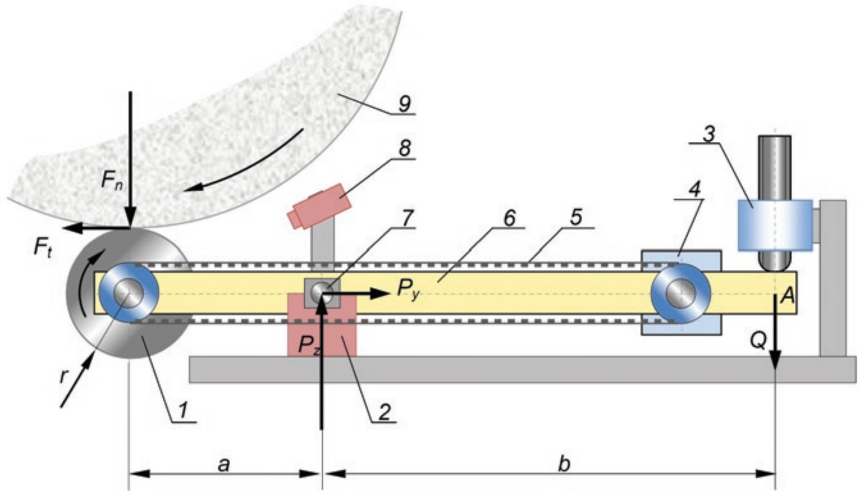


Fig. 1. Diagram of the device testing grinding wheel's operational properties 1 - sample, 2 - dynamometer Kistler type 9317B, 3 - actuator, 4 - sample drive motor, 5 - toothed belt, 6 - lever, 7 - lever axle, 8 - pyrometer, 9 - grinding wheel

The following can be singled out in the design of the device (fig. 1):
- a belt-driven sample assembly 5 placed on the rolling bearing of double-arm lever 6,
- lever axle 7 mounted directly on the piezoelectric, three-force dynamometer 2,
- pyrometer 8 fixed on the lever perpendicular to the surface of the sample 1,
- actuator 3 with lever thrust screw.

Placing the dynamometer under the lever's axle enabled the evaluation of the grinding forces. The distribution of forces occurring during grinding (Fig. 1) and their projection on the direction of the lever, indicate that the horizontal force component P_y registered by the dynamometer gets the value of the tangential grinding force F_t. However, the normal component of the grinding force F_n can be determined on the basis of equation (12), determined from the conditions of the moment equilibrium referring to the point of application of external force Q (point A in Figure 1):

$$F_n = \frac{P_z \cdot b - F_t \cdot r}{a + b} \tag{12}$$

The described device has been patented [37] and the stand construction, automatic control system and application of the grinding measurement cycle are presented in the article [36]

4. Evaluation of indicators of grinding wheels operational properties

In order to assess whether the designed formulas of indicators and the method of obtaining the data necessary to determine them allow obtaining information useful for objective evaluation of wheel performance, tests were performed in various grinding conditions, i.e.:
- grinding with various adjustable parameters on the test device,
- grinding of flat and cylindrical surfaces,
- dry and wet grinding.

The tests were performed on SPG 30x80 (PONAR-Głowno) plane grinders. A sample of cold work tool steel (145Cr6 - 65HRC) was being machined with a vitrified alumina grinding wheel, 1-350x20x127-99A46J7VE01-35. Each test was repeated 3 times due to the grinding parameters set out in the test plan. Maintaining the same cutting ability of grinding wheel during testing was ensured by performing a dressing of wheel before each test. For dressing, a single-grain diamond dresser with a specified active diamond width b_d was used. The dressing feed rate f_d was selected in order to maintain the same value of the overlap ratio k_d = 1,5 (equation (13)). The dressing infeed a_d was set at the lever of 0,01mm/pass:

$$k_d = \frac{b_d}{f_d} \tag{13}$$

During the tests, following measurements were made:
- the magnitude of forces loading the dynamometer in the Z and X directions. On the ground of these measurements the components of the grinding forces were determined. The dynamometer load value, resulting from the machining process, was calculated as the difference of the dynamometer measurements, which were recorded in the final part of the test during stable grinding and when there was no contact between the grinding wheel and the object (Fig. 2a),
- increase of the temperature of the sample's surface layer (Fig. 2b),
- wear of the wheel after each test (Fig. 2c),
- sample's roughness (the average of 5 measurements was used for calculations).

Wear of the grinding wheel was assessed basing on a profilogram obtained from a groove shape representation, which was created on the circuit of the grinding wheel as a result of wear processes occurring during grinding of a sample with a smaller width than the grinding wheel. The measurement was carried out with a Hommel-Wave measuring device. An exemplary result of profiling two grinding wheels is shown in Figure 2c.

Studies on the indicators (9) - (11) within the vriability range of grinding parameters were carried out on the test device described in p.3 - fig.3.

The tests were performed according to the planned experiment. The sample feed a_e and the peripheral speed of the v_p sample were taken as input values. The research was carried out according to the 32-trivalent, two-parameter, full plan, which was generated in the DOE module of the Statistica program. A random option of its layout was chosen to avoid systematic errors in the results of the conducted ex-

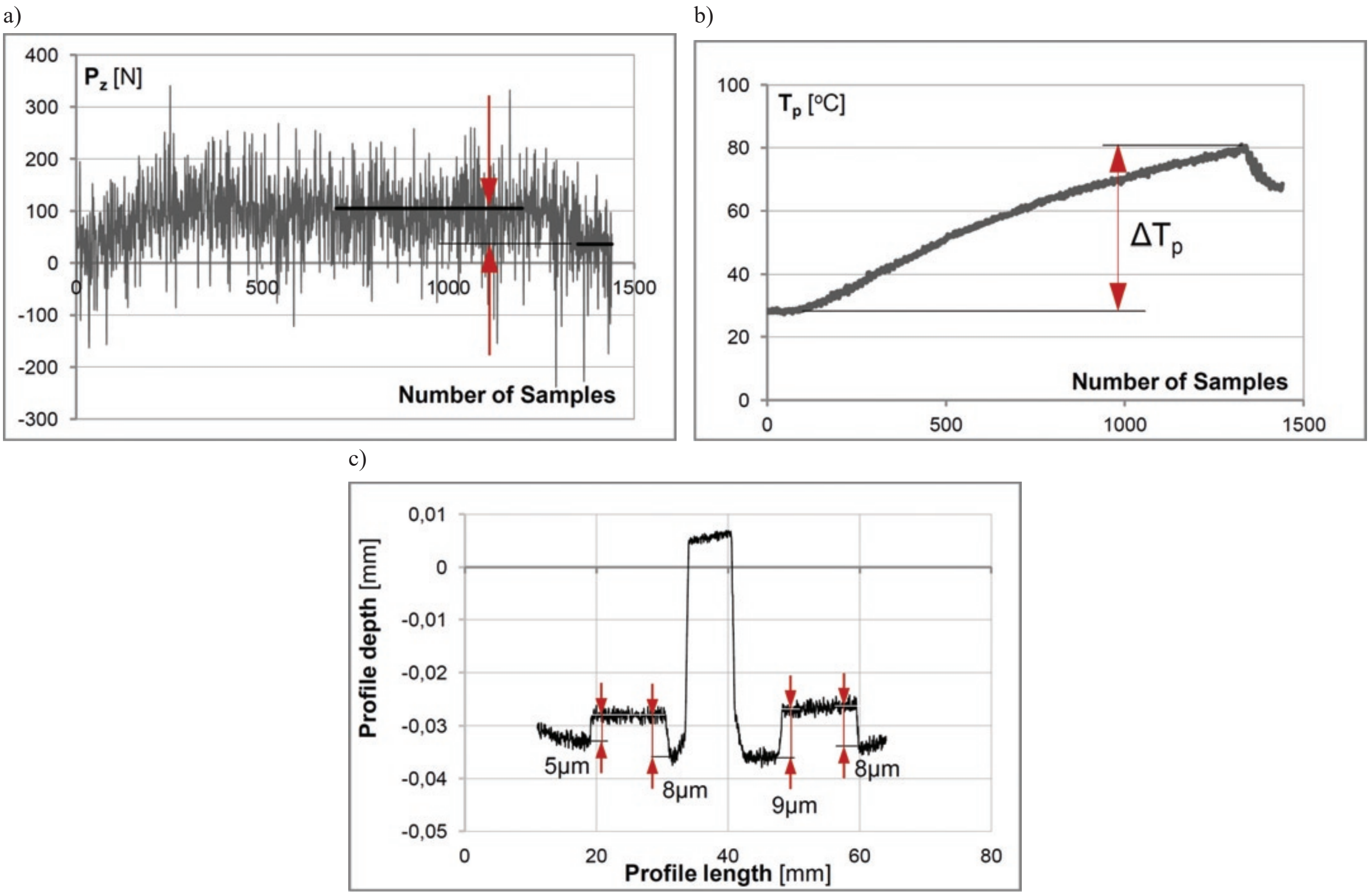


Fig. 2. Measurement results: a) load force of the dynamometer P_z, b) surface layer temperature of the sample, c) wheel wear

periments. The tests consisted of performing 1-minute trials of plunge grinding of cylindrical samples. The values of grinding parameters used in the tests are summarized in table 1. During the grinding process no cooling liquid was used.

Figure 4 presents graphs and equations of functions approximating the values of K_w, K_j, K_e indices which parameters are infeed a_e

Table 1. Values of the test input parameters

Parameter	Value
a_e [mm/rev]	0,005; 0,0125; 0,02
v_p [m/s]	0,15; 0,225; 0,3
v_s [m/s]	26

and velocity v_p. Approximation of test results was made using power functions, often used in modeling abrasive machining phenomena. These functions accurately reflect the main tendencies in the relations between variable inputs and process results, and are not susceptible to accidental deviations of measured values. The CurveExpert Professional v.1.5 program was used to generate the form of functions and their graphic presentations.

The presented graphs show that the indices calculated according to the proposed formulas react properly to the change of the adjustable parameters of the grinding process.

The value of the specific tangential grinding force included in the grinding efficiency indicator K_w (9) does not disturb its expected increase after the application of more intensive grinding conditions. This means that for the specific settings of the grinding process, greater values of the indicator will be shown in cases of those pairs of

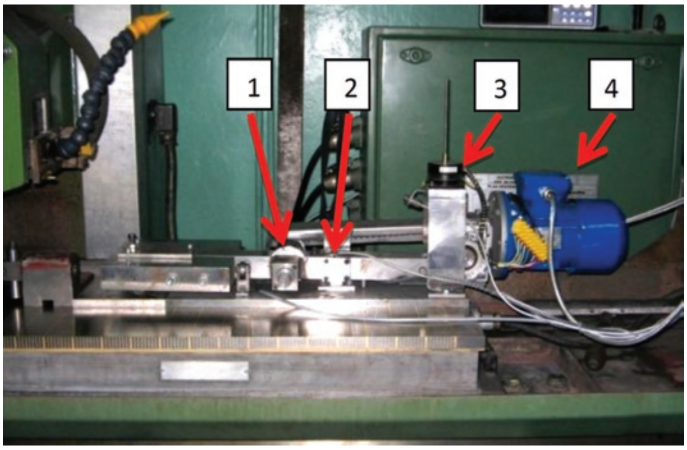


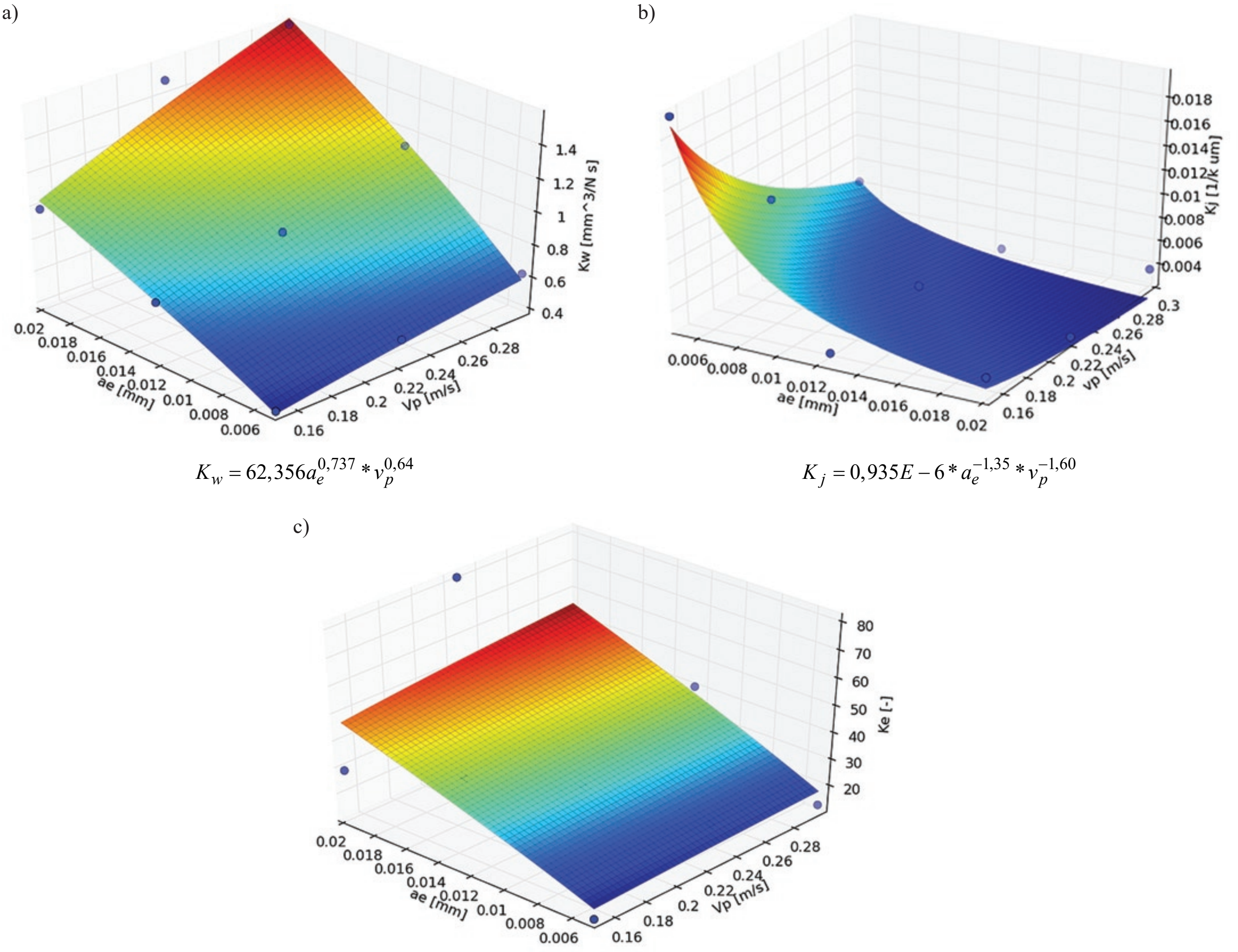
Fig. 3. Device for testing the wheel's operational properties set on the grinder's table SPG 30x80 1 - specimen, 2 - pyrometer, 3 - actuator, 4 - three-phase motor

grinding wheel and material, when the machining allowance is more easily removed.

Similarly, the grinding quality indicator K_j (10) takes highest values for the most benign machining conditions, that provide the fewest stresses and roughness of the surface layer.

The K_e (*G*-ratio) indicator was confirmed to increase its value in the response to more intensive machining with high cutting parameters.

The basis for determining how the indicators behave when using different types of grinding, i.e. when machining cylindrical surfaces and planes, were the results of tests carried out on the specially designed device and directly on the surface grinder, which was equipped



$K_w = 62{,}356 a_e^{0{,}737} * v_p^{0{,}64}$

$K_j = 0{,}935E-6 * a_e^{-1{,}35} * v_p^{-1{,}60}$

$K_e = 1807{,}97 a_e^{0{,}821} * v_p^{0{,}232}$

Fig. 4. Relation between the indicators of the wheel operational properties and the grinding parameters (specimen material 145Cr6, grinding wheel 99A46J7VE01)

with a Kistler dynamometer (model 9275) and the same type of pyrometer as in a test device. In both cases, the same values of the adjustable parameters of the grinding process were used (Table 1). In the surface grinding tests cuboidal samples were used with the treated surface's dimensions of 10x250mm. The grinding tests were carried out without the use of a cooling liquid.

Figure 5 presents charts showing the obtained results. The adjustable parameters of the grinding process, i.e. a_e, v_p, v_s, were replaced with one h_{eq} parameter (equivalent chip thickness (14)), which characterizes its intensity:

$$h_{eq} = \frac{a_e \cdot v_p}{v_s} \quad [mm] \tag{14}$$

where:

a_e – feed rate of the sample [mm],
v_p – sample's peripheral speed [m/s],
v_s – grinding wheel peripheral speed [m/s].

The results of the tests show a significant difference in the values of the indices describing the operational properties of the wheel used for surface and cylindrical plunge grinding, (Figure 5). This variety results from different grinding wheel operating conditions in both cases. The heat generated during cylindrical grinding increases the temperature of the sample much faster, as the grinding wheel remains in constant contact with the machined surface. In the case of surface grinding, when the grinding wheel is outside the machining zone, the heat accumulated by the sample and the grinding wheel is dispersed. However, the bigger contact area of the grinding wheel with the workpiece and the intermittent cutting process during surface grinding, cause more intensive wear of the grinding wheel. The causes are the higher loads that the abrasive grains experience in the chip-forming zone and when hitting of the grinding wheel on the edge of the specimen after reverse.

The above processes have their reflection in the values of the determined indicators. The increase in the temperature of the cylindrical sample in increasingly intensive grinding conditions is not compensated by the lower value of the grinding force, thus the grinding quality index K_j is getting lower and lower and approaching the value achieved during surface grinding. Surprisingly, there is only a slight increase in the indicator value despite setting the adjustable paratemers at their highest values. According to formula (10), in this case it must have had a disproportionately greater increase in the normal grinding force in relation to the other values used in the indicator. It can be assumed that this was a consequence of the inability to machine such chip thickness, which was further the consequence of increased friction and increased wear of the grinding wheel. The economic

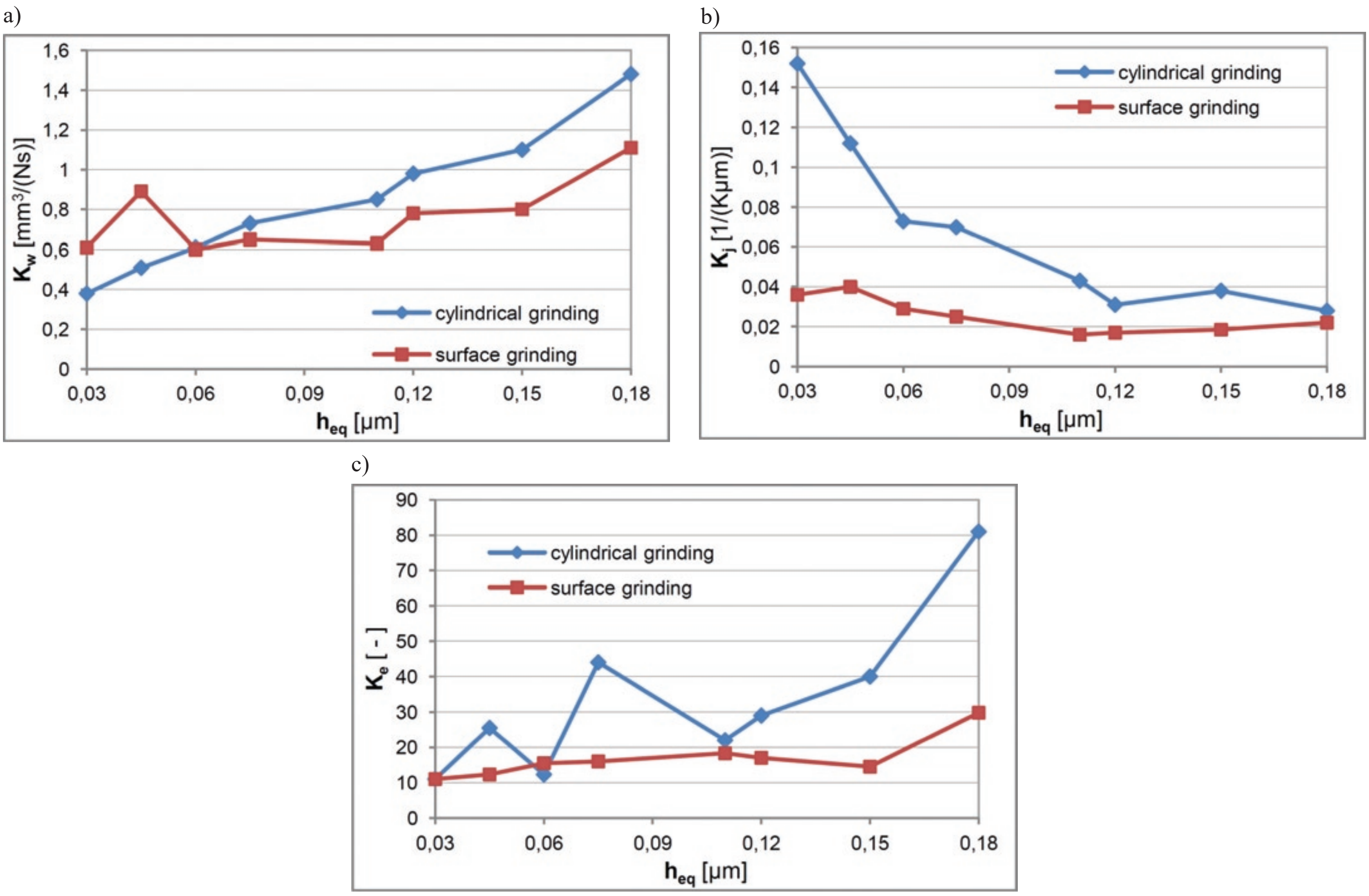


Fig. 5. Relation between the values of grinding wheel's operational indicators K_w, K_j and K_e and grinding parameters during cylindrical and surface grinding.

efficiency index K_e confirms this presumption, assuming a lower value for these grinding conditions.

As can be seen, graphs showing the indicators of wheel properties evaluation obtained in cylindrical grinding research are characterised by the grater dynamics of changes. This means that the indicators are characterized by greater sensitivity to changes in the processing conditions and that the proper method of grinding has been used in the developed method.

Tests evaluating indicators in the field of dry grinding and using cooling liquid (CL) were carried out in plane grinding conditions. Due to the inability to perform measurements with the pyrometer during grinding using CL, no qualitative index K_j was determined, which formula has an ΔT_p factor that determines the temperature increase of sample's surface layer. For further analysis, the K_w and K_e indexes were determined and the relation between their values and the grinding adjustable parameters is presented in Figure 6.

Both indicators were of higher, more favourable values in grinding operations carried out without the use of CL due to the grinding wheel operational properties. The use of a cooling liquid reduces the friction between the wheel and the workpiece, making it difficult to the chip formation. Lowering of the friction forces between the cutting edge and the workpiece reduces the tangential stress τ, that arises in the chip-forming zone. Thus, in order to form a chip, it is required to increase the depth of cutting edge indentation in the material (15) [32]:

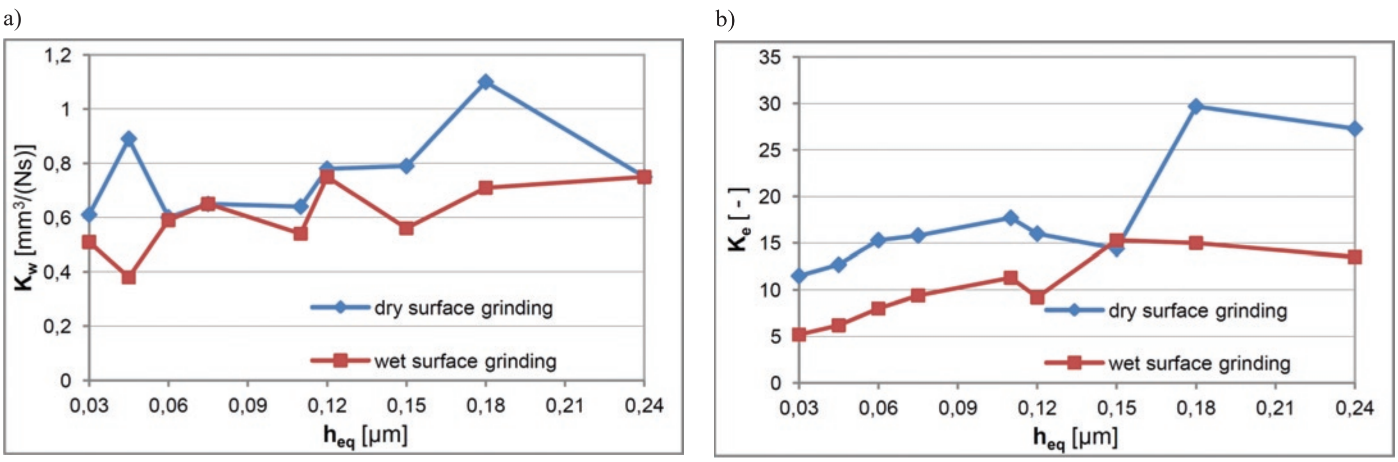


Fig. 6. Relation between the values of grinding wheels' operational indicators K_w and K_e and grinding parameters when dry and wet surface grinding

$$h_\mu = \rho_k \left\{ 1 - (cos\eta)\sqrt{0,5\left(1+\frac{\tau}{k}\right)} + (sin\eta)\sqrt{0,5\left(1-\frac{\tau}{k}\right)} \right\} \tag{15}$$

where:

- h_μ – threshold depth of indentation [μm],
- ρ_k – the radius of the sphere described on the abrasive grain [μm],
- η – cutting angle [°],
- k – material yield stress during shear [MPa],
- τ – shear stress [MPa].

Due to the spatial distribution of the cutting edges on the active surface of the grinding wheel, the number of cutting edges that will be able to cut into the required depth decreases. Therefore, efficiency indices K_w and the tool utilization efficiency K_e when dry grinding takes on more favorable, higher values. The higher K_e value when dry grinding also shows that limiting the wear of abrasive by using CL does not compensate for the higher grinding efficiency that is achieved during dry grinding.

The studies discussed above have shown that irrespective of the grinding conditions adopted, the proposed set of indicators and the method of their determination correctly evaluate the course of the grinding process in three aspects relevant to the use and are useful for assessing the wheel performance. Therefore, it should be recognized that the presented indicator formulas are an effective tool for evaluating both the course and results of grinding for a specific grinding / material pair in the tested range of adjustment values in the grinding process, i.e. infeed per revolution of the sample against the grinding wheel a_e and peripheral speed of the sample v_p.

5. Comparative research on the operational properties of grinding wheels during grinding of steel

The applicability of the discussed method was verified in terms of the choice of grinding wheels for the machining task. The objective of the studies was to check the response of indicators K_w, K_j, K_e to changes in the type of grinding wheel used for machining a particular type of material. The course and results of the grinding process carried out with two grinding wheels selected according to the recommendations described in the manufacturer's catalog were evaluated [13]. These are grinding wheels made of friable and monocrystalline alumina with varying grit size, hardness and structure. The tests were carried out for a group of four materials that are normally applied on parts undergoing hardening and grinding.

Table 2 includes the sample materials and characteristics of the recommended grinding wheels. In addition to the grinding wheel characteristics, there is the designation in brackets, which is used later in the article. Research was conducted using the same experimental and grinding conditions as in the evaluation of indicators.

5.1. Analysis of comparative study results

Figures 7 and 8 show graphs of index values K_w, K_j, K_e, which were calculated based on the measured values obtained in the tests. In the charts descriptions grinding wheel characteristics were simplified, giving only their grain, hardness and structure. The title of the graph contains a type of grinding material.

The graphs presented in Figure 7, showing the results of multicriteria evaluation of two types of grinding wheels used for grinding of steel 145Cr6, indicate that in most of the cutting parameters ranges, the quality of the surface layer of sample K_j is slightly higher after grinding with a grinding wheel 46J7. Additionally, due to being harder the grinding wheel 46J7 showed a lower relative wear of K_e. However, the softer grinding wheel 46I8 more easily and efficiently removed the K_w material, which was probably due to its more intense self-sharpening and a more open structure. The final choice of the grinding wheel and grinding conditions of the 145Cr6 steel can be made by the user, taking into account preferable operational characteristics: the quality provided by the grinding wheel 46J or the efficiency provided by the grinding wheel 46I8.

Grinding tests of the 100Cr6 steel with 46I8 and 46J7 grinding wheels unequivocally indicate the advantage of the 46I8 grinding wheel with the more open structure. Most of the machining parameters of 46I8 are characterized by higher values of grinding indices compared to the grinding wheels 46J7. Only grinding using the lowest infeed and rotational speed of the sample gives an advantage to the grinding wheel 46J7 in terms of the efficiency K_w and quality of the machining K_j.

The Figure 8 exemplifies the charts referring to the grinding results of C45 and 42CrMo4 steel using 46L and 60K grinding wheels. Analysis of these graphs indicates that 60K grinding wheel, within the small intervals of parameters values, gives better results compared to C45 steel. Considering more intensive grinding, the 46L grinding wheel gains advantage giving similar surface layer quality K_j and relatively less wear K_e. The analysis of grinding process of the 42CrMo4 low hardness steel indicated that the 60K grinding wheel should be recommended and considered as more useful within the whole range of tested grinding parameters. Operational properties indicators of this grinding wheel take lower values compared to the competitive tool only in individual cases.

Summing up, in many cases the indicators of grinding wheels operational properties have similar values, which proves similar usefulness of the tested tools for machining tasks. Differences in indicators value within a specific set of grinding parameters give the possibility of better choice of machining conditions for a particular material.

Table 2. Characteristics of samples and grinding wheels used in the tests

Material	Sample's hardness	Grinding wheel / marking
Alloy cold-work tool steel 100Cr6	55 HRC	1-350x20x127-M463I8VE01NPB5-35 (46I8)
		1-350x20x127-99A46J7VE01-35 (46J7)
Alloy tool steel 145Cr6	65 HRC	1-350x20x127-M463I8VE01NPB5-35 (46I8)
		1-350x20x127-99A46J7VE01-35 (46J7)
Non-alloy quality steel C45	40 HRC	1-350x20x127-99A46L7VE01-35 (46L)
		1-350x20x127-99A60K7VE01-35 (60K)
Alloy special steel 42CrMo4	20 HRC	1-350x20x127-99A46L7VE01-35 (46L)
		1-350x20x127-99A60K7VE01-35 (60K)

6. Summary

The current challenges standing against production processes, resulting from the growing requirements for product quality, minimizing costs and ensuring the feasibility of implementation, necessitate constant improvement and modernization. Technological operations included in production processes should be transformed towards efficiency and productivity improvement. A way to achieve these goals is multicriteria optimization, which in terms of the grinding process assessment may consist of the grinding wheel operational properties, characterizing the interactions between all process factors. The method presented in the article shows that using short test one can get knowledge about dependencies occurring

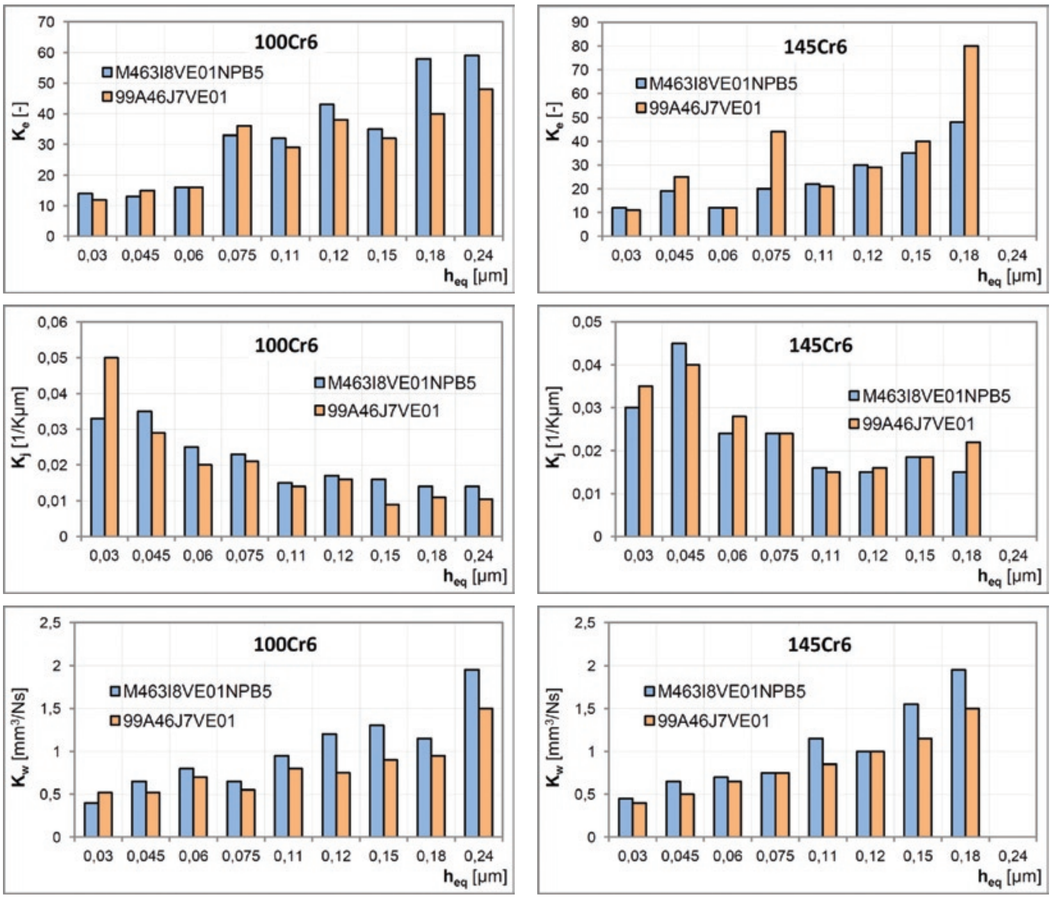


Fig. 7. Operational properties indices for evaluation of M463I8VE01NPB5 and 99A46J7VE01 grinding wheels used for grinding of 100Cr6 and 145Cr6 steels

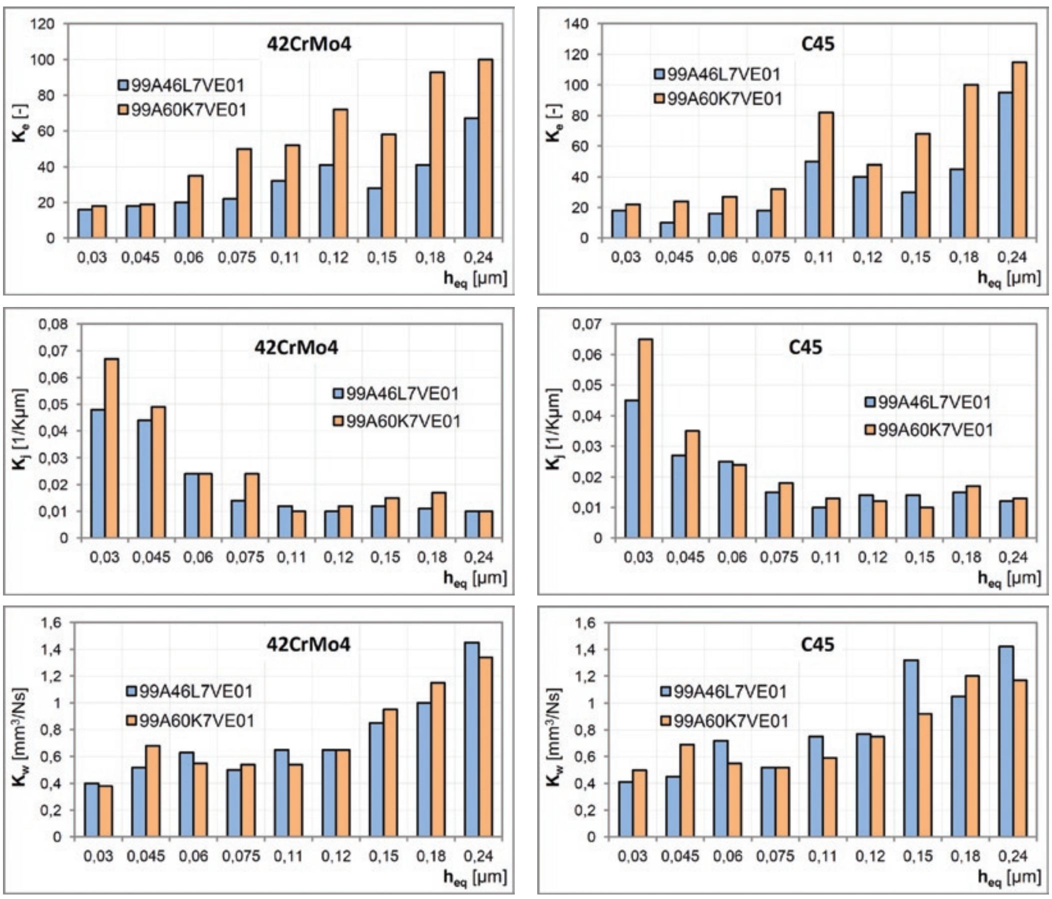


Fig. 8. Operational properties indices for evaluation of 99A46L7VE01 and 99A60K7VE01 grinding wheels used for grinding of 42CrMo4 and C45 steels

in the grinding process, which gives the opportunity to improve at least one of the process characteristics regarding efficiency, costs and quality. The research carried out showed that:

- the range of indicators and formulas adopted correctly characterize the operational properties of grinding wheels, relating them to the process efficiency and cost as well as the quality of the surface layer of the treated surface,
- the grinding variant used in the test device (cylindrical grinding) ensures high sensitivity of the indicators to changes in the grinding conditions, which enables a more unambiguous assessment of the tested grinding wheel's operational properties,
- results of tests on grinding of cylindrical surfaces in dry grinding conditions allow to conclude on the grinding wheel properties in other grinding conditions such as different grinding method or application of cooling liquid.
- due to the varied reaction of the grinding wheel's operational properties to the change of the adjustable parameters of the grinding process, the grinding wheel should be selected based on the values of all presented indicators so that the final decision takes into account the important for the user criteria for the optimized grinding operation.

References

1. Adibi H, Rezaei Ahmed S M, Sarhan A D. Analytical modeling of grinding wheel loading phenomena. International Journal of Advanced Manufacturing Technology 2013; 68: 473-485, https://doi.org/10.1007/s00170-013-4745-z.
2. Axinte D, Butler-Smith P, Akgun C, Kolluru K. On the influence of single grit microgeometry on grinding behavior of ductile and brittle materials. International Journal of Machine Tools and Manufacture 2013; 74: 12-18, https://doi.org/10.1016/j.ijmachtools.2013.06.002.
3. Brinksmeier E, Heinzel C,Wittmann M. Friction, Cooling and Lubrication in Grinding CIRP Annals - Manufacturing Technology 1999; 48: 581-598.
4. Burakowski T, Wierzchoń T. Surface Engineering of Metals: Principles, Equipment, Technologies. CRC Press, Boca Raton, 1999.
5. Dębkowski R. Analiza komputerowa obrazu mikroskopowego w zastosowaniu do oceny zużycia czynnej powierzchni ściernicy. Materiały XXIII Naukowej Szkoły Obróbki Ściernej, Rzeszów, 2000.
6. Gołąbczak A, Koziarski T. Assessment method of cutting ability of CBN grinding wheels. International Journal of Machine Tools and Manufacture 2005; 45: 1256-1260, https://doi.org/10.1016/j.ijmachtools.2005.01.008.
7. Guo C, Chand R H. Grindability and Mechanical Property of Ceramics. Proceedings of the 20th Annual Conference on Composites, Advanced Ceramics, Materials, and Structures-A: Ceramic Engineering and Science Proceedings, 2008: 214-219.
8. Hahn R S. On the mechanics of the grinding process under plunge cut conditions. Journal of Engineering for Industry 1966; 1: 72-79, https://doi.org/10.1115/1.3670895.
9. Herzenstiel P, Aurich J C. CBN-grinding wheel with a defined grain pattern -extensive numerical and experimental studies, Machining Science and Technology 2010; 14: 301-322, https://doi.org/10.1080/10910344.2010.511574.
10. Jackson M J, Mills B. Microscale wear of vitrified abrasive materials. Journal of Materials Science 2004; 39: 2131-2143, https://doi.org/10.1023/B:JMSC.0000017776.67999.86.
11. Kacalak W, Lipiński D, Rypina Ł, Szafraniec F, Tandecka K, Bałasz B. Performance evaluation of the grinding wheel with aggregates of grains in grinding of Ti-6Al-4V titanium alloy. International Journal of Advanced Manufacturing Technology 2018; 94: 301-314, https://doi.org/10.1007/s00170-017-0905-x.
12. Karpuschewski B, Wehmeier M, Inasaki I. Grinding monitoring system based on power and acoustic emission sensors. Annals of the CIRP 2000; 49: 235-240, https://doi.org/10.1016/S0007-8506(07)62936-9.
13. Katalog materiałów ściernych. Koło: Andre Abrasives, 2017.
14. Kato T, Fuji H. Temperature measurement of workpieces in conventional surface grinding. Journal of Manufacturing Science and Engineering 2000; 122: 297-303, https://doi.org/10.1115/1.538918.
15. Klocke F, Brinksmeier E, Weinert K. Capability Profile of Hard Cutting and Grinding Processes. CIRP Annals 2005; 54: 22-45, https://doi.org/10.1016/S0007-8506(07)60018-3.
16. König W, Messer J. Influence of the Composition and Structure of Steels on Grinding Process. Annals of the CIRP 1981; 30: 547-552, https://doi.org/10.1016/S0007-8506(07)60165-6.
17. Koziarski A, Golabczak A. The Assessment of the Grinding Wheel Cutting Surface Condition after Dressing with Single Point Diamond Dresser. International Journal of Machine Tools Design and Research 1985; 25: 313-325, https://doi.org/10.1016/0020-7357(85)90032-0.
18. Lipiński D, Kacalak W, Tomkowski R. Methodology of Evaluation of Abrasive Tool Wear with the Use of Laser Scanning Microscopy. Scanning 2014; 36: 53-63, https://doi.org/10.1002/sca.21088.
19. Malkin S, Cook N H. The wear of grinding wheels. Part 1: attritious wear. Transactions of ASME, Journal of Engineering for Industry 1971; 93: 1120-1128, https://doi.org/10.1115/1.3428051.
20. Malkin S, Guo C. Thermal analysis in grinding. CIRP Annals - Manufacturing Technology 2007; 56, 760-782.
21. Malkin S, Guo C. Thermal analysis of grinding. Annals of the CIRP 2007; 56: 760 - 782, https://doi.org/10.1016/j.cirp.2007.10.005.
22. Malkin S. The wear of grinding wheels. Part 2: fracture wear. Transactions of ASME, Journal of Engineering for Industry 1971; 93: 1129-1133, https://doi.org/10.1115/1.3428052.
23. Marinescu I D, Hitchiner M, Uhlmann E, Rowe W B. Inasaki I. Handbook of Machining with Grinding Wheels. CRC Press, Boca Raton, 2007.
24. Marinescu I D, Rowe W B, Dimitrov B, Inasaki I. Tribology of abrasive machining processes. William Andrew, Inc., Norwich, 2004.
25. Maslov E N.: Teorija šlifovanija materialov, Moskva: Mašinostroenie, 1974.
26. Mofdi M, Linghchi M, Zhang T. Applied mechanics in grinding. Part 7: residual stresses induced by the full coupling of mechanical deformation, thermal deformation and phase transformation. International Journal of Machine Tools and Manufacture 1999; 39: 1285-1298, https://doi.org/10.1016/S0890-6955(98)00094-7.
27. Nadolny K. Wear phenomena of grinding wheels with sol-gel alumina abrasive grains and glass-ceramic vitrified bond during internal cylindrical traverse grinding of 100Cr6 steel. International Journal of Advanced Manufacturing Technology 2015; 77: 83-98, https://doi.org/10.1007/s00170-014-6432-0.

28. Rasim M, Mattfeld P, Klocke F. Analysis of the grain shape influence on the chip formation in grinding. Journal of Materials Processing Technology 2015; 226: 60-68, https://doi.org/10.1016/j.jmatprotec.2015.06.041.
29. Rowe W B. Principles of modern grinding technology. Elsevier, 2009.
30. Setti D, Ghosh S, Rao P V. A method for prediction of active grits count in surface grinding. Wear 2017; 382-383: 71-77, https://doi.org/10.1016/j.wear.2017.04.012.
31. Sieniawski J, Nadolny K. The effect upon grinding fluid demand and workpiece quality when an innovative zonal centrifugal provision method is implemented in the surface grinding of steel CrV12. Journal of Cleaner Production 2016; 113: 960-972, https://doi.org/10.1016/j.jclepro.2015.11.017.
32. Steffens K. Beschreibung eines Gleitlinienfelds für die Deutung der Spanbildung beim Schleifen. Industrie-Anzeiger 1979; 19.
33. Stephenson D J. Three Dimensional Finite Element Simulation of Transient Heat Transfer in High Efficiency Deep Grinding. Annals of the CIRP 2004; 53: 259-262, https://doi.org/10.1016/S0007-8506(07)60693-3.
34. Tönshoff K H, Friemuth T, Becker J C. Process monitoring in grinding. Annals of the CIRP, 2002; 51: 551-571, https://doi.org/10.1016/S0007-8506(07)61700-4.
35. Uhlmann E, Lypovka P, Hochschild L, Schröer N. Influence of rail grinding process parameters on rail surface roughness and surface layer hardness. Wear 2016; 366-367:287-293, https://doi.org/10.1016/j.wear.2016.03.023.
36. Urbaniak M, Leonarcik R, Szajder M. Stanowisko do oceny szlifowalności materiałów. Naukowa Szkoła Obróbki Ściernej, Gdańsk, 2011.
37. Urbaniak M, Skowron M, Leonarcik R. Patent nr 213086 na wynalazek pt.: Urządzenie pomiarowe, 2013.
38. Urbaniak M. System oceny użytkowych właściwości ściernic. Łódź: Zeszyty Naukowe P.Ł., 2002.
39. Wegener K, Hoffmeister W, Karpuschewski B, Kuster F, Hahmann W C, Rabiey M. Conditioning and monitoring of grinding wheels. CIRP Annals 2011; 60: 757-777, https://doi.org/10.1016/j.cirp.2011.05.003.
40. Yu H, Lu Y, Wang J. Study on wear of the grinding wheel with an abrasive phyllotactic pattern. Wear 2016; 358-359: 89-96, https://doi.org/10.1016/j.wear.2016.04.007.

Rafał LEONARCIK
Sonoco Poland - Packaging Services Sp. z o.o.
Nowy Jozefów 70, 94-406 Łódź, Poland

Mirosław URBANIAK
Department of Technology
Jacob of Paradies University
Fryderyk Chopin 52 Street, 66-400 Gorzów Wielkopolski, Poland

Ryszard DĘBKOWSKI
Institute of Machine Tools and Production Engineering
Lodz University of Technology
Stefanowskiego 1/15 Street, 90-924 Łódź, Poland

E-mails: rafal.leonarcik@sonoco.com,
murbaniak@ajp.edu.pl, ryszard.debkowski@p.lodz.pl

Article citation info:
ŚWIDER J, ZBILSKI A. Power losses and their properties for low range of a robot electric motor working conditions as the part of energy effectiveness research. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 542–548, http://dx.doi.org/10.17531/ein.2018.4.5.

Jerzy ŚWIDER
Adrian ZBILSKI

POWER LOSSES AND THEIR PROPERTIES FOR LOW RANGE OF A ROBOT ELECTRIC MOTOR WORKING CONDITIONS AS THE PART OF ENERGY EFFECTIVENESS RESEARCH

STRATY MOCY ORAZ ICH WŁASNOŚCI W NISKIM ZAKRESIE WARUNKÓW PRACY SILNIKA ELEKTRYCZNEGO ROBOTA W KONTEKŚCIE BADAŃ EFEKTYWNOŚCI ENERGETYCZNEJ*

Power losses are one of many factors affecting the energy effectiveness of production processes, however despite this, commonly investigated ranges of power losses do not explain how they change in the stages being different from a typical driving mode. This investigation focuses on low working conditions of a robot electric motor and the properties of power losses changes while going from a driving mode into a stand-still mode of electric motor work. Apart from determined values of power maps components, this work shows how to manage with technical limitations in performing measurements of industrial robot electrical states at the industrial conditions, like high disturbances, noise and limited range of robot axis angle position.

***Keywords**: energy effectiveness, industrial robot, electric motor, power maps, power losses, disturbances, noise, harmonics.*

Straty mocy są jednym z wielu czynników wpływających na efektywność energetyczną procesów produkcyjnych, jednak pomimo tego, najczęściej badane zakresy strat mocy nie określają sposobu ich zmian w trybach pracy odmiennych od typowej pracy napędowej. Opisane badania zostały skoncentrowane na niskim zakresie warunków pracy silnika robota przemysłowego oraz na własnościach zmian postaci strat mocy podczas przechodzenia ze stanu pracy napędowej do pracy statycznej. Oprócz wyznaczonych wartości komponentów map mocy, w pracy przedstawiono techniczne rozwiązania umożliwiające wykonywanie pomiarów stanów elektrycznych robota w warunkach przemysłowych, którymi były zniekształcenia, zakłócenia oraz ograniczony zakres pozycji kątowych badanego przegubu robota.

***Słowa kluczowe**: efektywność energetyczna, robot przemysłowy, silnik elektryczny, mapy mocy, straty mocy, zniekształcenia, zakłócenia, harmoniczne.*

1. Introduction

Energy effectiveness of a production processes is a part of a machines building and exploitation discipline, whose significance was already noticed and is constantly being developed [3, 4, 9, 10, 11]. Authors of this paper focused an attention in their work on all factors, having an influence on the final electrical energy, consumed by the manipulation and transportation machines and particularly consumed by industrial robots.

This investigation is an extension of the work with power losses in the FANUC AM100iB robot electric motors, presented in [20], and introduces some partial improvement into a wider energy effectiveness research.

The object of this investigation was a Fanuc AM100iB robot (Fig. 1), and more specifically its first electrical motor – working at steady-state at low range conditions, that are low speed and under low load. Because the robot motors are powered by a power electronical amplifier, a significant power distortion and noise, produced by a diodebridge rectifiers and PWM technology, an internal inaccuracies, electrical unbalances or system nonlinearities such as transformer saturation had to be included [7, 12].

The limitation this work has to faced with, was a necessity to keep the investigated motor being connected to the original amplifier as the only one available power source and a control unit. Simultaneously significantly distorted electrical voltages waveforms caused high values fluctuations and the motor shafts could rotate only in a limited range because of an original robot control unit safety systems. As a

Fig. 1. General view of the investigated Fanuc AM100iB robot

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

consequence the steady-states had to be recognised and analysed in a separate off-line mode. This circumstance and requirements reconstructed real exploitation conditions and can explain what measurement equipment and measurement attitude had to be considered to achieve reliable results.

In order to properly determine electrical power values for low range conditions, the proper measurement methods for noise and distorted waveforms had to be considered. From among of many known power calculation solutions, allowing for considering high disturbances, the IEEE 1459-2010 standard [8] has been chosen. The selection was affected by a pragmatic oriented goals, which were possibilities to keep the results being mostly comparable with measurements, performed by currently manufactured power analysers [6]. The most common alternative power calculation methods proper for unbalanced systems, nonsinusoidal, disturbed and noised conditions and mainly used for scientific works are Budeanu's, Fryze's and Czarnecki's methods [1, 5, 15, 18, 19].

2. The goals of investigations

The all of robot dynamical parameters are ones of its many properties, which affect the energy effectiveness of its work. In this research case, the attention was focused on identifying the power losses in the first electrical motor for the low range working conditions, particularly for low speed and low load range. A separate cognition of the electrical properties behavior while going from a driving work mode into a stand-still mode was cared about. Moreover, also those signal processing techniques were evaluated, which the measurement utilities could be extended by, in order to automatise measurement processes.

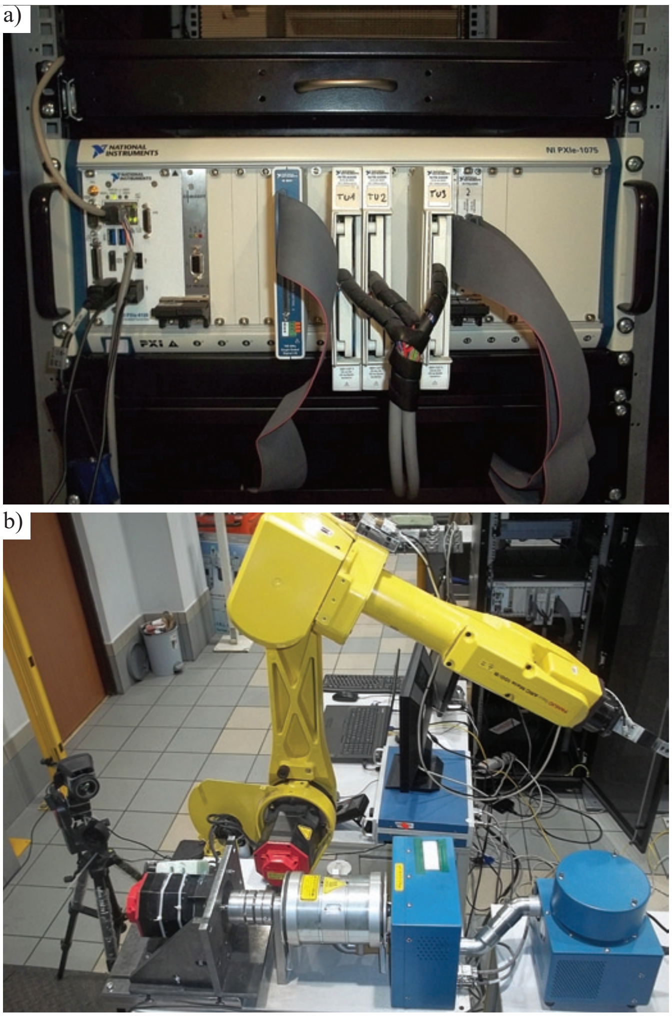


Fig. 2. a) NI PXIe-1075 system with measurement cards; b) General view of the investigated robot electric motor

3. Measurement equipment

For performing an electrical power analyses and for acquisition of the kinematic parameters of the investigated Fanuc AM100iB robot, the specialised measurement equipment was designed and built up (Fig. 2a). The apparatus consisted of NI 6581, NI 3xTB-4300B and 2x NI PXIe-6363 cards allowed for high frequency sampling of the electrical voltage and current waveforms in all power lines and for monitoring and decoding digital data containing shafts angle positions, generated by the *Fanuc Pulse Coders*. However, the alternative solution for determining kinematic parameters of electric motors shafts by a direct measurement of a robot arm angle positions - available to recalculate into shafts' angle positions, could also be considered and is widely described in the following paper [13] for TCP robot point. The all acquisition devices were synchronised and controlled by a NI PXIe-1075 system and the LabVIEW SCADA application. The mechanical load of motor shaft and mechanical power was created and measured by the hysteresis brake MAGTROL HD-715 (Fig. 2b).

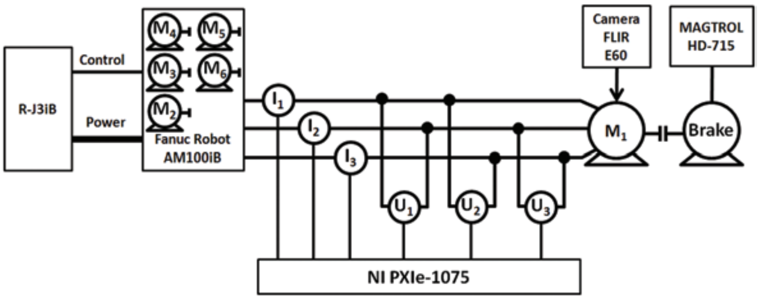


Fig. 3. Wiring diagram

The wiring diagram and the mechanical connection scheme is shown in Fig. 3. The example raw data for an electrical voltage and current waveforms collected by this equipment show the graphs in Fig. 4. The specimen of the recorded signal is designated by the black fat line, which simultaneously indicates the rms value. The signals waveforms shapes result from the limited range of an angle position the motor shaft was able to rotate.

The longer recording duration of the specimen in steady-state, the most accurately the disturbances and noise were reduced from the final power calculation and the better final quality was achieved. Mainly the specimen's recording duration influenced on the fundamental harmonic frequency identification correction from the value numerically identified by FFT analysis. Additionally the longest specimen's recording duration, the more accurate angle phases shifts between electrical voltages and currents were recognised. From the other site, the maximum specimen's recording duration was limited by the limited range of an available angle rotation.

The wiring temperature was monitored in an indirect measurement by the use of the FLIR E60 thermographic camera. The all measurements were performed for the circumstances when the motors cover's temperature was in the range between 42-48°C (Fig. 5).

4. Power calculation

The distortion and noise level in an electrical voltage and current waveforms is evaluated by the index called the total harmonic distortion (THD), which according to IEEE 519-1992 is defined as the ratio of the rms of the harmonic content to the rms of fundamental quantity and is expressed as a percent of the fundamental harmonic [17]. The total harmonic distortions – *THD*, for an electrical voltage and cur-

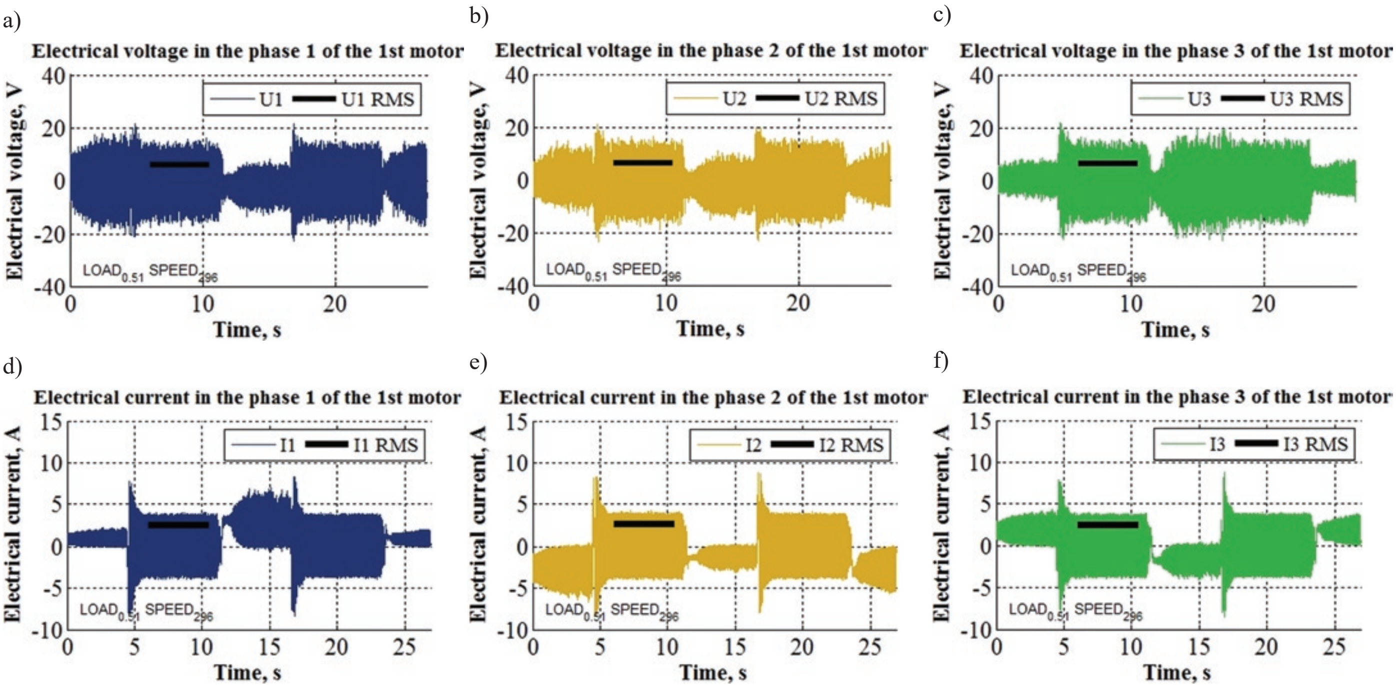


Fig. 4. Raw data

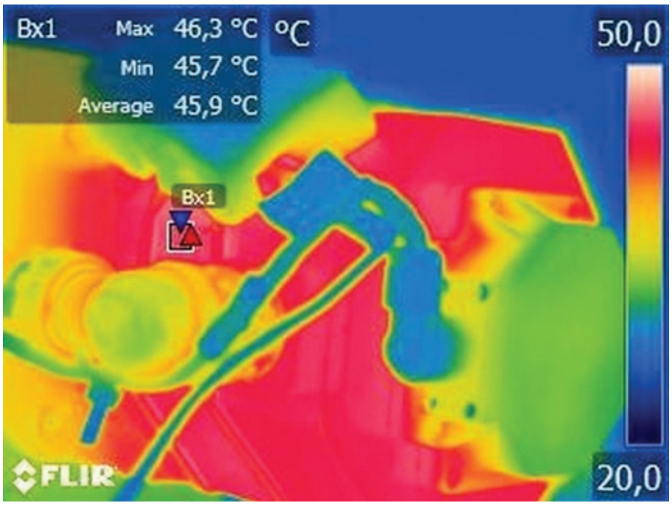


Fig. 5. Thermographic view of the electric motor cover

rent, in an accordance with the newest standard [7, 8] are expressed respectively by the equations (1) and (2):

$$THD_V = \frac{\sqrt{\sum_{k=2}^{\infty} V_{krms}^2}}{V_{1rms}} \cdot 100\% \tag{1}$$

$$THD_I = \frac{\sqrt{\sum_{k=2}^{\infty} I_{krms}^2}}{I_{1rms}} \cdot 100\% \tag{2}$$

V_{krms} and I_{krms} mean an rms value of respectively harmonic components of an electrical voltage and electrical current noise, V_1 and I_1 mean an rms value of respectively a fundamental harmonic of electrical voltage and electrical current values.

In electrical grids working with frequencies 50/60 Hz the most significant harmonics affecting the measurements are the 3rd, 5th and 7th [7]. However, because the fundamental frequencies, the number of significant harmonic components and their values varied in a wide range and changed in different working conditions, therefore instead of a pure harmonic distortion measurement the combined total harmonic distortion plus noise method (THD+N) for determining the values of THD_V and THD_I was finally used (3) [2, 14, 16, 22]:

$$THD + N = \frac{\sum_{n=2}^{\infty} \left(S_{harm} + S_{noise}\right)_{rms}}{S_{rms}} \tag{3}$$

S_{harm} means the value of an electric voltage or current harmonic distortion, S_{noise} means the value of an electric voltage or current harmonic noise and S_{rms} represents the rms value of a total electric voltage or current waveform.

To determine the fundamental electric voltage and electric current harmonics frequencies and to determine their magnitudes values and to evaluate their quality, the Fast Fourier Transform - FFT and a pattern sinusoidal signals were used. Pattern sinusoidal signals were manually tuned to the analised waveforms, extracted from the measured signals (4) and used for determining the proper sample index, which the specimens had to begin from. They were also used for determining a phase shifts between electrical voltage and current waveforms. Additionally the pattern signals were used for the correction of the fundamental frequencies values recognised by FFT if that was required. The waveforms of measured parameters became the waveforms of the total distortion and noise signals after extracting the fundamental signal harmonics from them (5). The resultant normalised distortion, noise and fundamental harmonic waveforms are shown in Fig. 6.

$$S_1 = S - S_p \tag{4}$$

$$\sum_{n=2}^{\infty} S_{harm} + S_{noise} = S - S_1 \tag{5}$$

The example spectrum analyses of electrical voltage and current waveforms for low working condition range are shown in Fig. 7. Based on the graphs in Fig. 6 and Fig. 7 the rms of total

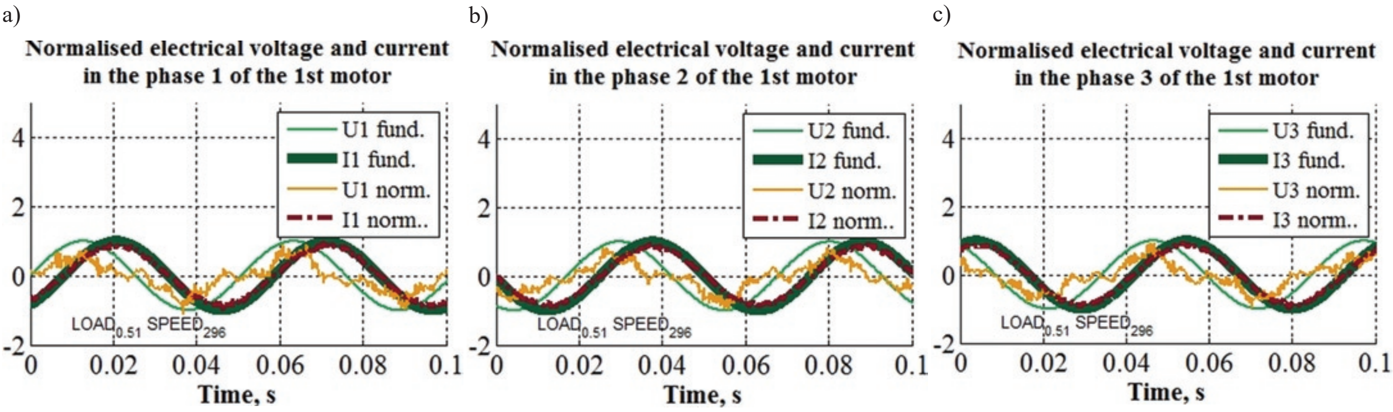


Fig. 6. Normalised electrical voltage and current waveforms and sinusoidal patterns of their fundamental harmonics in each phase of the 1st robot motor

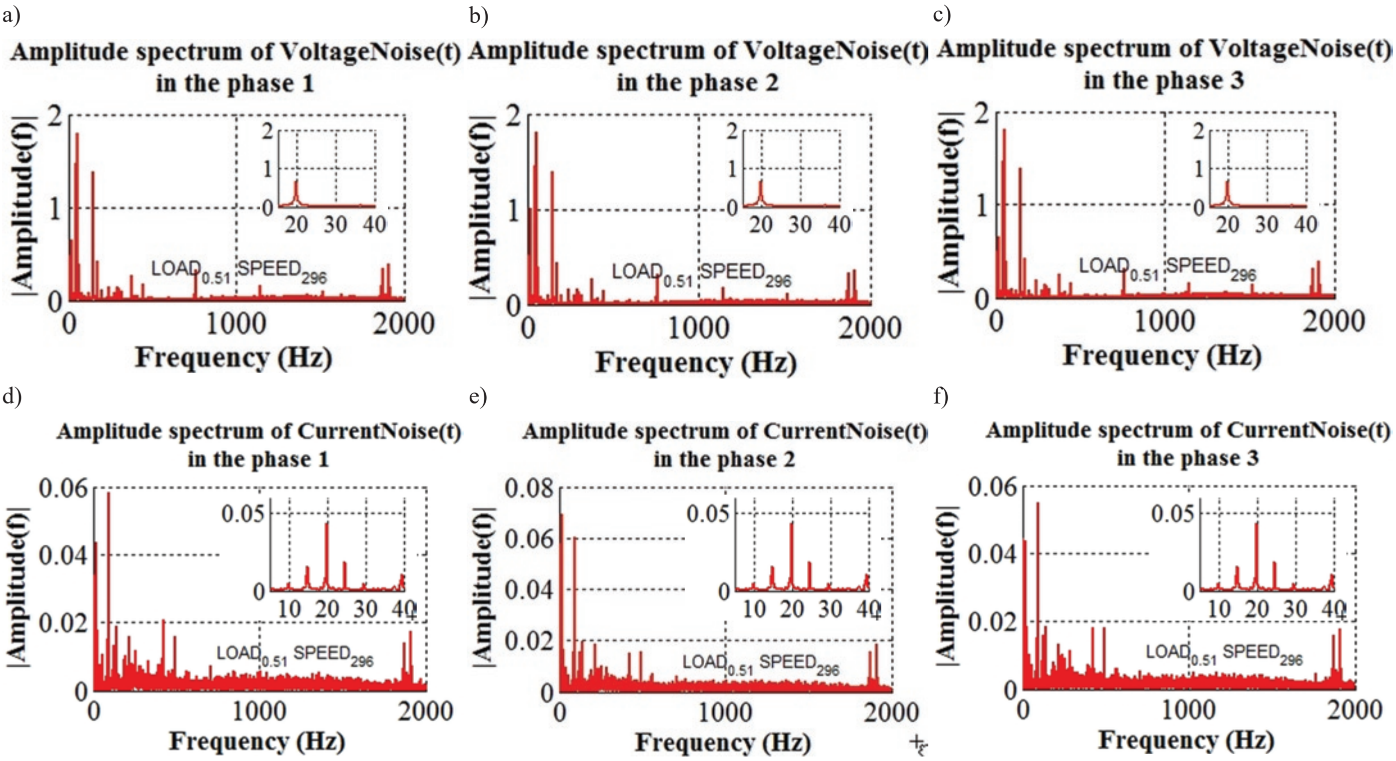


Fig. 7. Spectrum analyses for electrical voltage and current waveforms. In the top right corners the reduced fundamental harmonics are shown

disturbances and noise waveforms, the fundamental frequencies and their amplitudes were identified as well as the quality of disturbances and noise reduction was evaluated. The example relation between disturbances, noise and fundamental frequencies are shown in Fig. 8.

An electrical active power in each phase, measured by a 3V3A/Three-voltage three-current method [23], used for three-wire systems and line-to-line nonsinusoidal, unbalanced cases was calculated by (6), (7), (8) [8]:

$$P_A = V_A \cdot I_A \cdot PF_{True\,A} \tag{6}$$

$$P_B = V_B \cdot I_B \cdot PF_{True\,B} \tag{7}$$

$$P_C = V_C \cdot I_C \cdot PF_{True\,C} \tag{8}$$

V_A, V_B and V_C mean an rms value of an electrical voltage in the phase *A, B* and *C* respectively, I_A, I_B and I_C mean an rms value of an electrical current in the phase *A, B* and *C* respectively, P_A, P_B and P_C mean an electrical power in the phase *A, B* and *C* respectively, PF_{true} means true power factor (9), consisted of PF_{disp} - displacement power factor (10) and PF_{dist} - distortion power factor (11) like in [7, 8]:

$$PF_{true} = PF_{displ} \cdot PF_{dist} \tag{9}$$

$$PF_{displ} = \cos(\delta - \theta) \tag{10}$$

$$PF_{dist} = \frac{1}{\sqrt{1+(THD_V/100)^2} \cdot \sqrt{1+(THD_I/100)^2}} \tag{11}$$

where δ and θ mean voltage and current phase shifts respectively, THD_V and THD_I mean electrical voltage and current distortion and noise respectively.

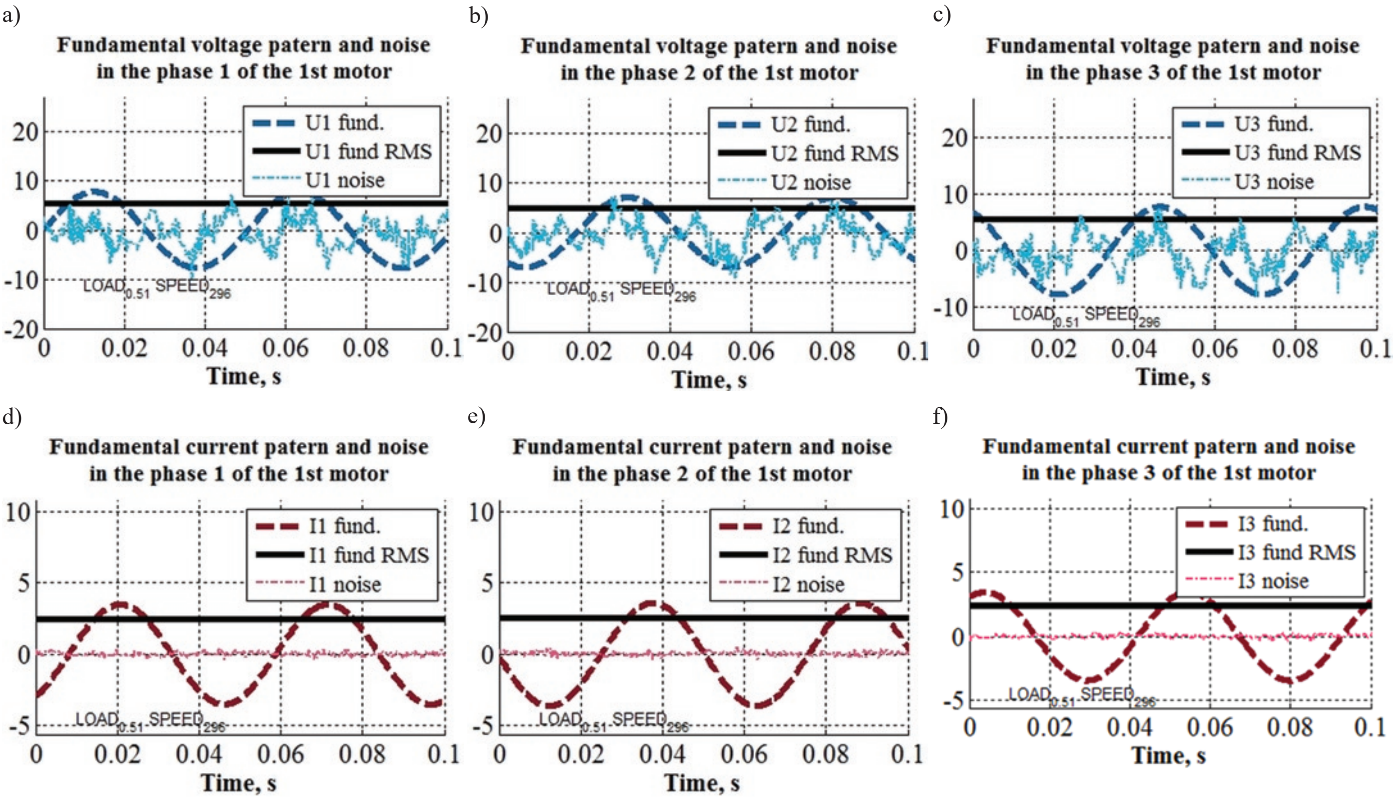


Fig. 8. Example of distortion, noise and fundamental harmonics waveforms comparison for low working conditions

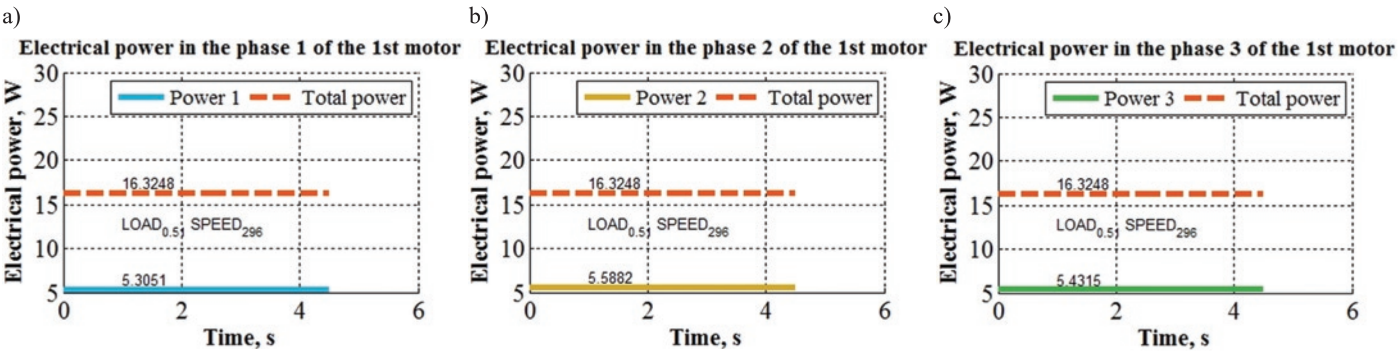


Fig. 9. Selected results for a total electrical power and electrical power in each phase for low working conditions

The total active electrical power was calculated by the equation (12) (Fig. 9), inter alia used for the three-wire *star* systems, for which the following balance $I_A+I_B+I_C=0$ is true [8]:

$$P_T = P_A + P_B + P_C \tag{12}$$

The power losses map (13) (Fig. 10a) was calculated as a difference between the total active electrical power (12) (Fig. 10c) measured indirectly by the NI PXIe-1075 system and the mechanical power (Fig. 10b) measured by the hysteresis MAGTROL HD-715 brake:

$$P_L(\tau_1,q_1) = P_T(\tau_1,q_1) - P_M(\tau_1,q_1) \tag{13}$$

The energy efficiency map was calculated as the ration between input and output power (14) (Fig. 10d):

$$(\tau_1,q_1) = \frac{P_M(\tau_1,q_1)}{P_T(\tau_1,q_1)} \tag{14}$$

Because of its comprehensive range of information, the power losses map (Fig. 10a) is the most significant input data for modelling electrical power losses, being the part of the total energy consumption, evaluated in energy effectiveness analysis of industrial production processes performed by investigated robot [20, 21, 24].

Performed investigations showed that the change of electrical states while changing work from the driving into stand-still mode is very abrupt, and in the case of investigated motor takes places between 0 and 6 rpm (Fig. 11). Additionally there is also a visible gap above 6 rpm, recorded for every load case (Fig. 11). Those facts have to be included in the numerical robot model being under development by the authors of this paper.

5. Conclusion

The measurement equipment was a proper selection for managing with the limitations and accuracy level requirements resulted from necessity to keep the investigated electric motor being connected to the original control unit and resulted from low range working condition being under investigation. The apparatus allowed for recording a raw data with its harmonics up to very high frequencies and for providing

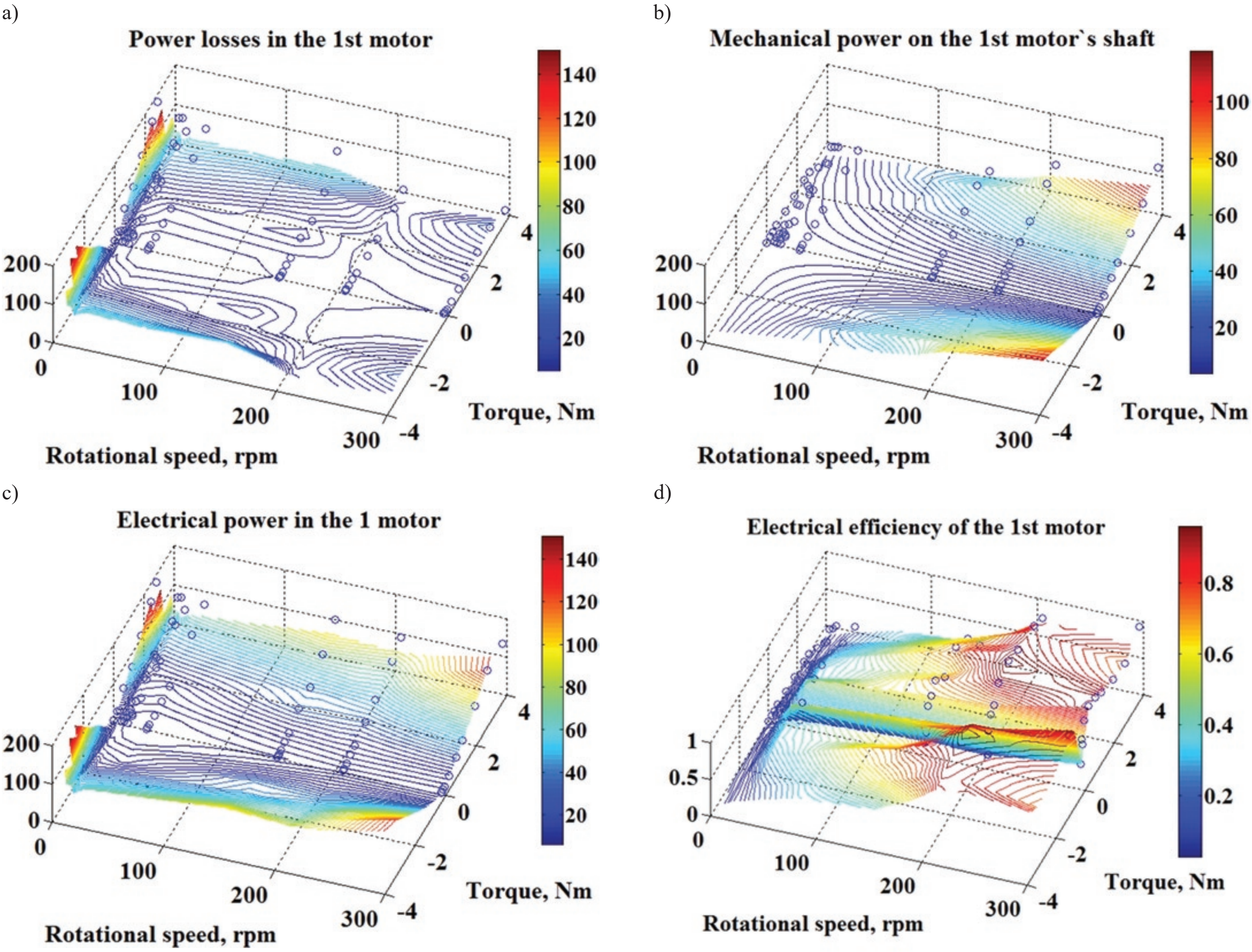


Fig. 10. Power maps for low working conditions

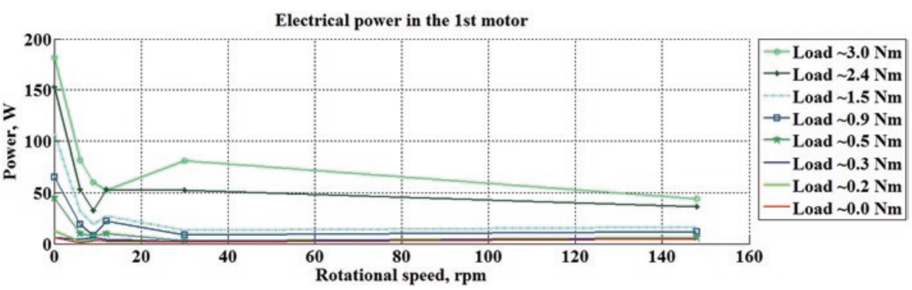


Fig. 11. Active electrical power in the 1st motor with visible abrupt electrical state change close to 6 rpm and with a gap above 6 rpm

high quality shaft's angle position. This solution could be compared to six power analysers working with extra feature to decode the digital data with values of kinematic parameters of the motor shafts.

Spectrum analyses showed that the pattern signals for notch filtering of sampled signals used for this work did not perfectly separate the fundamental harmonics from the disturbances and noise. As a consequence it extended a measurements uncertainty and decreased the final results' quality at a data processing stage. Nevertheless the level of already achieved reduction did not significantly influence on the results, what was evaluated by total power changes for more and more better reductions. Despite this the quality of band-stop filter for notch filtering of the fundamental harmonics application should be evaluated yet and compared with so far used approach before further research of energy effectiveness.

Investigations of the low range working conditions showed that the electrical states between the driving mode and the stand-still mode change very abruptly and in the investigated motor case take place between 0 and 6 rpm and additionally there is a gap visible above 6 rpm, recorded for every load case.

Even more accurate investigation of the closest area of 6 rpm is required to avoid low resolution results and a total measurement uncertainty should be also evaluated in the further work.

The work has a cognitive significance, and its results explain which direction the further research should be focused on. Because an industrial robot is an example of highly non-linear system, therefore more unbalanced working conditions and higher nonlinear loads than those ones used and existed in this investigation must be expected in typical exploitation circumstances.

Acknowledgement

The project was financed from the funds of National Science Centre awarded on the basis of a decision number DEC-2011/03/B/ST8/04631.

References

1. Akagi H, Watanabe E H, Aredes M. Instantaneous Power Theory and Applications to Power Conditioning 2017: 27, https://doi.org/10.1002/9781119307181.
2. Audio Precision. More about THD+N and THD 2013.
3. Belda K, Burget P. Physical Modelling of Energy Consumption of Industrial Articulated Robots. 15th International Conference on Control, Automation and Systems (ICCAS 2015).
4. Brossog P M, Kohl J, Merhof J, Spreng S, Franke J. Energy Consumption and Dynamic Behavior Analysis of a Six-axis Industrial Robot in an Assembly System. CIRP 2014; 23: 131-136, https://doi.org/10.1016/j.procir.2014.10.091.
5. Bucci G, Ciancetta F, Fiorucci E, Ometto A. Survey about classical and innovative definitions of the power quantities under nonsinusoidal conditions. International Journal of Emerging Electric Power Systems 2017; 18(3): 1-16, https://doi.org/10.1515/ijeeps-2017-0002.
6. Fluke 434-II/435-II/437-II. Users Manual 2012. Rev.1 06/12.
7. Grady W M, Gilleskie R J. Harmonics and how they relate to power factor. Proc. of the EPRI Power Quality Issues & Opportunities Conference (PQA'93), San Diego, CA 1993: 3-10.
8. IEEE Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions. IEEE Std 1459TM-2010; (Revision of IEEE Std 1459-2000).
9. Liu A, Liu H, Yao B, Xu W, Yang M. Energy consumption modeling of industrial robot based on simulated power data and parameter identification. Advances in Mechanical Engineering 2018; 10 (5): 1-11, https://doi.org/10.1177/1687814018773852.
10. Mohammeda A, Schmidt B, Wanga L, Gaoc L. Minimizing Energy Consumption for Robot Arm Movement. Procedia CIRP 2014; 25: 400-405, https://doi.org/10.1016/j.procir.2014.10.055.
11. Paesa K, Dewulfa W, Elsta K V, Kellensb K, Slaetsa P. Energy efficient trajectories for an industrial ABB robot. CIRP 2014; 15: 105-110, https://doi.org/10.1016/j.procir.2014.06.043.
12. Park O S, Park J W, Bae C B, Kim J M. A Dead Time Compensation Algorithm of Independent Multi-Phase PMSM with Three- Dimensional Space Vector Control. Journal of Power Electronics 2013; 13 (1): 1, https://doi.org/10.6113/JPE.2013.13.1.77.
13. Płaczek M, Piszczek Ł. Testing of an industrial robot's accuracy and repeatability in off and online environment. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (3): 455-464, https://doi.org/10.17531/ein.2018.3.15.
14. Rane. Audio specifications 2000; 2.
15. Revuelta P S, Litrán S P, Thomas J P. Active Power Line Conditioners: Design, Simulation and Implementation for Improving Power Quality 2015: 25.
16. Rhode&Schwarz. Harmonic Distortion Measurements in the Presence of Noise Application Note. June 17, 2011-0e.
17. Siemens. Harmonics in power systems, causes, effects and control 2013: 3.
18. Stevanović D, Petković P. A single-point method based on distortion power for the detection of harmonic sources in a power system. Metrology and Measurement Systems 2014; XXI (1): 3-14.
19. Szubert K. Power Measurement when waveform are distorted. Jakosc i uzytkowanie energii elektrycznej - Electrical Power Quality and Utilisation 1996; 2 (2);
20. Świder J, Zbilski A. Modelling of the power of losses in the Fanuc AM100iB robot drives and in its power electronic systems. Modelowanie inzynierskie - Engineering modeling 2013; 17(48): 138-142.
21. Świder J, Zbilski A. The modelling and analysis of a partial loads in the Fanuc AM100iB robot joints. International Journal of Modern Manufacturing Technologies 2013; V (2): 89-96.
22. Texas Instruments. Digital Audio Measurements 2001; SLAA114: p. 10.
23. Yokogawa. WT310/WT310HC/WT330 Digital Power Meter Getting Started Guide 2013: 2-25.
24. Zbilski A. Doctoral dissertation: Method for energy intensity analysis of manipulation and transportation processes 2014.

Jerzy ŚWIDER
Adrian ZBILSKI
Silesian University of Technology
Faculty of Mechanical Engineering
Konarskiego St, 44-100 Gliwice, Poland

E-mails: jerzy.swider@polsl.pl, adrianzbilski@gmail.com

Article citation info:
SKAČKAUSKAS P, SOKOLOVSKIJ E. Modification and reliability estimation of vector based Dubins path approach for autonomous ground vehicles path re-planning. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 549–557, http://dx.doi.org/10.17531/ein.2018.4.6.

Paulius SKAČKAUSKAS
Edgar SOKOLOVSKIJ

MODIFICATION AND RELIABILITY ESTIMATION OF VECTOR BASED DUBINS PATH APPROACH FOR AUTONOMOUS GROUND VEHICLES PATH RE-PLANNING

POPRAWA I OCENA NIEZAWODNOŚCI OPARTEJ NA WEKTOROWEJ METODZIE TRAJEKTORII DUBINSA DLA KOREKTY TRAJEKTORII POJAZDÓW AUTONOMICZNYCH

Due to global purposes to ensure growth of a competitive and sustainable transport system, also to solve traffic safety and environmental problems, various engineering solutions are being sought out. It can be assumed that autonomous vehicles are the technology, which will ensure the positive change in the transport system. Even though many studies successfully advanced toward realisation of autonomous vehicles, a significant amount of technical and policy framework problems still has to be solved. This paper addresses the problem of predefined path feasibility and proposes an effective methodology for a path to follow re-planning. The proposed methodology is composed of three parts and is based on the Dubins path approach. In order to modify the vector based Dubins path approach and to ensure the path feasibility, the optimisation problem was solved. A cost function with different inequality constraints was formulated. The performance and reliability of the proposed methodology were analysed and evaluated by carrying out an experimental research while using the autonomous test vehicle.

***Keywords**: autonomous ground vehicle, Dubins path, path re-planning, path following, performance, reliability estimation.*

Dla zapewnienia rozwoju konkurencyjnego i zrównoważonego systemu transportowego, oraz w celu rozwiązania problemów związanych z bezpieczeństwem ruchu i środowiskiem, poszukiwane są różne rozwiązania techniczne. Można założyć, że autonomiczne pojazdy są technologią, która zapewni pozytywną zmianę w systemie transportowym. Mimo że wiele badań z powodzeniem dotyczyło realizacji autonomicznych pojazdów, należy jeszcze rozwiązać wiele problemów technicznych i prawnych. W niniejszym dokumencie poruszono problem predefiniowanej wykonalności ścieżki i zaproponowano skuteczną metodologię dla ścieżki do śledzenia ponownego planowania. Proponowana metodologia składa się z trzech części i opiera się na metodzie Dubinsa. Aby zmodyfikować metodę trajektorii Dubinsa i zapewnić optymalną trajektorię, w publikacji rozwiązano zadanie optymalizacji. Sformułowana funkcja celu z różnymi nieliniowymi ograniczeniami. Skuteczność i niezawodność proponowanej metodologii została przeanalizowana i oceniona po przeprowadzeniu eksperymentalnych badań z wykorzystaniem autonomicznego pojazdu badawczego.

***Słowa kluczowe**: samochód autonomiczny, trajektoria Dubinsa, przeplanowanie trajektorii, śledzenie trajektorii, skuteczność, ocena niezawodności.*

1. Introduction

In research papers a number of investigations and solutions on how to increase mobility, passenger comfort, traffic safety or cut carbon emissions, etc. can be found. For example, technological innovations for traffic safety improvement are proposed in [8, 36], reliability improvement propositions of urban / commercial transport and transportation are given in [28-32], decision making methods in automotive industry are reviewed in [34] and so on. However, to achieve all of the mentioned goals, a technological breakthrough is necessary. In [18] it is stated that the main breakthrough in the transport system will be brought by the technology of autonomous vehicles. As also pointed out in [22], nowadays it is assumed that the autonomous ground vehicles (AGVs) have a great potential to be widely applied in a variety of fields, such as road transportation, agriculture, planetary exploration, military purpose and so on. Despite the scientific advances that have been made throughout the last decade, there are still a number of problems and challenges, related to reliability of the autonomous vehicles, human-vehicle interaction, path planning and following, control systems, driving stability, policy framework, etc. To realize this essential improvement of the current transport system, all the mentioned problems must be solved, i.e., the AGVs reliability and safety must be evident. Thus, as stated in [6], reliability of the autonomous vehicles as a new aim for quality and reliability engineering can be described.

2. Scientific background

One of the major factors, on which depends reliability and a successful integration of the AGVs into the transport system, is accurate and safe path following. Based on [13], the autonomous path following problem can be divided into two key steps: extraction of the desired path to follow and lateral / longitudinal control of the vehicle. According to [25], the most fundamental requirement for a path is feasibility – it must be possible for the planned path to be executed by the AGV. However, in most of the research works it is assumed that the predefined path to follow is safe and feasible and the main focus is the lateral / longitudinal control of the AGV. Such assumption becomes incorrect in real driving scenarios, when an on-board planner or a human cannot evaluate or does not have sufficient information about obstacles, road network, mobility and dynamic limits of the vehicle. According to [15], a human driver does not have a precise path in mind when driving; instead, s/he would normally have a global sense how they should drive, reach the destination, avoid an obstacle, etc. Under such circumstances, safe deviation from the predefined path and the path re-planning becomes an important task. Based on researches described in different sources, three conditions can be singled out, under which the path must be re-planned: 1. Avoiding dangerous situations, like spinning out or obstacle hitting (described in [7]); 2. Sharp turns and discontinuities in the path which may compromise the integrity of the AGV (described in [35]), and as singled out in [25] – 3. Neglected constraints and mobility of the vehicle.

In various literature different algorithms can be found, which are related to the planning, re-planning and optimisation of the paths of the AGVs. In the majority of cases, the constraints for the planned, re-planned or optimised path are minimum path length, minimum energy spent to follow the path, shortest time distance, etc. One of the pioneering works about smooth path generation based on energy minimization was presented in [12] by Delingette. Delingette [12] developed a method based on the usage of intrinsic splines with polynomial curvature profile which allowed to solve general geometric constraints. Liang and Liu [23] developed the shortest path planning method, which searched for a minimal length path from all the paths generated by a linear programming optimisation. The main advantage of the developed method – it can be applied while considering backward motion capability. However, in both works [12, 23], the developed algorithms were theoretically based, experimental investigations were not performed. Arokiasami [3] proposed a vector based path generation method, which follows a geometric approach to path-generation and path-tracking and is suitable for aerial and ground autonomous vehicles. Gupta [16] used a sampling based model predictive optimisation for feasible path planning and pointed out that a single-objective path length minimization can lead to trajectories that have unnecessary energy consumption. One drawback of the used method – it consists of kinematic, dynamic and power models of the AGV and is rather complex. Castillo [9] described the use of a genetic algorithm for the point-to-point path planning, while using single-objective and multi-objective optimisation (minimization of path length and difficulty). In [11] the path planning problem was also formulated as a multi-objective problem, while focusing on the energy consumption and the path safety. Although multi-objective optimisation is more suitable and useful for difficult path planning problems than single-objective optimisation, from researches provided in [9, 11] it is notable that the complexity of multi-objective path planning problems is very important in the efficiency of the algorithm performance. Finding an optimal solution in path planning can become difficult, thus, real time application of the algorithm can be complicated. To move in the shortest and collision-free path in environments with obstacles Han and Seo [17] suggested a methodology based on a surrounding point set algorithm. An advantage of the proposed methodology is that, when narrow spaces exist in the optimal path, the proposed approach does not fail to place points on the related space. Krishnan [20] defined the path length minimization problem in polar and not in the Cartesian coordinate system. Authors in [20] state that the migration from the Cartesian coordinate system to the polar coordinate frame has improved the efficiency of the proposed algorithm, however, in order to ensure better performance, environment limitations should be addressed while using an online guidance system that can make finer corrections to the trajectory on a real time basis. Human-vehicle interaction during the path planning was referred in [33] by Receveur. Receveur [33] described a multi-criteria path optimisation, listed the criteria that need to be minimized in technical order of importance, but also stated that a human can have different preferences for the path planning. Though the algorithm described by Receveur effectively generated a close to human-like trajectory, the author still suggests that the algorithm should be improved by periodically recalculating the optimal trajectory. It can be seen that the path planning, re-planning and optimisation problems have been studied using various methodologies which were theoretically and experimentally validated using various car-like mobile robots or other AGVs. Davoodi [11] noticed that since there are many types of robots and AGVs with different abilities and constraints, it is nearly impossible to provide a unique exact definition for the expression "optimal path" in the path planning context. For example, a typical vehicle cannot instantaneously move orthogonally to the wheel plane. The nonholonomic and other constraints, associated with their movement and mechanical design, need to be considered. Furthermore, as mentioned above, the main focus in such research works is the minimization of selected parameters, with less or no concern to the path feasibility, which leads to undesired deviations from the path and other negative effects. That is why one of the major issues associated with the path planning and the feasibility of the path is related to the mobility of the vehicle. Because of these reasons, the aim of this work is to develop a reliable and effective methodology for re-planning of sharp turns in a path, predefined by an on-board planner or a human, while focusing on the constraints and the mobility of the vehicle.

3. Methodology

In this paper the proposed methodology is composed of three parts. First of all, to develop a methodology for the re-planning of sharp turns in a path, while focusing on the constraints and the mobility of the vehicle, a bicycle-like kinematic model of the AGV is considered. Secondly, the vector based Dubins path approach for the path re-planning is given and explained. And finally, in order to increase the path feasibility and safety, a cost function with inequality constraints for the vector based Dubins path approach is proposed.

3.1. Kinematic vehicle model

In the considered kinematic model, it is assumed that the two wheels on the front and on the rear axles are collapsed into a single wheel, respectively located at the midpoints of both axles. The front wheel can be steered, and the orientation of the rear wheel is fixed. The main feature of the kinematic model of the AGV is the presence of the nonholonomic constraints. According to [1], the bicycle-like kinematic models of the AGVs are widely used because of their low parameter dependency and the usage of such models is a quite standard assumption in the literature where analysis of different control strategies is performed [2] or new path planning and following methods as in [4, 5], are developed. It is assumed that the AGVs have planar motion, therefore the vector of the generalized Cartesian coordinates of the centre of mass is:

$$Q = \begin{bmatrix} x \\ y \\ \theta \end{bmatrix}, \tag{1}$$

where x and y are the position coordinates of the centre of mass; θ is the vehicle orientation with respect to the fixed frame (X,Y) of the Cartesian coordinates.

Considering the variables indicated in the Eq. (1), the kinematic model of the vehicle can be represented as:

$$\begin{cases} \dot{x} = v \cdot \cos\theta \\ \dot{y} = v \cdot \sin\theta \, , \\ \dot{\theta} = \dfrac{v}{L} \tan\delta \end{cases} \tag{2}$$

where L is the AGV wheelbase; δ is the steering angle of the front wheels; v is the linear velocity of the vehicle.

The average steering angle of the front wheels can be described geometrically as:

$$\delta = \tan^{-1}\left(\frac{L}{r}\right), \tag{3}$$

where r is the radius of the path of the vehicle.

In the methodology proposed below, the kinematic model is hereinafter used for expressing the desired movement coordinates, while evaluating the constraints of the vehicle accordingly.

3.2. Vector based Dubins Path Approach

Based on research works [14, 19, 26], in which path planning problems with a turning radius motion constraint are addressed, it can be stated that the Dubins path approach for the AGVs, often referred to as the Dubins vehicle, is one of the most effective, widely used and modified classic methods for the optimal path planning. When applying this method in the general case, finding the optimal path involves checking for 6 possibilities, which consist of concave or convex circle segments and / or straight line segments. In this work, the purpose of applying the Dubins path approach is not to generate a large number of paths that eventually will be discarded, but, based on the path waypoints / intersections predefined by an on-board planner or a human, to solve the optimisation based path planning problem and find the optimal Dubins path possibility for path re-planning. The waypoints predefined by an on-board planner or a human can be converted into Dubins path by inserting a filleted circular arc near the intersection waypoints. A general idea of the vector based Dubins path approach for path re-planning, further used in this work, originally was presented in [19] for unmanned aerial vehicles. The scheme of a path re-planning example using the vector based Dubins path approach is presented in Fig. 1.

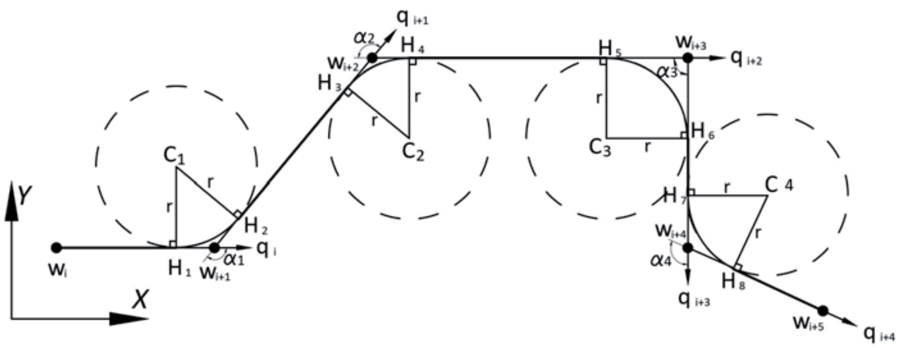


Fig. 1. Scheme of path re-planning using the vector based Dubins path approach

In the provided scheme (Fig. 1) w_i are referred to as path waypoints predefined by a human or an on-board planner (x and y coordinates) which are connected by the straight line segments and form a global path – path with sharp turns to follow. The AGVs movement directions in straight line segments are shown by unit vectors q_i, which are expressed:

$$q_i = \frac{w_{i+1} - w_i}{\left\| w_{i+1} - w_i \right\|}, \tag{4}$$

$$q_{i+1} = \frac{w_{i+2} - w_{i+1}}{\left\| w_{i+2} - w_{i+1} \right\|}. \tag{5}$$

Based on [19], since the AGV needs a command to switch from the straight to a circular moving segment, the coordinates, where the transition happens, have to be defined. To define these coordinates, the angles α_i of the sharp turns have to be expressed:

$$\alpha_i = \cos^{-1}\left(-q_i^T \cdot q_{i+1}\right). \tag{6}$$

Additionally, the transition distances k_i from the predefined waypoints w_i to the transition points H_i have to be found:

$$k_i = -\left(\frac{r}{\tan\frac{\alpha_i}{2}}\right) q_{i+1}. \tag{7}$$

Then, the transition points (coordinates) H_i where the movement changes from the straight line into a circular regime are expressed:

$$H_i\left(w_{i+1} - k_i, q_i\right) \tag{8}$$

$$H_{i+1}\left(w_{i+1} + k_i, q_{i+1}\right). \tag{9}$$

The coordinates of the circular arcs centre points – Dubins circles C_i, in which the AGV moves during the turning-circular regime, are calculated:

$$C_i = w_{i+1} - \left(\frac{r}{\sin\frac{\alpha_i}{2}}\right) \frac{q_i - q_{i+1}}{\left\| q_i - q_{i+1} \right\|}. \tag{10}$$

The same approach is used for all the waypoints / intersections in the predefined path.

3.3. Modification of Vector based Dubins Path Approach

According to [21], the Dubins path approach for the AGVs is based on three assumptions: 1. Vehicle moves at a constant velocity; 2. Vehicle cannot move in reverse, and 3. Vehicle has a minimum turning radius. From [37] it is clear that, with the mentioned assumptions, the steady state cornering condition is considered. In research works such as [14, 19, 25, 27], where the Dubins path approach is applied and modified, the assumption that a vehicle always moves in a minimum turning radius is considered to be an advantage of this method. In this paper, the assumption that a vehicle always moves in a minimum turning radius is deemed not as an advantage but as a disadvantage. This proposition can be explained by the fact that the steering angle of the front wheels change is not a discrete function. After reaching the transition points where moving has to switch from the straight line into the turning-circular regime, the steering angle of the front wheels at the same time point does not change from neutral to the desired and vice versa. The change of the vehicle steering angle of the front wheels depends on the angular velocity of the steering wheel turning, the steering ratio and other mechanical constraints. The same proposition can be applied to steer-by-wire systems. This means that at the beginning of the turning-circular regime, the actual steering angle of the front wheels is not equal to the predefined steering angle of the front wheels and, because of this reason, the undesired deviations from the predefined path appear and are always increasing. If the turning radius is minimum, then the amount of time during which the desired value of the steering angle is reached will be longer, and, respectively, the undesired deviations will be growing a longer amount of time and will be larger. That is why it is needed to optimise the turning radius, evaluating all the mentioned variables. To solve this problem and modify the vector based Dubins path approach, a cost function, which ensures the least possible deviations from the predefined straight line segments, is formulated:

$$\min f = \Sigma\left(x - x_p\right)^2 + \Sigma\left(y - y_p\right)^2 + \Sigma\left(\theta - \theta_p\right)^2, \qquad (11)$$

where x_p, y_p, θ_p respectively, are the coordinates of the predefined path.

Because the optimised parameter is turning radius r, also seeking to evaluate the nonholonomic constraints of the AGV, Eq. (11) must be rearranged. The rearrangement is done based on the kinematic AGV model, by inserting Eq. (3) into Eq. (2), integrating the obtained expression and inserting the result into Eq. (11). After rearranging, the cost function expression is obtained:

$$\min f = \Sigma\left(x - x_p\right)^2 + \Sigma\left(y - y_p\right)^2 + \Sigma\left(\left(\frac{v \cdot \tan\left(\tan^{-1}\left(\frac{L}{r_i}\right)\right) \cdot t_i}{\sqrt{1 + \frac{l_r^2 \cdot \tan\left(\tan^{-1}\left(\frac{L}{r_i}\right)\right)^2}{L}}}\right) - \theta_p\right)^2, \qquad (12)$$

where r_i is the optimised turning radius for every sharp turn in the re-planned path; l_r is the distance from the centre of mass to the rear axle; t_i is the demanded movement time from one transition point to the next transition point.

In the non-modified Dubins path approach it is assumed that the turning radius r for every re-planned sharp turn is the same (Fig. 1, Fig. 4, part b). Seeking to achieve better re-planning results, in the proposed methodology it is considered that, based on the straight line segments lengths, vehicle velocity, etc., the turning radius r_i is optimised separately for every sharp turn, which is re-planned. Consequentially, based on the statement mentioned above, that the proposed cost function has to be applied separately to every sharp turn that is being re-planned, it is clear that Eq. (12) is only suitable for the re-planning of the first sharp turn. For the rest of the sharp turns re-planning, Eq. (12) has to be appended:

$$\min f = \Sigma\left(x - x_p\right)^2 + \Sigma\left(y - y_p\right)^2 + \Sigma\left(\left(\frac{v \cdot \tan\left(\tan^{-1}\left(\frac{L}{r_i}\right)\right) \cdot t_i}{\sqrt{1 + \frac{l_r^2 \cdot \tan\left(\tan^{-1}\left(\frac{L}{r_i}\right)\right)^2}{L}}}\right) + \theta_{i-1} - \theta_p\right)^2, \qquad (13)$$

where θ_{i-1} is the vehicle orientation at the previous transition point.

In this methodology, lower and upper bounds for the turning radius optimisation problem, in which iterations must stay, are defined as a maximal steering angle of the front wheels $-\delta_{max} \le \delta \le \delta_{max}$, which cannot be exceeded due to technical vehicle limits / capacity. Respectively, the lower and upper bounds for Eq. (12) and Eq. (13) are $-r_{max} \le r \le r_{max}$. Seeking to ensure path integrity and safety, also to ensure the reliability and simplicity of the method, two different inequality constraints are proposed. In the case of the first inequality constraint, to utilise all of the six optimal Dubins path possibilities, such inequality constraint is formulated:

$$\sqrt{\frac{r_i^2 \cdot q_i(x)}{\tan\left(\frac{\alpha_i}{2}\right)^2} + \frac{r_i^2 \cdot q_i(y)}{\tan\left(\frac{\alpha_i}{2}\right)^2}} - \frac{q_{i-s}}{2} \le 0, \qquad (14)$$

where $q_i(x, y)$ respectively, are the length of the corresponding unit vector in the x and y directions; q_{i-s} is the length of the shorter straight line segment, connected to the waypoint / sharp turn, which is re-planned.

In this inequality constraint case, the transition distances k_i from the predefined waypoints w_i cannot be less than half of the shorter straight line segment, connected to the waypoint / sharp turn, which is re-planned. That is why, depending on the shortest straight line segments lengths, all of the six optimal Dubins path possibilities are available.

During the optimisation of every turning radius r_i, it is necessary to evaluate that the steering angle of the front wheels change is not a discrete function. It is accepted that, in the kinematic model, the steering angle of the front wheels change is described as a time varying function:

$$\delta_i(t) = \frac{\pm d \cdot \omega \cdot t}{S_R}, \qquad (15)$$

where d is the direction of the steering angle of the front wheels (negative – to the right, positive – to the left); ω is the angular velocity of the steering wheel turning; t is time; S_R is the steering ratio.

When using the described methodology and separately optimising every turning r_i, the transition distances k_i from the predefined waypoints w_i, the transition points H_i, the coordinates of the circular arcs centre points C_i also have to be recalculated separately while applying the already described vector based Dubins path approach. Because the steering angle of the front wheels change is not a discrete function, in order to ensure feasibility of the re-planned path, the turning of the steering wheel must be started not at the exact same time as the first transition point H_i is reached. The transition point T_1 for the start of the turning of the steering wheel is expressed:

$$T_1 = \sqrt{\left(H_1(x) - w_1(x)\right)^2 + \left(H_1(y) - w_1(y)\right)^2} - \frac{\tan^{-1}\left(\frac{L}{r_1}\right) \cdot v \cdot S_R}{\omega}, \qquad (16)$$

where $H_1(x, y)$ respectively, are the x and y coordinates of the recalculated first transition point; $w_1(x, y)$ respectively, are the x and y coordinates of the first waypoint.

Based on Eq. (16), the second inequality constraint case can be formulated by revising Eq. (14):

$$\sqrt{\frac{r_i^2 \cdot q_i(x)}{\tan\left(\frac{\alpha_i}{2}\right)^2} + \frac{r_i^2 \cdot q_i(y)}{\tan\left(\frac{\alpha_i}{2}\right)^2}} - \left(\frac{q_{i-s}}{2} - \frac{\tan^{-1}\left(\frac{L}{r_i}\right) \cdot v \cdot S_R}{\omega} \right) \leq 0. \quad (17)$$

In this inequality constraint case, the transition distances k_i from the predefined waypoints w_i will be less than half of the shorter straight line segment, connected to the waypoint / sharp turn, which is re-planned. Then, only four of the optimal Dubins path possibilities will be utilised, which necessarily involves the concave / convex circle segments connection with the straight line segments.

Based on each optimised turning radius r_i , the time bounds for every steering angle of the front wheels change, when the constant value of the steering angle must be maintained, is expressed as:

$$T_{UB\_i} = \frac{\tan^{-1}\left(\frac{L}{r_i}\right) \cdot S_R}{\omega}. \quad (18)$$

At the transition point, where the moving of the AGVs has to shift from the turning-circular regime into the straight line regime, the time bounds for the steering angle of the front wheels change from desired to neutral are also determined by applying Eq. (18).

Taking into consideration various researches, described in [9, 11, 14, 24], which analyse different algorithms for path optimisation, it can be stated that genetic algorithm as a robust popular heuristic search method has been extensively used to solve various single-objective and multi-objective path planning problems in discrete and continuous spaces by many authors. Thus, in this paper, to solve the described optimisation problem, a genetic algorithm was used.

4. Experimental research

In order to investigate which inequality constraint of the proposed methodology is more effective and reliable, also to execute the experimental research, the described modified vector based Dubins path approach for the path re-planning was designed in the *MATLAB/Simulink* software package. For the lateral AGV control purposes, the kinematic-based controller described in [1] was used and also designed in the *MATLAB/Simulink* software package. Basic level of vehicle autonomy, i.e., autonomous steering, was implemented by using an automated steering device, developed in the *Vilnius Gediminas technical university*, with *Arduino* microcontroller (software and hardware), mounted on the test vehicle (Fig. 2). During the experimental research, the modified vector based Dubins path approach for the path re-planning and the kinematic-based controller developed in *MATLAB / Simulink* software were connected in real time with the *Arduino* microcontroller of the automated steering device while using the universal asynchronous receiver-transmitter (UART) based communication system. To satisfy the real time communication condition and to not overload the data buffers of the used communication system, the serial signal sending and receiving rates were 10 Hz. The step size of the mathematical operations solver used in the controller was fixed at 0.1 s. In [10] a similar method to achieve the basic level of vehicle autonomy, while using the *Arduino* microcontroller, is described.

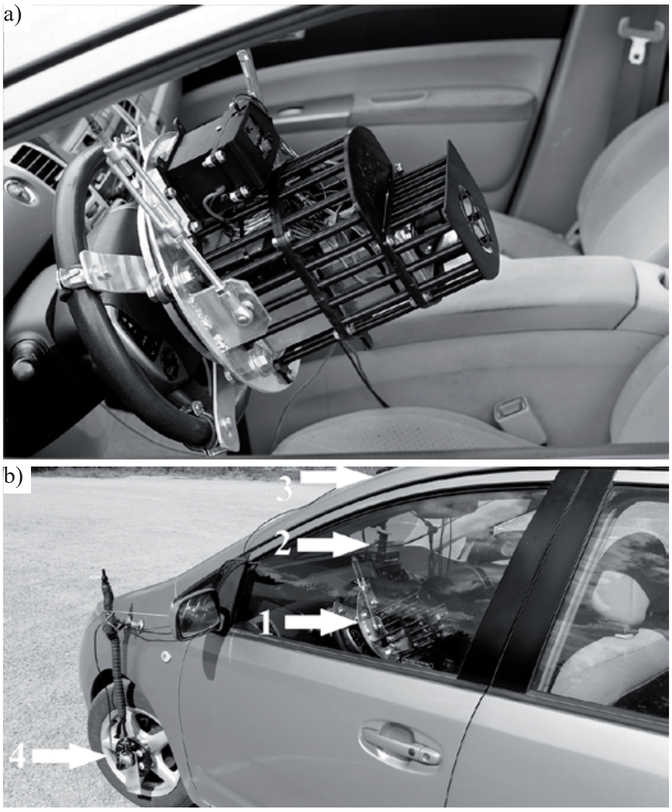


Fig. 2. Experimental research: a – automated steering device mounted on the test vehicle; b – autonomous test vehicle with 1 – automated steering device, 2 – GPS tracker, 3 – GPS antenna, 4 – angular velocity sensor

Seeking to perform reliability analysis of the proposed methodology, an experimental research was executed while re-planning the same path, predefined by a human, using different proposed inequality constraints (Fig. 3). The length of the longer side of the predefined path was 45 m, the length of the shorter side of the predefined path was 32 m. The angular velocity of the front wheels, the angle of the steering wheel, and the steering angle of the front wheels were the input parameters in the used controller. The actual movement coordinates were recorded with a GPS tracker, that was installed in the autonomous vehicle (Fig. 2, part b), upon the frequency of 100 Hz.

During the path following, an approximately constant velocity of 2.5–3.05 m/s of the test vehicle was maintained. The velocity of the test vehicle was controlled by a human supervisor for safety reasons.

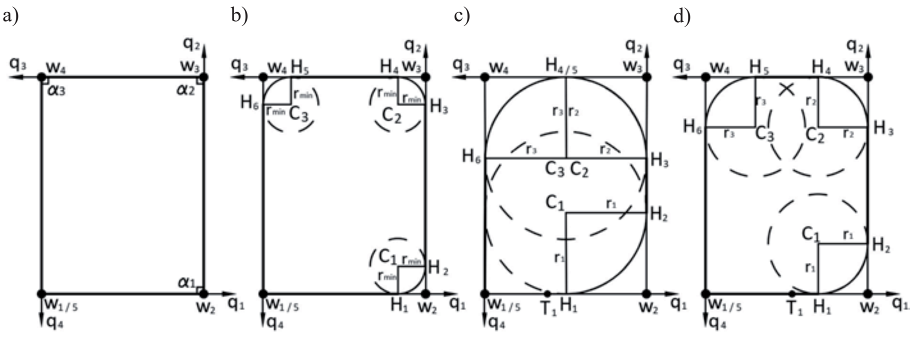


Fig. 3. Basic schemes of the predefined and re-planned paths: a – the predefined path; b – the path re-planned while using the non-modified Dubins path approach; c – the path re-planned while using Eq. (14); d – the path re-planned while using Eq. (17)

The human supervisor did not interact with the steering wheel at all. The experimental researches were done under equal conditions in an enclosed lot on a dry asphalt, without vertical slope surface at the temperature of 15–20°C. To ensure quantitative reliability analysis of the developed methodology, overall, 32 experimental drives, i.e., 16 experimental drives per each inequality constraint case, were performed.

5. Results and discussion

During all the experimental drives in the first inequality constraint case, the recorded moving trajectories were similar (Fig. 4, part b), and, respectively, during all the experimental drives in the second inequality constraint case, the recorded moving trajectories were also similar (Fig. 5, part b). Although in Fig. 4, part b and Fig. 5, part b some inaccuracies can be seen in the recorded moving trajectories, actual deviations or other undesired errors did not occur during the path following. The inaccuracies that can be seen in the recorded moving trajectories can be explained by the low performance of the used GPS tracker in some parts of the path. Due to these reasons, seeking to clearly describe the performance and reliability of the proposed algorithm, for further investigation of the proposed methodology, a specific experimental drive from each inequality constraint case was selected.

In Fig. 4 (part a) and Fig. 5 (part a) the values of the steering angle of the front wheels are presented, which were predefined while using the described modified Dubins path approach with different inequality constraints. Respectively, the actual values of the steering angle of the front wheels are also presented, which were recorded in real time as feedback data for the used kinematic-based controller during both cases of the experimental drives. For proper understanding of the results given in Fig. 4 and Fig. 5, it must be pointed out that during the moving in the turning-circular regime, the change of the steering angle of the front wheels was linear. Discretisation of the feedback signal is seen due to the selected serial signal receiving rate value (10 Hz). i.e., the frequency of the feedback data receiving into the data buffer. Discretisation of the feedback signal would be less visible if the serial signal receiving rate value was higher, however, in that case, the real time communication condition would not be satisfied because of the overloaded data buffer of the used *Arduino* microcontroller. Similarly, due to the selected serial signal receiving rate value and the discretised feedback signal, in both figures (Fig. 4, part a and Fig. 5, part a) it can be seen that there is a slight delay (~0.5 s) between the predefined values and the recorded feedback data values of the actual steering angle of the front wheels. Though, when the feedback signal is delayed, the control signal is formulated based on the delayed data and, respectively, is inaccurate, the significance of the mentioned delay to the path tracking in any case of the experimental drives was not observed.

The re-planning of the first sharp turn (point w_2, Fig. 3) does not clearly visibly reflect the difference between the re-planned first turning radius r_1, while using different proposed inequality constraints. However, as it should be, the turning radius r_1 (16.0 m), which was re-planned while using the case of the first inequality constraint (Eq. 14), is larger than the turning radius r_1 (14.8 m), which was re-planned while using the case of the second inequality constraint (Eq. 17). This can be explained by the assumption proposed above that, in the case of the first inequality constraint, seeking to use all of the six optimal Dubins path possibilities, the transition distance k_1 from the predefined waypoint w_2 cannot be less than half of the shorter straight line segment, connected to the waypoint / sharp turn (from w_1 till w_2), which is re-planned (w_2). This means that the transition point H_1 cannot be closer to the predefined waypoint w_2 than the middle point of the straight line segment from w_1 until w_2. In the case of the second inequality constraint, as already described, seeking to use only four of the optimal Dubins path possibilities, the transition point H_1 must be closer to the predefined waypoint w_2 than half of the shorter straight line segment, connected to the waypoint / sharp turn, which is re-planned (w_2). Thus, in the first re-planned sharp turn, and, respectively, in other re-planned sharp turns, moving in the turning-circular regime starts later while using the case of the second inequality constraint. Although the turning radiuses are different, from the presented results (Fig. 4 and Fig. 5) it can be seen that in both inequality constraint cases, the re-planned first sharp turn is feasible for the AGV.

The difference in the re-planned path, while using different inequality constraints, becomes more visible when second and third sharp turns are being re-planned. Because, in the re-planning of the second and the third sharp turns, the shorter straight line segment coincide – from w_3 till w_4, in the case of the first inequality constraint, the transition points H_4 and H_5 also coincide (Fig. 3, part c). Due to the reasons that all six optimal Dubins path possibilities can be used and that the transition points H_4 and H_5 also coincide, at the joint transition point $H_{4/5}$ the steering angle of the front wheels is constant. Seeking to move in the re-planned path, the constant value of the steering angle (0.16 rad) is maintained until the transition point H_6, which defines the change of moving from the turning-circular regime to the straight line regime (Fig. 4, part a). While the case of the first inequality constraint was used, four changes of the steering angle

a)

b)

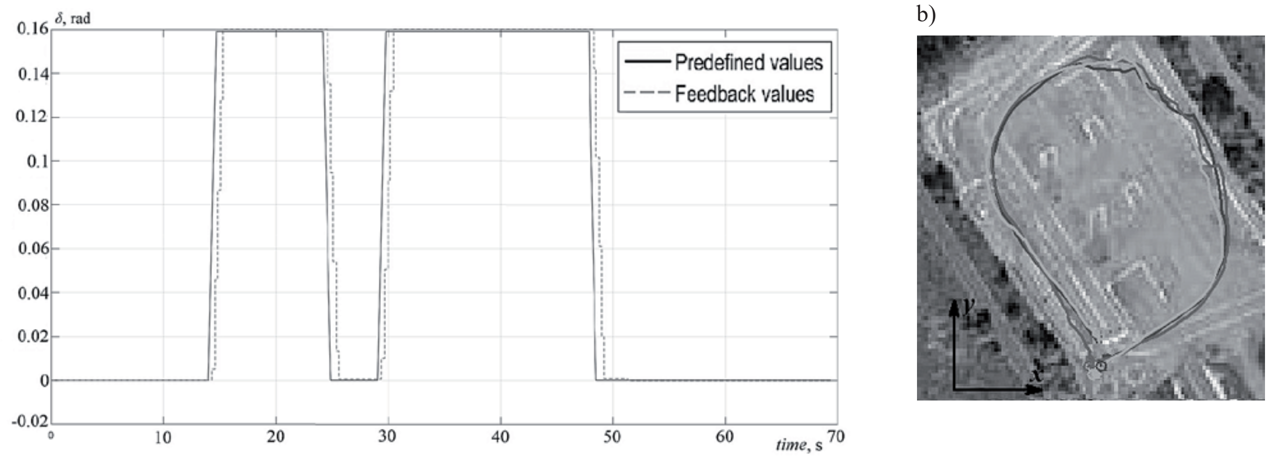


Fig. 4. Data of the performed experimental drives while using Eq. (14): a – the predefined values and the experimental drive feedback data values of the actual steering angle of the front wheels; b – example of the performed experimental drives while moving in a path, re-planned while using Eq. (14)

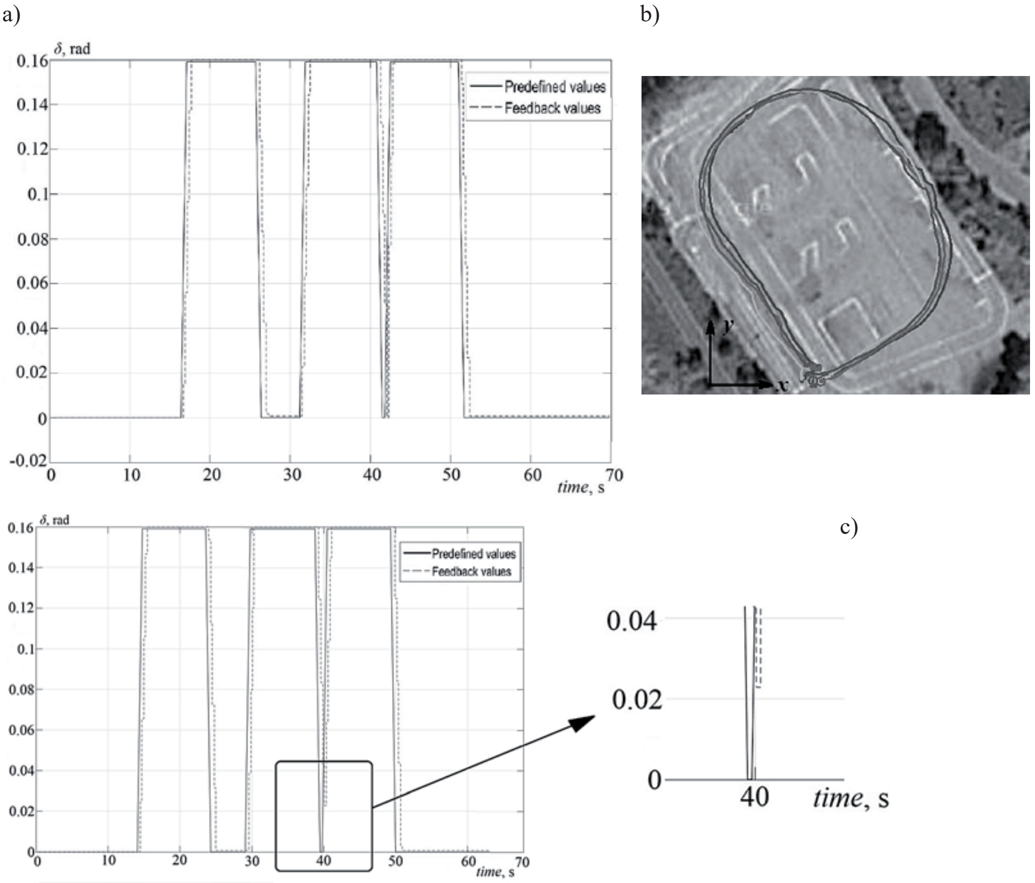


Fig. 5. Data of the performed experimental drives while using Eq. (17): a – the predefined values and the experimental drive feedback data values of the actual steering angle of the front wheels; b – example of the performed experimental drives while moving in a path, re-planned while using Eq. (17); c – the experimental drive feedback data values of the actual steering angle of the front wheels when front wheels were not returned to neutral position

of the front wheels were performed and the re-planned trajectory was feasible for the AGV (Fig. 4). In the case of the second inequality constraint, due to the assumptions mentioned above, the transition points H_4 and H_5 do not coincide. That is why, after reaching the transition point H_4, the front wheels are returned to the neutral position seeking to move in the straight line regime and, after reaching the transition point H_5, the turning of the steering wheels is being performed again to move in the turning-circular regime (Fig. 5, part a). While the case of the second inequality constraint was used, six changes of the steering angle of the front wheels were performed (Fig. 5). However, during the analysis of the results it was observed that when the moving in straight line segment time, due to the AGV velocity, is shorter than time, determined by using Eq. (18), then there is no possibility to accurately return the front wheels into the neutral position (Fig. 5, part c). When the AGV velocity was lower and the moving in the straight segment time was longer than time, determined by using Eq. (18), the front wheels were returned into the neutral position, as predefined (Fig. 5, part a). Based on the results, which are presented in Fig. 4 and Fig. 5, it can be stated that the reliability and performance of the proposed modified vector based Dubins path approach is proper and effective – both inequality constraints can ensure a feasible re-planned path, however, there is an additional condition. When the moving in the straight line regime time, due to the AGV velocity, is shorter than time, necessary to reach the predefined steering wheels angle, which is determined by using Eq. (18), the condition to ensure re-planned path feasibility is not satisfied. Seeking to clearly describe the reliability and possible use of the proposed inequality constraints, it must be noted that both inequality constraints are suitable for path re-planning in unstructured roads, i.e., in cases where the turning radius is not constrained by external factors, like obstacles, road lines, etc. For example, while investigating the performance of various autonomous ground vehicles controllers, etc. Respectively, the proposed inequality constraints are not suited for a path, which is based on a real road network, re-planning. That is because the inequality constraints are built upon the length of the shorter straight line segment. To further improve the proposed methodology, with a view to ensure its use for re-planning of a path based on a real road network, the possibility to use both inequality constraints, which would be based not on the length of the shorter straight line segment but on road network limitations, must be considered and, respectively, developed.

6. Conclusion

In this work the development and reliability estimation of a methodology for re-planning of sharp turns in a path, while focusing on the constraints and the mobility of the vehicle, is presented. The conclusions obtained from the work are as follows:

1. Vector based Dubins path approach has been modified by applying a cost function with two different inequality constraints, this way developing a methodology for re-planning of sharp turns in a path, predefined by an on-board planner or a human. Because the non-modified Dubins path approach is based on the assumption that the AGV has a minimum turning radius, the developed methodology allows to solve the optimisation problem and to find an optimal and reliable turning radius, which would ensure path feasibility, while taking into consideration the primary path with sharp turns, the AGV velocity, the angular velocity of the the steering wheel turning, the steering ratio and the wheelbase of the AGV.
2. To ensure quantitative reliability analysis and estimation of the developed methodology, experimental drives were performed. It was observed that, during the experimental drives, the real time communication condition was satisfied, while the developed methodology was used for path re-planning. Thus, it can be stated that the developed methodology is effective and does not require many computational resources.
3. During all the experimental drives in the first inequality constraint case, and, respectively, during all the experimental drives in the second inequality constraint case, actual deviations or other undesired errors did not occur during the path following. However, during the result analysis it was estimated that the reliability and performance of the developed methodology, while using different proposed inequality constraints, were not similar. When the first inequality constraint

was used for path re-planning, the reliability and performance of the developed methodology was proper and effective in all the cases of the experimental drives. Thus, it can be stated that the developed methodology with the first inequality constraint can reliably ensure a feasible re-planned path to follow. However, it was found out that the developed methodology with the second inequality constraint can reliably ensure a feasible re-planned path to follow only in the cases, when the moving in the straight line regime time, due to the AGV velocity, is longer than time, necessary to reach the predefined steering wheels angle. Such inaccuracy can lead to undesired deviations from the path and other negative effects, when the distance between the waypoints is respectively large.

In conclusion it can be stated that the developed methodology is reliable and effective, however it is important to improve the proposed methodology by considering the modification of both inequality constraints by basing them not on the length of the shorter straight line segment but on road network geometry. By realizing such an improvement, the developed methodology would not only be useful while seeking to ensure the path feasibility for the performance evaluation of various autonomous ground vehicles controllers, but also while developing high performance path planning algorithms for a safe, reliable and successful integration of the autonomous ground vehicles into the transport system.

References

1. Alcala E, Puig, V, Quevedo, J, Escobet, T, Comasolivas R. Autonomous Vehicle Control Using a Kinematic Lyapunov-Based Technique with LQR-LMI Tuning. Control Engineering Practice 2018; 73: 1–12, https://doi.org/10.1016/j.conengprac.2017.12.004
2. Alcala E, Sellart L, Puig V, Quevedo J, Saludes J, Vazquez D, Lopez, A. Comparison of Two Non-Linear Model-Based Control Strategies for Autonomous Vehicles. 24th Mediterranean Conference on Control and Automation 2016: 846–851, https://doi.org/10.1109/MED.2016.7535921
3. Arokiasami W A, Vadakkepat P, Tan K C, Srinivasan D. Vector Directed Path Generation and Tracking for Autonomous Unmanned Aerial / Ground Vehicles. IEEE Congress on Evolutionary Computation (CEC) 2016: 1375–1381, https://doi.org/10.1109/CEC.2016.7743949
4. Bascetta L, Cucci D A, Mattuecci M. Kinematic trajectory Tracking Controller for an All-Terrain Ackermann steering Vehicle. IFAC-PapersOnLine 2016; 49(15): 13–18, https://doi.org/10.1016/j.ifacol.2016.07.600
5. Bayar G, Bergerman M, Konukseven E I, Koku A B. Improving the Trajectory Tracking Performance of Autonomous Orchard Vehicles Using Wheel Slip Compensation. Biosystems Engineering 2016; 146: 149–164, https://doi.org/10.1016/j.biosystemseng.2015.12.019
6. Brombacher A. (Re)liability of Self-driving Cars. An Interesting Challenge!. Quality and Reliability Engineering International 2014; 30(5): 613–614, http://dx.doi.org/10.1002/qre.1707
7. Brown M, Funke J, Erlien S, Gerdes J C. Safe Driving Envelopes for Path Tracking in Autonomous Vehicles. Control Engineering Practice 2017; 61: 307–316, https://doi.org/10.1016/j.conengprac.2016.04.013
8. Bullough J D, Skinner N P. High visibility reflective sign sheeting materials: field and computational evaluations of visual performance. Transport 2018; 33(2): 344–352, http://dx.doi.org/10.3846/16484142.2016.1230889
9. Castillo O, Trujillo L, Melin P. Multiple Objective Genetic Algorithms for Path-planning Optimization in Autonomous Mobile Robots. Soft Computing 2007; 11(3): 269–279, https://doi.org/10.1007/s00500-006-0068-4
10. Coombes M, Eaton W, MacAree O, Chen W H. Development of a Generic Network Enabled Autonomous Vehicle System. UKACC International Conference on Control 2014: 621–627, https://doi.org/10.1109/CONTROL.2014.6915211
11. Davoodi M, Panahi F, Mohades A, Hashemi S N. Clear and Smooth Path Planning. Applied Soft Computing 2015; 32: 568–579, https://doi.org/10.1016/j.asoc.2015.04.017
12. Delingette H, Hebert M, Ikeuchi K. Trajectory Generation with Curvature Constraint Based on Energy Minimization. Proceedings of IEEE / RSJ International Workshop on Intelligent Robots and Systems 1991: 6 p., https://doi.org/10.1109/IROS.1991.174451
13. Fu Y, Li H, Kaye M. Design and Lyapunov Stability Analysis of a Fuzzy Logic Controller for Autonomous Road Following. Mathematical Problems in Engineering 2010: 20 p., https://doi.org/10.1155/2010/578406
14. Gottlieb Y, Manathara J G, Shima T. Multi-Target Motion Planning Amidst Obstacles for Autonomous Aerial and Ground Vehicles. Journal of Intelligent & Robotic Systems 2017: 1–22, https://doi.org/10.1007/s10846-017-0684-5
15. Gu T, Dolan J M. On-Road Motion Planning for Autonomous Vehicles. Intelligent Robotics and Applications 2012: 588–597, https://doi.org/10.1007/978-3-642-33503-7_57
16. Gupta N, Ordonez C, Collins E G. Dynamically Feasible, Energy Efficient Motion Planning for Skid-Steered Vehicles. Autonomous Robots 2017; 41(2): 453–471, https://doi.org/10.1007/s10514-016-9550-8
17. Han J, Seo Y. Mobile Robot Path Planning with Surrounding Point Set and Path Improvement. Applied Soft Computing 2017; 7: 35–47, https://doi.org/10.1016/j.asoc.2017.03.035
18. Kala R. On-Road Intelligent Vehicles. Motion Planning for Intelligent Transportation Systems. Kidlington: Elsevier, 2016.
19. Kikutis R, Stankūnas J, Rudinskas D, Masiulionis T. Adaptation of Dubins Paths for UAV Ground Obstacle Avoidance When Using a Low Cost On-Board GNSS Sensor. Sensors 2017; 17(10): 23 p., https://doi.org/10.3390/s17102223
20. Krishan J, Rajeev U P, Jayabalan J, Sheela D S. Optimal Motion Planning Based On Path Length Minimisation, Robotics and Autonomous Systems 2017; 94: 245–263, https://doi.org/10.1016/j.robot.2017.04.014
21. LaValle S M. Planning Algorithms. Cambridge University Press: 2006.
22. Li X, Sun Z, Cao D, Liu D, He H. Development of a new integrated local trajectory planning and tracking control framework for autonomous ground vehicles. Mechanical Systems and Signal Processing 2017; 87: 118–137.
23. Liang T C, Liu J S. A Bounded-Curvature Shortest Path Generation Method for Cal-Like Mobile Robot Suing Cubic Spiral. Proceedings of 2004 IEEE / RSJ International Conference on Intelligent Robot and Systems 2004: 2819–2824, https://doi.org/10.1109/IROS.2004.1389836
24. Liang Y, Xu L. Global Path Planning for Mobile Robot Based Genetic Algorithm and Modified Simulated Annealing Algorithm. Proceedings of the FirstACM/SIGEVO Summit on Genetic and Evolutionary Computation 2009: 303–308, https://doi.org/10.1145/1543834.1543875
25. Macharet D G, Neto A A, Neto V F, Campos M F M. Nonholonomic Path Planning Optimization for Dubins' Vehicles. IEEE International Conference on Robotics and Automation 2011: 4208–4213, https://doi.org/10.1109/ICRA.2011.5980239

26. Macharet D G, Neto A A, Neto V F, Campos M F M. Dynamic Region Visit Routing for Vehicles with Minimum Turning Radius. Journal of Heuristics 2017; 24(1): 83–109, https://doi.org/10.1007/s10732-017-9359-4
27. Medeiros A C, Urrutia S. Discrete Optimization Methods to Determine Trajectories for Dubins' Vehicles. Electronic Notes in Discrete Mathematics 2010; 36: 17–24. https://doi.org/10.1016/j.endm.2010.05.003
28. Millo F, Cubito C, Rolando L, Pautasso E, Servetto E. Design and development of an hybrid light commercial vehicle. Energy 2017; 136: 90–99, http://dx.doi.org/10.1016/j.energy.2016.04.084
29. Młynarski S, Pilch R, Smolnik M, Szybka J. Methodology of network systems reliability assessment on the example of urban transport. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20(2): 278–283, http://dx.doi.org/10.17531/ein.2018.2.14.
30. Mousavi M, Pakniyat A, Wamg T, Boulet B. Seamless dual brake transmission for electric vehicles: design, control and experiment. Mechanism and Machine Theory 2015; 94: 96–118, https://doi.org/10.1016/j.mechmachtheory.2015.08.003
31. Pielecha I, Cieślik W, Szałek A. Operation of electric hybrid drive systems in varied driving conditions. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (1): 16–23, http://dx.doi.org/10.17531/ein.2018.1.3.
32. Pravilonis T, Sokolovskij E, Surblys V. The analysis of deformations of the frame in a basic minibus and a low floor minibus. Mechanics 2017; 23(3): 365–369, http://dx.doi.org/10.5755/j01.mech.23.3.14480
33. Receveur J B, Victor S, Melchior P. Multi-Criteria Trajectory Optimization for Autonomous Vehicle, IFAC Papers Online 2017; 50(1): 12520–12525. https://doi.org/10.1016/j.ifacol.2017.08.2063
34. Renzi C, Leali, F, Di Angelo L. A review on decision-making methods in engineering design for the automotive industry. Journal of Engineering Design 2017; 28(2): 118–143, http://doi.org/10.1080/09544828.2016.1274720
35. Reuter J. Mobile Robots Trajectories with Continuously Differentiable Curvature: An Optimal Control Approach. Proceedings of the IEEF / RSJ Intl. Conference on Intelligent Robots and Systems 1998: 38–43. https://doi.org/10.1109/IROS.1998.724593
36. Surblys V, Ślaski G, Pikosz H. The usage of a laser height sensors for estimating road unevenness profile. The archives of automotive engineering = Archiwum motoryzacji. Warszawa: Wydawnictwo Naukowe Przemysłowego Instytutu Motoryzacji 2018; 79(1): 95–106, http://dx.doi.org/10.14669/AM.VOL.79.ART7
37. Žuraulis, V.; Sokolovskij, E.; Matijošius, J. The opportunities for establishing the critical speed of the vehicle on research in its lateral dynamics. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2013; 15(4): 312–318.

Paulius SKAČKAUSKAS
Edgar SOKOLOVSKIJ
Department of Automobile Engineering
Vilnius Gediminas Technical University
J. Basanavičiaus 28, LT-03224 Vilnius, Lithuania

E-mail: paulius.skackauskas@vgtu.lt, edgar.sokolovskij@vgtu.lt

Article citation info:
DUAN R, LIN Y, ZENG Y. Fault diagnosis for complex systems based on reliability analysis and sensors data considering epistemic uncertainty. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 558–566, http://dx.doi.org/10.17531/ein.2018.4.7.

Rongxing DUAN
Yanni LIN
Yining ZENG

FAULT DIAGNOSIS FOR COMPLEX SYSTEMS BASED ON RELIABILITY ANALYSIS AND SENSORS DATA CONSIDERING EPISTEMIC UNCERTAINTY

DIAGNOZOWANIE BŁĘDÓW W SYSTEMACH ZŁOŻONYCH NA PODSTAWIE ANALIZY NIEZAWODNOŚCI ORAZ DANYCH Z CZUJNIKÓW Z UWZGLĘDNIENIEM NIEPEWNOŚCI EPISTEMICZNEJ

This paper presents an information fusion method to diagnose system fault based on dynamic fault tree (DFT) analysis and dynamic evidential network (DEN). In the proposed method, firstly, it uses a DFT to describe the dynamic fault characteristics and evaluates the failure rate of components using interval numbers to deal with the epistemic uncertainty. Secondly, qualitative analysis of a DFT is to generate the characteristic function via a traditional zero-suppressed binary decision diagram, while quantitative analysis is to calculate some importance measures by mapping a DFT into a DEN. Thirdly, these reliability results are updated according to sensors data and used to design a novel diagnostic algorithm to optimize system diagnosis. Furthermore, a diagnostic decision tree (DDT) is obtained to guide the maintenance workers to recover the system. Finally, the performance of the proposed method is evaluated by applying it to a train-ground wireless communication system. The results of simulation analysis show the feasibility and effectiveness of this methodology.

Keywords: *dynamic fault tree, dynamic evidential network, interval numbers, sensors data, diagnostic importance factor.*

W artykule przedstawiono metodę fuzji informacji służącą do diagnozowania błędów systemu w oparciu o analizę dynamicznego drzewa błędów (DFT) oraz dynamiczną sieć dowodową (DEN). W proponowanej metodzie, pierwszym krokiem jest wykorzystanie DFT do opisania dynamicznych charakterystyk błędów oraz ocena intensywności uszkodzeń komponentów przy użyciu liczb przedziałowych, która rozwiązuje problem niepewności epistemicznej. Krok drugi stanowi jakościowa analiza DFT, która polega na wygenerowaniu funkcji charakterystycznej za pomocą tradycyjnego binarnego diagramu decyzyjnego typu "zero-suppressed" (w którym zostały wyeliminowane wszystkie węzły, których krawędź „1" prowadzi do liścia „0"), oraz analiza ilościowa polegająca na obliczeniu pewnych miar ważności poprzez odwzorowanie DFT w DEN. W kroku trzecim, otrzymane wyniki niezawodnościowe aktualizuje się zgodnie z danymi z czujników a następnie wykorzystuje do stworzenia nowego algorytmu diagnostycznego do optymalizacji diagnostyki systemu. Powstaje diagnostyczne drzewo decyzyjne (DDT), które stanowi dla pracowników utrzymania ruchu wytyczną w procesie odzyskiwania systemu. Działanie proponowanej metody oceniano poprzez zastosowanie jej do diagnostyki systemu łączności radiowej pociąg–ziemia. Wyniki analizy symulacyjnej wskazują na możliwość praktycznego wykorzystania i skuteczność omawianej metodologii.

Słowa kluczowe: *dynamiczne drzewo błędów, dynamiczna sieć dowodowa, liczby przedziałowe, dane z czujników, czynnik ważności diagnostycznej.*

1. Introduction

With the rapid development of science and technology, application of high dependability safeguard techniques have improved the performance of modern systems greatly on the one hand, but increased the complexity of these systems on the other hand, which significantly raises some challenges in fault diagnosis. These challenges are failure dependency of components and epistemic uncertainty. Usually, some methods of fault tolerance are used to improve the system reliability. The behaviours of components in this system, such as failure priority, functional dependent failures, and sequentially dependent failures should be taken into account. In addition, high reliability makes it extremely difficult to obtain complete fault data because these systems may still be in the early life cycle, which results in the epistemic uncertainty. Thus, the work of fault diagnosis has attracted more attention than before. The aim of a fault diagnosis system is to quickly detect and identify the root causes of these failures based on some information such as sensors data and operator experience by using some models and algorithms. Several efficient fault diagnosis approaches have been proposed for a variety of systems over the last few decades. Doguc et al. proposed a new fault diagnosis method based on the real-time reliability analysis [7]. Bayesian network (BN) was used to calculate the system reliability, and the real-time system reliability was monitored and compared with the previous values. If the deviations exceeded the set threshold, a heuristic efficient algorithm was used to locate the failed component which had the greatest changes between the prior probability and posterior probability. In the literature [3], a real-time fault diagnosis method for complex systems using object-oriented BN was proposed. It included an off-line BN construction phase and an on-line fault diagnosis phase. However, the construction of BN model requires a large amount of fault data. In [5], a fault diagnosis approach based on the fuzzy neural network and fault tree was proposed. Fuzzy neural network was used to train the relationship between faults and symptoms. Fault tree was used to describe the logical relationship between faults and symptoms. In [13], a new method was proposed to diagnose the bearing fault using evidence network and support vector machine. The fault model construction was established using a data-driven method, and the evidence theory was used to solve the conflicting results from different layer models to increase the diagnosis accuracy. However, the above methods are based on the data-driven fault method which needed lots of fault data and cannot deal with the epistemic uncertainty. A fault diagnosis method for safety instrumentation system based on the fault tree and BN was proposed [6]. It used the static fault tree to construct the fault model of safety instrument system and mapped the fault tree into BN to calculate the importance measure which was used to design the diagnosis algorithm. Nevertheless, this method is unable to describe the dynamic fault characteristics and fails to deal with the epistemic uncertainty. In work of [1], DFT was introduced to model the dynamic fault behaviours and diagnostic importance factor (DIF) was calculated to determine the diagnostic sequence. However, this method determined the diagnosis sequence only by components' DIF, and usually caused minimal cut sets ((MCS)) with a smaller DIF to be checked first, thereby influencing the diagnosis result. Tao et al. presented an improved fault diagnosis method which took components' DIF and MCS's DIF into account to avoid that case [23]. In order to improve the diagnosis efficiency, Assaf et al. proposed a method to incorporate the evidence information from sensors into the diagnostic process based on the DFT [2]. However, the solution for DFT is based on Markov chains, which is ineffective in handing large DFT and modelling power capabilities. Furthermore, it cannot update the reliability results according to the evidence data from sensors, which affects the diagnostic efficiency. Therefore, Duan et al. presented an efficient diagnostic algorithm which used DFT to establish a system failure model and calculated reliability parameters using a discrete time Bayesian network (DTBN) [8]. This approach not only can avoid the state space explosion, but also can incorporate sensor information to update reliability results. Nevertheless, DTBN is an approximate method to solve DFT and there is a contradiction between the accuracy and computational complexity. Furthermore, these diagnosis methods are usually assumed that the failure rates of the components are expressed in crisp values describing their reliability characteristics and cannot cope with the epistemic uncertainty. So, a fuzzy DFT analysis was introduced, which can deal with the uncertainty and model the dynamic fault characteristics [12, 17]. Nevertheless, the solution for the fuzzy DFT was still based on the Markov chains. To overcome these shortcomings, a new fault diagnosis algorithm based on fuzzy set and DFT analysis was proposed [10]. The fuzzy information obtained by fuzzy set theory and domain expert was transformed into quantitative information to obtain the fuzzy failure rates of components. DTBN was used for quantitative analysis. Nevertheless, it is usually difficult to determine the corresponding membership function of each language value. To this end, Duan et al. proposed a new fault diagnosis for complex systems based on dynamic evidential network and multi-attribute decision making [11]. It used interval numbers to express the failure rates of the basic events and obtained the optimal diagnosis sequences based multi-attribute decision making with interval numbers. However, this method failed to incorporate the sensors data to optimize the diagnosis process.

In summary, fault diagnosis methods based on reliability analysis have some following limitations:

(1) Traditional fault diagnosis methods based on reliability analysis generally use a static fault tree or DFT to construct fault model and assume that the failure rates of all events are crisp values, which cannot deal with epistemic uncertainty. Although some researchers put forward the possibility theory [21, 25], fuzzy set theory [4, 15], imprecise probability [18], interval analysis [27] and evidence theory [28], these theories were only used for the reliability analysis and risk assessment and were not further applied to the fault diagnosis. Furthermore, Markov chains and DTBN are usually used to solve DFT. Markov chains have a bad state space explosion problem and the inability to update the posterior probability of the component based on sensors data. The DTBN based solution for DFT has the contradiction between computational accuracy and computational complexity. That is, its computational accuracy is related to the size of time granularity n. As n increases, the conditional probability table has an exponential growth [26]. Although the solution proposed in [16] can solve the problem of calculation accuracy to a certain extent, it cannot fuse the sensors information for backward reasoning.

(2) From the aspect of sensors information fusion, Traditional method appends a sensor layer for capturing evidence onto the DFT without impacting the reliability analysis, and the sensor layer uses static gates to represent evidence information. However, evidence information is only used to update qualitative information to reduce the number of suspected MCS and fails to update the quantitative information, thus unable to reflect the contribution of components to the system failure.

(3) In the view of the diagnosis algorithm, the algorithms based on reliability analysis generally only take the importance measures or posterior probability of components into account [1, 9]. Furthermore, the importance measures are usually crisp values and cannot be used to make decisions under uncertainty.

Motivated by the problems motioned above, this paper proposes an information fusion method to diagnose system fault based on DFT and DEN. DFT is used to establish the system fault model to describe the dynamic fault characteristics. Interval numbers are used to describe the failure rate of components to deal with epistemic uncertainty. Furthermore, an efficient zero-suppressed binary decisions diagrams is used to obtain all MCSs, and a DFT is mapped into a DEN to calculate the reliability parameters. In addition, evidence information from sensors is incorporated to update the qualitative information and quantitative parameters, which are used to design the fault diagnosis algorithm. Finally, a train-ground wireless communication system is given to demonstrate the efficiency of this proposed method.

The remainder of this article is organized as follows. Section 2 presents the model construction and qualitative analysis of DFT. Section 3 introduces the dynamic evidence network and provides a quantitative analysis method by mapping a DFT to a DEN. A novel approach is proposed to incorporate the evidence information to update the reliability results, and an efficient diagnosis algorithm is given in Section 4. Section 5 is devoted to a simple illustration example of the proposed approach. Some conclusions and future research recommendations are given in the final section.

2. DFT

2.1. Model Construction of DFT

Fault tree is a deductive method to decide the potential causes that may cause the occurrence of a predefined undesired event, generally denoted as the top event. DFT extends a static fault tree to describe the dynamic failure behaviours such as priorities of failure events, spares, and sequence-dependent events. Dynamic gates in DFT include the priority AND gate (PAND), the functional dependency gate (FDEP), the sequence enforcing gate (SEQ), the cold, hot, and warm spare gates (CSP, HSP, WSP). The model construction of the fault tree usually requires an in depth knowledge of the system and its components. It includes the construction of a network topology and the failure rates estimation of components. The former can resort to fault mode and effect analysis and the latter needs to obtain lots of fault data, which is almost impossible to estimate precisely the failure rates of the basic events in the practical engineering application. In this paper, interval numbers are used to describe the failure rates of the basic events based on the expert elicitation and some data sheet at the design stage.

2.2. Qualitative analysis of a DFT

The qualitative analysis of a fault tree can be used to obtain the MCS. Algebraic simplification is the most effective method to solve MCS, but it is not suitable for solving DFT. Zero-suppressed binary decisions diagrams, introduced by Tang, separate timing constraints and logic constraints and convert a DFT into a static fault tree [24]. This algorithm generates the MCS of the corresponding static fault tree using several set operations and then it can be expanded into minimal cut sequences if we consider the timing constraints.

Let S_1, S_2 be the input of MCS for AND gate and MCS for OR gate respectively, several set operations are as follows:

$$\begin{aligned} &S_c = S_1 \cap S_2, D_1 = S_1 - S_c, D_2 = S_2 - S_c \\ &U = D_1 \cup D_2, P = D_1 * D_2, D_3 = U - P \\ &MCS_{OR} = S_c \cup D_3, MCS_{AND} = S_c \cup P \end{aligned} \tag{1}$$

where D, S_c, U, and P respectively represent set difference, set intersection, set union, and set product. MCSOR and MCSAND are the output of MCS- OR and MCS- AND respectively.

The MCS generation algorithm is implemented recursively during the depth-first left-most traversal of a fault tree. Firstly, it generates the MCS of the inputs of a connection gate, and then executes some operations to combine the MCS of the inputs into the MCS of the output of the connection gate. Finally, all the minimal cut sequences from the MCS can be obtained by considering the timing constraints [24].

2.3. Quantitative analysis of a DFT

Quantitative analysis of a DFT is mainly to calculate the system reliability and some importance measures. DIF is the most frequently used importance measure and is also the cornerstone of diagnosis method based on reliability. From a diagnostic point of view, it allows us to discriminate between components by their importance. It is well known to us all that components with a larger DIF value should be diagnosed first. It can assure a minimal number of system checks while bringing back the system. Reliability parameters are calculated by converting a DFT into a DEN which is introduced in Section 3.

3. DEN

3.1. EN

D-S evidence theory has a unique ability in the expression of epistemic uncertainties. The evidence theory can be well compatible with the theory of probability. EN consists of BN and D-S evidence theory and includes both advantages [14]. It is a popular analysis tool for representing and managing epistemic uncertainties. An EN is a directed acyclic graph (DAG) used to represent system's uncertain knowledge and system logic in artificial intelligence. An EN is defined as $EN = <G, P>$, where $G = <N, A>$ represents a network graph and $N = \{N_1, N_2, \cdots, N_k\}$ represents a set of nodes. A node can be a basic variable or an abstraction of a system or component, such as system reliability, component status. A is a set of arcs, which indicate direct conditional relations between the connected nodes. P represents some network parameters in EN. Each network parameter represents the belief distributions that are distributed to a node, and each node $X_i \in N$ has a corresponding conditional belief table. The parent node of node X_i is set to $Pa(X_i)$ and their relationship is expressed in the formula $P(X_i \mid Pa(X_i))$.

3.2. DEN

A DEN extends an EN with adding a temporal dimension. This new dimension is managed by defining different nodes to model variables with respect to different time slices. A DEN includes an initial network and some temporal transition networks. Each time slice corresponds to a static EN, and the time slices are made up of a directed acyclic graph $G_T = <V_T, E_T>$ and the corresponding conditional probabilities. The V_T and E_T are respectively nodes of time T and directed arcs. A directed arc links two variables belonging to different time slices and E_T^{tmp} is used to denote the temporal transition network of time slices. Then E_T^{tmp} can be determined by:

$$E_T^{tmp} = \{(a,b) \mid a \in V_{T-1}, b \in V_T\}, T_0 \le T \le T_0 + N\Delta T \tag{2}$$

where T_0 is an initial network.

In the DEN model, G_T depends solely upon the present state and the previous state. Thus, the following equation is obtained:

$$P(G_T \mid G_{T-\Delta T}, ..., G_{T_0}) = P(G_T \mid G_{T-\Delta T}) \tag{3}$$

In addition, we define these impacts as transition-belief masses between the focal elements of the variable at time step k and those at time step k+1 and the CBT relative to inter-time slices is calculated by Equation 4:

$$m(X_{k+1} \mid X_k) = \begin{bmatrix} m(G_1^{X_{K+1}} \mid G_1^{X_K}) & \cdots & m(G_Q^{X_{K+1}} \mid G_1^{X_K}) \\ \vdots & \ddots & \vdots \\ m(G_1^{X_{K+1}} \mid G_Q^{X_K}) & \cdots & m(G_Q^{X_{K+1}} \mid G_Q^{X_K}) \end{bmatrix} \tag{4}$$

3.3. System reliability model of DEN

In evidence theory, $\Theta = \{W_i, F_i\}$ is the knowledge framework of the component i and the focal elements are defined by:

$$2^{\Theta} = \{\{\varnothing\}, \{W_i\}, \{F_i\}, \{W_i, F_i\}\} \tag{5}$$

where $\{W_i\}$ and $\{F_i\}$ denote the working state and failure state respectively. The state of $\{W_i, F_i\}$ corresponds to the epistemic uncertainty.

Belief measure *(Bel)* defines the lower bound of the probabilities that the focal element exists, and plausibility measure *(Pl)* defines the upper bound of the probabilities that the focal element exists. The basic belief assignment on the system state expresses an epistemic uncertainty, where *Bel* and *Pl* measures are not equal and bound the system reliability. Therefore, the basic probability assignment (BPA) of component i can be computed as:

$$\begin{cases} m(\{W_i\}) = Bel(\{W_i\}) \\ m(\{F_i\}) = 1 - Pl(\{W_i\}) \\ m(\{W_i, F_i\}) = Pl(\{W_i\}) - Bel(\{F_i\}) \end{cases} \tag{6}$$

Presumably, the upper and lower bounds of the component's failure probability is equivalent to the BPA in the DEN:

$$\begin{cases} m(\{W_i\}) = 1 - \overline{P(x)} \\ m(\{F_i\}) = \underline{P(x)} \\ m(\{W_i, F_i\}) = \overline{P(x)} - \underline{P(x)} \end{cases} \quad (7)$$

where $Bel(\{F_i\}) = \underline{P(x)}$ and $Pl(\{F_i\}) = \overline{P(x)}$.

3.4. DFT analysis based on DEN

3.4.1. Converting a static logic gate into a DEN

Static logic gates mainly include three gates, AND gate, OR gate and voting gate. This section takes an OR gate for example and provides the schemes to map an OR gate into a DEN. When any of the input components X_i (i=1,…, n) of an OR gate fails, the output of the gate fails too. Fig. 1 shows an OR gate and the equivalent DEN. Table 1 gives the conditional probabilities of node *A (T+ΔT)* in the DEN. Equation 8 gives the conditional probabilities of output node *E (T+ΔT)*. A more detailed description of this work can be found in [20].

Table 1. The conditional probabilities of node A (T+ΔT).

A(*T*)	*A*(*T*+*ΔT*)		
	{*W*}	{*F*}	{*W,F*}
{*W*}	m_A(W)	m_A(F)	m_A(W,F)
{*F*}	0	1	0
{*W,F*}	0	m_A(F)	1- $m_A(F)$

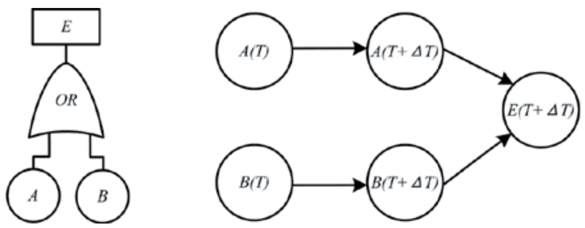


Fig. 1. An OR gate and the equivalent DEN

$$\begin{cases} P(E=1 \mid A(T+\Delta T)=0, B(T+\Delta T)=1)=1 \\ P(E=1 \mid A(T+\Delta T)=1, B(T+\Delta T)=0)=1 \\ P(E=1 \mid A(T+\Delta T)=1, B(T+\Delta T)=1)=1 \\ P(E=1 \mid A(T+\Delta T)=1, B(T+\Delta T)=\{0,1\})=1 \\ P(E=1 \mid A(T+\Delta T)=\{0,1\}, B(T+\Delta T)=1)=1 \\ P(E=\{0,1\} \mid A(T+\Delta T)=0, B(T+\Delta T)=\{0,1\})=1 \\ P(E=\{0,1\} \mid A(T+\Delta T)=\{0,1\}, B(T+\Delta T)=0)=1 \\ P(E=\{0,1\} \mid A(T+\Delta T)=\{0,1\}, B(T+\Delta T)=\{0,1\})=1 \\ P(E=1 \mid A(T+\Delta T)=0, B(T+\Delta T)=0)=0 \end{cases} \quad (8)$$

3.4.2. Converting a dynamic logic gate into a DEN

Some dynamic logic gates are introduced to model the functional and sequential in the DFT. These logic gates include PAND, SEQ, FDEP and spare gates. An FDEP gate will be used to describe how the dynamic logic gates are mapped into DEN. An FDEP gate includes a trigger event and some dependent basic events. The trigger event can be a basic event or an output of another gate in the DFT. The occurrence of a trigger event will force all basic events to occur, which means all basic events functionally depend upon the trigger event. Fig. 2 shows an FDEP gate and the equivalent DEN. Table 2 and Table 3 show the conditional probabilities of the node *A*(*T*+Δ*T*) and *E*(*T*+Δ*T*) respectively.

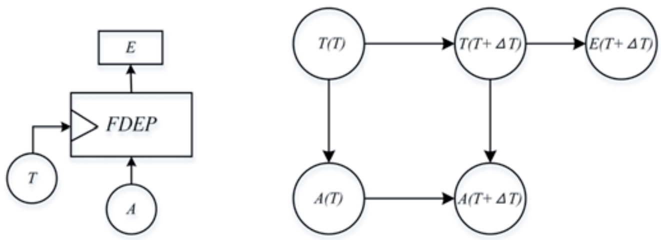


Fig. 2. An FDEP gate and the equivalent DEN

Table 2. The conditional probabilities of the node A (T+ΔT)

T(*T*+*ΔT*)	*A*(*T*)	*A*(*T*+*ΔT*)		
		{*W*}	{*F*}	{*W,F*}
{W}	{W}	m_A(W)	m_A(F)	m_A(W,F)
{*W*}	{*F*}	0	1	0
{*W*}	{*W,F*}	0	0	1
{*F*}	{*W*}	0	1	0
{*F*}	{*F*}	0	1	0
{*F*}	{*W,F*}	0	1	0
{*W,F*}	{*W*}	0	0	1
{*W,F*}	{F}	0	1	0
{*W,F*}	{*W,F*}	0	0	1

Table 3. The conditional probabilities of the node E(T+ΔT).

T(*T*+*ΔT*)	*E*(*T*+*ΔT*)		
	{*W*}	{*F*}	{*W,F*}
{*W*}	1	0	0
{*F*}	0	1	0
{*W,F*}	0	0	1

3.4.3. Calculating reliability results

After DFT model of a system is built, it can be mapped into the equivalent DEN using the approach mentioned above. Reliability results of system can be obtained by resorting to the DEN inference algorithm. Reliability parameters mainly include system unreliability and DIF, which can be used to develop a diagnosis algorithm.

The unreliability of a system is calculated by the following equation:

$$P_S = [\underline{P_S}, \overline{P_S}] = [Bel(\{F_S\}) \quad Pl(\{F_S\})] \quad (9)$$

where $[Bel(\{F_S\}), Pl(\{F_S\})]$ represents the failure probability of a system.

DIF is usually defined as the probability that a basic event has occurred given that the top event has also occurred. The DIF of a component *i* is given by:

$$DIF_i = P(i \mid S) = [Bel(\{F_{i|S}\}), Pl(\{F_{i|S}\})] \quad (10)$$

where i is a component in the system S; $P(i \mid S)$ is the probability that the basic event i has occurred given the top event has occurred.

Similarly, the DIF of a MCS n is defined by:

$$DIF_{MCS_n} = P(MCS_n|S) = \frac{P(MCS_n)}{P(S)} \tag{11}$$

where $P(S)$ is the unreliability of the system S; $P(MCS_n|S)$ is the failure probability that the MCS n has occurred given the top event has occurred.

For convenience, we calculate the value $P(MCS_n)$ instead of DIF_{MCS_n} and use it to design the diagnosis algorithm in the following section.

3.5. Importance sorting using possibility-based NSG ranking approach

Based on above analysis, we can obtain the interval value of DIF which can be used to develop an efficient diagnosis algorithm in order to reduce the diagnosis cost. As is known to all, components with a larger DIF are more important from a diagnostic point of view. Thus, the importance ranking of components will be very important for determining a diagnosis sequence. Nevertheless, these interval values are not sufficient to rank components and should be converted into a probability measure. In this paper, a possibility-based NSG ranking method, developed by Nakahara et al. is used to rank DIF of components expressed by interval numbers [19, 22]. This method can be used to compare the DIF of components to provide a guidance for system diagnosis.

For interval numbers $a=[a^-,a^+]$ and $b=[b^-,b^+]$, l(a) and $l(b)$ respectively denote the lengths of the intervals $a=[a^-,a^+]$ and $b=[b^-,b^+]$, it calculated as follows:

$$l(a)=a^+-a^-, l(b)=b^+-b^- \tag{12}$$

Then the possibility of $[a]\ge[b]$ can be defined as:

$$\begin{aligned} p([a]\ge[b]) &= \min\{0,1-\max(\frac{a^+-b^-}{l(a)+l(b)},0)\} \\ &= \begin{cases} 1 & a^- \ge b^+ \\ \dfrac{a^+-b^-}{l(a)+l(b)} & a^+ > b^- and\ a^- < b^+ \\ 0 & a^+ \le b^- \end{cases} \end{aligned} \tag{13}$$

A possibility-based NSG ranking method includes the following steps.

Step 1: For a set of interval numbers $a_i=[a_i^-,a_i^+]$ i=1,2,$\cdots$,n, compare them with each other, and then the corresponding possibility $p_{ij}=p(a>b)$ can be obtained. So we can establish the probability matrix $P=(p_{ij})_{n\times n}$, which is given by:

$$P=\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \tag{14}$$

Step 2: Denote $\lambda_i=\sum_{j=1}^{n} p_{ij}$ as the row sum of the possibility matrix P and $\lambda=(\lambda_1\ \lambda_2\ \cdots\ \lambda_n)^T$ as the corresponding row sum vector.

Step 3: Calculate the ranking vector ω=(ω_i) is given by:

$$\omega_i=\frac{1}{n(n-1)}(\lambda_i+\frac{n}{2}-1) \quad i=1,2\cdots n \tag{15}$$

According to the ranking vector of the possibility matrix P, the interval numbers $a_i=[a_i^-,a_i^+]$ i=1,2,$\cdots$,n can be sorted based on the value of ω_i.

4. Fault Diagnosis Method based on Reliability Analysis and Sensors data

4.1. Model construction of diagnostic sensors

When a system fails, usually several evidence information from sensors can be observed too, and this may be utilized to improve the efficiency of the diagnosis algorithm. In general, the more the number of sensors used to monitor the system, the higher the diagnostic efficiency of the system. However, too many sensors will increase system costs on the one hand, but on the other hand, it will reduce the reliability of the diagnostic system. So a tradeoff between the good points against the bad should be taken into account. Besides, sensors might fail and false information can misguide the diagnosis process. For simplicity, we assume that sensors never fail in the paper. To optimize the diagnosis process, a diagnostic sensors model is constructed to update the qualitative and quantitative information. As we all know, the DEN created from DFT has no evidence nodes representing the evidence information, thus, we need to add them in the DEN. Evidence nodes in the DEN provide links connecting it with the component in the DEN, which are monitored by sensors. The links are directed from the component to the evidence nodes. Evidence nodes in the DEN create a conditional probability table using the probability of producing the observation results. This diagnostic sensors model does not affect the system reliability analysis and can update the qualitative information and quantitative parameters according to sensors data.

4.2. Incorporating sensors data

4.2.1. Updating the system characteristic function

If sensors detect some failed components, we can use this evidence information to minimize the number of the diagnosed MCS. Since, examining a cut set that caused the system to fail then fixing the failed components in that cut set should recover the system, we can increase the efficiency of fault diagnosis by reducing the number of cut sets examined. The cut sets under evidence (CUE) is the set of all essential MCS obtained after evidence information removes some unsuspected cut sets. We can use evidence information from sensors to simplify the characteristic function of the system in order to obtain the CUE function using the algorithm in [2].

4.2.2. Updating DIF

In addition, we can use the evidence information from sensors to update DIF, which reflects objectively the contribution to the system failure. The DIF of the components under the evidence information conditions can be calculated using the Equation (16). Calculating DIF is very simple. We just input the corresponding evidence information to the DEN and obtain the DIF of components and CUE using the inference algorithm:

$$DIF_i' = P(i|S,E) = \frac{P(i,E,S)}{P(S)DIF_E} \tag{16}$$

where i, S and E represent a component, system and evidence information, respectively.

4.3. Fault diagnosis algorithm

The aim of fault diagnosis is to obtain the optimal check sequence to locate the fault as fast as possible using an efficient diagnosis algorithm. As it is known to all, the direct cause of the system failure is the failure of a CUE. So, we should check CUE one by one to locate the failed component in the system. Only when we finish checking a CUE can we do next. The sequence by which CUE is diagnosed depends on the corresponding DIF, while the sequence of components in the same CUE is determined by their DIF. The CUE with a larger DIF is checked first. Accordingly, the component with a larger DIF in a CUE is checked first. It can assure a minimal number of system checks while bringing the system back. The fault diagnosis algorithm, which incorporates sensors data, is as follows:

Step 1. List all CUEs and rank them according to their DIF.

Step 2. Select the CUE with a highest DIF value and diagnose the component X with a highest DIF in the same CUE.

Step 3. Split all CUEs into those with X and those without.

a) If X has failed test, we take all CUEs that include X
  - ➢ Diagnose all CUEs and the CUE with a higher DIF is checked first.
  - ➢ The component with a larger DIF in the same CUE is checked first.

b) If X has not failed test, we take the other CUEs
  - ➢ Select the CUE untested with a highest DIF value.
  - ➢ And recursively repeat Step2 - Step3.

4.4. Evaluation of diagnosis algorithm

The diagnosis algorithm can easily be described in the graphical DDT, which can help us recognize the failed components with a map. It is a directed acyclic graph composed of circular nodes and arcs linking parent nodes to child nodes. A node represents a component being tested. Arcs point to the next component to be tested; right arcs point to components within the same cutest as the parent node, and left arcs point to components which are not in the same cutest as the parent node. Moreover, when diagnostician reaches a node and tests the component at the node, the test either fails or passes. If the test fails, then the right arc is traversed indicating the need to repair the tested component in the parent node. If a test passes, then the left arc is traversed indicating that the cut sets which include the tested component in the parent node have not failed.

There are many indicators to evaluate the fault diagnosis algorithm. In this paper, we can evaluate the diagnostic efficiency with the help of the DDT. Traditional evaluation measures only take the test cost or the failure probability of components into account, and neglect the qualitative information and the importance factors. Thus, we use expected diagnostic cost (EDC) which incorporates the structure information, DIF and test cost into one measure for predicting diagnosis cost. This evaluation index takes the diagnosis accuracy as well as the diagnosis cost into account and also considers the relationship between component failure and system failure. Generally, the diagnostic cost is lower, the diagnostic approach is more efficient. EDC can be computed by:

$$EDC = \sum_{i=1}^{n} DIF_{CUE_i} cp_i, cp_i = \sum_{j=1}^{m_i} t_{c_j} \tag{17}$$

where DIF_{CUE_i} is the DIF of the i^{th} CUE; cp_i is the sum of all test cost from the top node to the i^{th} CUE's leaf node; t_{c_j} is the test cost of the node c_j.

5. A numerical example

Train-ground wireless communication system is a key subsystem of urban rail transit, and its reliability has been improved by the application of high technologies to ensure safe operation. Once breaking down, less causes the operation performance drop, more leads to a disaster. Therefore, an efficient diagnosis strategy should be taken to restore normal operation as soon as possible. A DFT model of a train-ground wireless communication system is shown in Fig.3. It is assumed that all components have the exponential distribution and interval failure rates of components expressed in interval values are shown in Table 4.

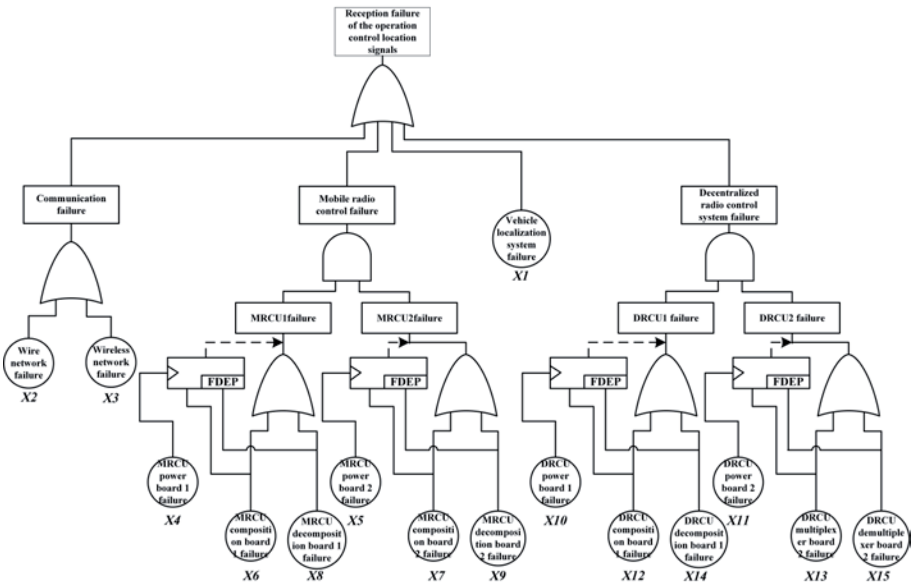


Fig. 3. DFT model of train-ground wireless communication system

Table 4. Failure rates of components are expressed in interval numbers

Components	Interval failure rates	Components	Interval failure rates
X1	[4.22e-6, 5.28e-6]	X8,X9	[5.49e-6, 6.71e-6]
X2	[5.94e-6, 7.26e-6]	X10,X11	[3.15e-5, 3.85e-5]
X3	[4.86e-5, 5.94e-5]	X12,X13	[6.12e-5, 7.48e-5]
X4,X5	[3.78e-5, 4.62e-5]	X14	[5.04e-5, 6.11e-5]
X6,X7	[6.48e-5, 7.92e-5]	X15	[5.04e-5, 6.11e-5]

Through the qualitative analysis of DFT mentioned above, the system characteristic function (the sum of all MCS) of train-ground wireless communication system is obtained:

$$F = X1 + X2 + X3 + X4X5 + X4X7 + X4X9 + X6X5 + X6X7 + X6X9 + \\ X8X5 + X8X7 + X8X9 + X10X11 + X10X13 + X10X15 + X12X11 + \\ X12X13 + X12X15 + X14X11 + X14X13 + X14X15$$

The DFT is mapped into a corresponding DEN for quantitative analysis. Assuming the task time T = 1000 h, the probability of system failure can be obtained using the inference algorithm and it is [0.08293, 0.10714]. In addition, the DIF of all components and MCSs can be calculated shown in Table 5 and Table 6 respectively.

Table 5. DIFs of all components

Components	DIF of components	Components	DIF of components
X1	[0.0508,0.0518]	X8,X9	[0.0857,0.0939]
X2	[0.0709,0.0722]	X10,X11	[0.0708,0.0756]
X3	[0.5681,0.5727]	X12,X13	[0.2012,0.2156]
X4,X5	[0.0751,0.0822]	X14	[0.1788,0.1914]
X6,X7	[0.1963,0.2148]	X15	[0.1788,0.1914]

Table 6. DIFs of all MCSs

MCSs	DIF of MCSs	MCSs	DIF of MCSs	MCSs	DIF of MCSs
X1	[0.0393,0.0635]	X6.X7	[0.0877,0.1675]	X10.X15	[0.0227,0.0432]
X2	[0.0553,0.0873]	X6.X9	[0.0384,0.0732]	X12.X11	[0.0255,0.0488]
X3	[0.4428,0.6954]	X8.X5	[0.0147,0.0281]	X12.X13	[0.0725,0.1383]
X4.X5	[0.0129,0.0246]	X8.X7	[0.0384,0.0732]	X12.X15	[0.0644,0.1224]
X4.X7	[0.0336,0.0642]	X8.X9	[0.0167,0.0321]	X14.X11	[0.0227,0.0432]
X4.X9	[0.0147,0.0281]	X10.X11	[0.0090,0.0172]	X14.X13	[0.0644,0.1224]
X6.X5	[0.0336,0.0642]	X10.X13	[0.0255,0.0488]	X14.X15	[0.0572,0.1084]

A possibility-based NSG sorting method is used to rank the DIF of components and the ranking vectors ω_i of matrices P can be computed as:

ω_i=(0.0333, 0.0402, 0.1, 0.0544, 0.0544, 0.0873, 0.0873, 0.0643, 0.0643, 0.0446, 0.0446, 0.0889, 0.0889, 0.0738, 0.0738)

So, the order of the components' DIF is obtained:

$$X3 > X12(X13) > X6(X7) > X14(X15) > X8(X9) > X4(X5) > X10(X11) > X2 > X1$$

Similarly, the ranking of all MCSs can also be obtained:

$$X3 > X6.X7 > X12.X13 > X12.X15(X13.X14) > X14.X15 > X2 > X6.X9(X8.X7) \\ > X1 > X4.X7(X6.X5) > X10.X13(X12.X11) > X10.X15(X14.X11) > X8.X9 \\ > X8.X5(X4.X9) > X4.X5 > X10.X11$$

We assume that a sensor monitors $X6$ and detects that it is in a work state. We can use this evidence information to simplify the characteristic function and obtain an updated system characteristic function:

$$F_{CUE} = X1 + X2 + X3 + X4X5 + X4X7 + X4X9 + X8X5 + X8X7 + \\ X8X9 + X10X11 + X10X13 + X10X15 + X12X11 + \\ X12X13 + X12X15 + X14X11 + X14X13 + X14X15$$

In addition, this evidence information can be input into the DEN and the corresponding evidence is as follows:

$$P(X6 = \{W\}) = 1, P(X6 = \{W, F\}) = P(X6 = \{F\}) = 0 \quad (18)$$

Using the DEN reasoning algorithm, the updating DIFs of components and CUEs are shown in Table 7 and Table 8 respectively.

Using the sorting method, we can get the order of components:

$$X3 > X12(X13) > X14(X15) > X7 > X2 > X10(X11) > X1 > X9 > X5 > X8 > X4$$

Based on the proposed diagnosis algorithm, we can get the DDT of train-ground wireless communication system without sensors information, shown in Fig. 4 and the corresponding DDT which incorporates sensors information into diagnosis process shown in Fig. 5.

Since the failure probability of CUE is expressed as an interval number, it cannot be directly used to calculate EDC. For convenience, assuming that all components have a unit test cost and test cost of components is independent, we calculate EDC using the median of the interval number in Equation (17). Table 9 shows the EDC of different diagnostic algorithms and indicates the proposed method is more efficient than others.

4. Conclusion

In this paper, a novel fault diagnosis approach for complex systems is presented based on DFT analysis and DEN, which aims to deal with two important issues that arise in engineering applications, such as failure dependency and epistemic uncertainty. For the challenge of failure dependency, a DFT is used to describe the dynamic fault behaviours. For the challenge of the epistemic uncertainty, the failure rates of components in complex systems are

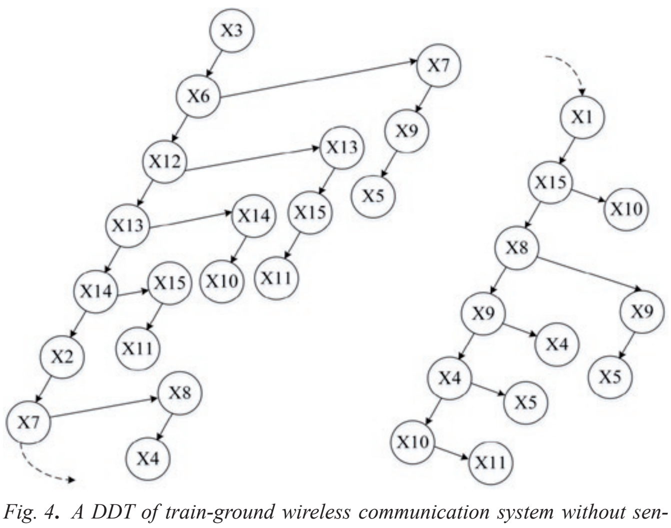


Fig. 4. A DDT of train-ground wireless communication system without sensors information.

Table 7. The updating DIFs of components

Components	DIF of components	Components	DIF of components
X1	[0.0571, 0.0581]	X8	[0.0125, 0.0137]
X2	[0.0797, 0.0809]	X9	[0.0449, 0.0541]
X3	[0.6381, 0.6420]	X10,X11	[0.0762, 0.0807]
X4	[0 , 0]	X12,X13	[0.2165, 0.2302]
X5	[0.0394, 0.0474]	X14	[0.1923, 0.2044]
X7	[0.1030, 0.1237]	X15	[0.1923, 0.2044]

Table 8. The updating DIFs of CUEs

CUEs	DIF of CUEs	CUEs	DIF of CUEs	CUEs	DIF of CUEs
X1	[0.0393,0.0635]	X8.X5	[0.0147,0.0281]	X10.X15	[0.0227,0.0432]
X2	[0.0553,0.0873]	X8.X7	[0.0384,0.0732]	X12.X11	[0.0255,0.0488]
X3	[0.4428,0.6954]	X8.X9	[0.0167,0.0321]	X12.X13	[0.0725,0.1383]
X4.X5	[0.0129,0.0246]	X10.X11	[0.0090,0.0172]	X14.X11	[0.0227,0.0432]
X4.X7	[0.0336,0.0642]	X10.X13	[0.0255,0.0488]	X14.X13	[0.0644,0.1224]
X4.X9	[0.0147,0.0281]	X12.X15	[0.0644,0.1224]	X14.X15	[0.0572,0.1084]

Table 9. EDC of Different diagnostic algorithms

Diagnostic algorithms	EDC
A diagnostic method proposed by Assaf [1]	7.638
A diagnostic method without sensors information	7.211
A diagnostic method with incorporating sensors information	5.519

expressed in interval numbers. Furthermore, qualitative analysis of a DFT is to generate the characteristic function via a zero-suppressed binary decision diagram, while quantitative analysis is to calculate some importance measures by converting a DFT into a DEN. In addition, these reliability results are updated according to the evidence information from sensors and used to design a novel algorithm to improve the diagnosis efficiency. Finally, a real example is given to demonstrate the feasibility and efficiency of the proposed method. This method takes full advantages of both DFT for modelling and DEN for the uncertainty inference, which is especially suitable to diagnose complex systems.

In the future work, we will focus on how the reliability of sensors influences the diagnosis efficiency.

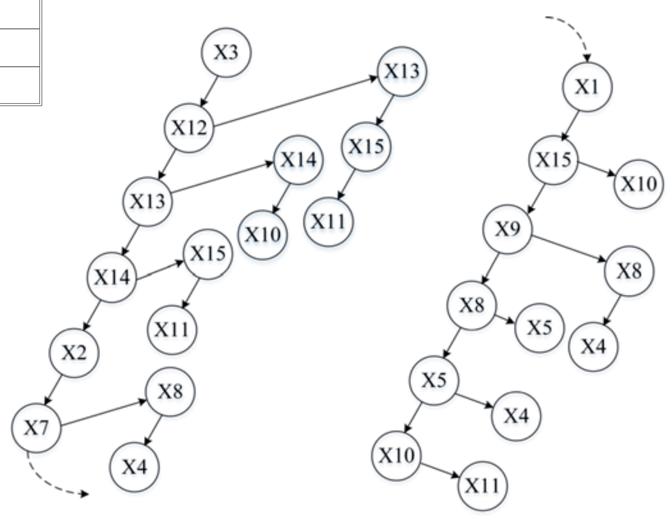


Fig. 5. A corresponding DDT which incorporates sensors information into diagnosis process.

Acknowledgement

This research was funded by China Postdoctoral Science Foundation [2015M580568], National Natural Science Foundation of China [71461021] and Graduate Innovation Foundation of Nanchang University [CX2017185].

References

1. Assaf T, Dugan J B. Design for diagnosis using a diagnostic evaluation measure. Instrumentation & Measurement Magazine 2006; 9(4): 37-43, https://doi.org/10.1109/MIM.2006.1664040.
2. Assaf T, Dugan J B. Diagnosis based on reliability analysis using monitors and sensors. Reliability Engineering & System Safety 2008;

93(4): 509-521, https://doi.org/10.1016/j.ress.2006.10.024.
3. Cai B, Liu H, Xie M. A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks. Mechanical Systems and Signal Processing 2016; 80: 31-44, https://doi.org/10.1016/j.ymssp.2016.04.019.
4. Cao Y, Li T, Wang J, Xie R, Wang X. Fuzzy dynamic fault tree analysis for electro-mechanical actuator based on algebraic model with common-cause failures. Automatic Control and Computer Sciences 2016; 50(2): 80-90, https://doi.org/10.3103/S0146411616020024.
5. Chen Y, Zhen Z, Yu H, Xu J. Application of Fault Tree Analysis and Fuzzy Neural Networks to Fault Diagnosis in the Internet of Things (IoT) for Aquaculture. Sensors 2017; 17(1): 1-15, https://doi.org/10.3390/s17010153.
6. Chiremsel Z, Said R N, Chiremsel R. Probabilistic Fault Diagnosis of Safety Instrumented Systems based on Fault Tree Analysis and Bayesian Network. Journal of Failure Analysis & Prevention 2017; 16(5): 747-760, https://doi.org/10.1007/s11668-016-0140-z.
7. Doguc O, Ramirez-Marquez J E. Using Bayesian approach for sensitivity analysis and fault diagnosis in complex systems. Journal of Integrated Design and Process Science 2009; 13(1): 33-48.
8. Duan R X, Dong D C, Zhao S M. Information Fusion method for system fault diagnosis based on dynamic fault tree analysis. Journal of Tongji University (Natural Science) 2011; 39(11): 1699-1704.
9. Duan R X, Fan J H. A novel diagnosis strategy for braking system of urban rail transit based on fuzzy set and dynamic fault tree. ICIC Express Letters 2014; 8(7): 1809-1814.
10. Duan R X, Zhou H L. Diagnosis strategy for micro-computer controlled straight electro-pneumatic braking system using fuzzy set and dynamic fault tree. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2014; 16(2): 217-223.
11. Duan R X, Hu L F and Lin Y N. Fault diagnosis for complex systems based on dynamic evidential network and multi-attribute decision making with interval numbers. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19(4): 580-589, https://doi.org/10.17531/ein.2017.4.12.
12. Huang H Z, Li Y F, Sun J, Yang Y J. Fuzzy Dynamic Fault Tree Analysis for the Solar Array Drive Assembly. Journal of Mechanical Engineering 2013; 49(19): 70-76, https://doi.org/10.3901/JME.2013.19.070.
13. Hui K H, Lim M H, Leong M S, Al-Obaidi S M. Dempster-Shafer evidence theory for multi-bearing faults diagnosis. Engineering Applications of Artificial Intelligence 2017; 57: 160-170, https://doi.org/10.1016/j.engappai.2016.10.017.
14. Jiang C, Zhang Z, Han X and Liu J. A novel evidence-theory-based reliability analysis method for structures with epistemic uncertainty. Computers & Structures 2013; 129: 1-12, https://doi.org/10.1016/j.compstruc.2013.08.007.
15. Kabir S, Walker M, Papadopoulos Y, Rüde E, Securius P. Fuzzy temporal fault tree analysis of dynamic systems. International Journal of Approximate Reasoning 2016; 77: 20-37, https://doi.org/10.1016/j.ijar.2016.05.006.
16. Khakzad N, Khan F, Amyotte P. Risk-based design of process systems using discrete-time Bayesian networks. Reliability Engineering and System Safety 2013; 109: 5-17, https://doi.org/10.1016/j.ress.2012.07.009.
17. Li Y F, Huang H Z, Liu Y, Xiao N C, Li H Q. A new fault tree analysis method: fuzzy dynamic fault tree analysis. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2012; 14(3): 208-214.
18. Liu Z, Li Y F, He L P, Yang Y Z, Huang H Z. A new fault tree analysis approach based on imprecise reliability model. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 2014; 228(4): 371-381, https://doi.org/10.1177/1748006X14520824.
19. Mi J, Li Y ., Yang Y J, Peng W, Huang H Z. Reliability assessment of complex electromechanical systems under epistemic uncertainty. Reliability Engineering & System Safety 2016; 152: 1-15, https://doi.org/10.1016/j.ress.2016.02.003.
20. Philippe W, Christophe S. Dynamic evidential networks in system reliability analysis: A Dempster-Shafer Approach. Proceedings of the 16th Mediterranean Conference on Control & Automation, Ajaccio, France 2008; 603-608.
21. Şahin R, Liu P. Possibility-induced simplified neutrosophic aggregation operators and their application to multi-criteria group decision-making. Journal of Experimental & Theoretical Artificial Intelligence 2017; 29(4): 769-785, https://doi.org/10.1080/0952813X.2016.1259266.
22. Sun H L, Yao W X. Comments on methods for ranking interval numbers. Journal of Systems Engineering 2010; 25(3): 18-26.
23. Tao J Y, Dong D C. An improved method for system fault diagnosis using fault tree analysis. Journal of Harbin Institute of Technology 2010; 42(1): 143-147.
24. Tang Z, Dugan J B. Minimal cut set / sequence generation for dynamic fault trees. Proceedings of Annual Symposium Reliability and Maintainability, Los Angeles, CA, USA, 2004; 207-213.
25. Wang D, Zhang Y, Jia X, Jiang P, Guo B. Handling Uncertainties in Fault Tree Analysis by a Hybrid Probabilistic-Possibilistic Framework. Quality and Reliability Engineering International 2016; 32(3): 1137-1148, https://doi.org/10.1002/qre.1821.
26. Yuge T, Yanagi S. Dynamic fault tree analysis using Bayesian networks and sequence probabilities. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences 2013; 96(5): 953-962, https://doi.org/10.1587/transfun.E96.A.953.
27. Zhang R, Zhang L, Wang N, Wang W. Reliability evaluation of a multi-state system based on interval-valued triangular fuzzy Bayesian networks. International Journal of System Assurance Engineering and Management 2016: 7(1), 16-24, https://doi.org/10.1007/s13198-015-0335-9.
28. Zhang X, Mahadevan S, Deng X. Reliability analysis with linguistic data: An evidential network approach. Reliability Engineering & System Safety 2017; 162: 111-121, https://doi.org/10.1016/j.ress.2017.01.009.

Rongxing DUAN
Yanni LIN
Yining ZENG
School of Information Engineering
Nanchang University, 999 Xuefu Rd., Honggutan new district
Nanchang, Jiangxi, China

E-mail: duanrongxing@ncu.edu.cn, Nikkilin@yeah.net, 1244942196@qq.com

Article citation info:
SANTANA JMM, SANTIAGO RLV, MOURA MC, LINS ID. Extended warranty of medical equipment subject to imperfect repairs: an approach based on generalized renewal process and Stackelberg game. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 567–578, http://dx.doi.org/10.17531/ein.2018.4.8.

João Mateus Marques de SANTANA
Rafael Luiz Velôzo SANTIAGO
Márcio das Chagas MOURA
Isis Didier LINS

EXTENDED WARRANTY OF MEDICAL EQUIPMENT SUBJECT TO IMPERFECT REPAIRS: AN APPROACH BASED ON GENERALIZED RENEWAL PROCESS AND STACKELBERG GAME

ROZSZERZONA GWARANCJA NA SPRZĘT MEDYCZNY PODLEGAJĄCY NIEPEŁNYM NAPRAWOM: PODEJŚCIE OPARTE NA UOGÓLNIONYM PROCESIE ODNOWY I MODELU STACKELBERGA

Due to its advanced technology, maintenance services of healthcare equipment have been commonly executed by the original equipment manufacturer (OEM), which can be characterized as a monopolist. In this context, hospitals require high availability of their equipment at a reasonable servicing cost, whereas OEM aims to maximize its profit by selling extended warranty (EW) services for multiple consumers. The issue of drawing a maintenance contract between OEM and hospitals has already been treated by adopting a Stackelberg's game. However, the "as good as new" and "as bad as old" assumptions are usually considered, which are rather difficult to observe in practice, especially for healthcare institutions and their technology-intensive equipment. Thus, we here adopt generalized renewal processes (GRP) for modelling imperfect repairs, and we develop a discrete event simulation method for finding the best strategies of each player: OEM sets the prices for EW and on-demand maintenance that optimize its profit, while hospitals choose which option they should hire. We also present an application example with real data gathered from an angiography device, which is used for mapping blood vessels and diagnosing heart diseases.

Keywords: *medical equipment; extended warranty; Stackelberg game; generalized renewal process.*

Ze względu na zaawansowanie technologiczne sprzętu medycznego, jego obsługą serwisową zazwyczaj zajmuje się producent sprzętu oryginalnego (OEM), co czyni go monopolistą w tym zakresie. Podczas gdy szpitalom zależy na wysokiej gotowości sprzętu przy rozsądnych kosztach obsługi, OEM dąży do maksymalizacji zysku poprzez sprzedaż rozszerzonej gwarancji na usługi serwisowe wielu klientom. Istnieją już badania, w których kwestię zawierania umowy o świadczenie usług serwisowych między OEM a szpitalami analizowano z zastosowaniem modelu Stackelberga. Jednak zwykle badania te zakładają, że stan po naprawie może być albo "jak fabrycznie nowy" albo"jak przed uszkodzeniem", co rzadko spotyka się w praktyce, zwłaszcza w przypadku placówek służby zdrowia i ich zaawansowanego technologicznie sprzętu. W związku z tym, w przedstawionej pracy, przyjęto uogólniony proces odnowy (GRP) do modelowania niepełnych napraw oraz opracowano metodę symulacji zdarzeń dyskretnych w celu znalezienia najlepszych strategii dla każdego gracza: OEM ustala ceny rozszerzonej gwarancji oraz konserwacji na żądanie, tak by zoptymalizować swój zysk; szpital natomiast ustala, którą opcję powinien wybrać. W pracy przedstawiono również przykład zastosowania omawianego podejścia z wykorzystaniem rzeczywistych danych zebranych z angiografu, który służy do obrazowania naczyń krwionośnych i diagnozowania chorób serca.

Słowa kluczowe: *sprzęt medyczny; rozszerzona gwarancja; gra Stackelberga; uogólniony proces odnowy.*

1. Introduction

Medical equipment plays an important role in modern healthcare institutions because they present the following purposes: diagnosis, disease prevention, monitoring and patient treatment. During the last decades, as technology has advanced, the maintenance of such equipment has become too complex to be done in-house. Therefore, this activity has been commonly assigned to the Original Equipment Manufacturer (OEM).

In this context, the importance of studying more deeply the warranty issue is reinforced by the role maintenance management plays in guaranteeing the quality of healthcare services. In diverse situations, human lives depend on the correct operation of medical devices. For instance, it is estimated that an amount of US$ 24.83 billion will be annually spent in the medical equipment maintenance market by 2022 [20]. However, maintenance outsourcing for medical equipment is yet to be fully explored by current research (Cruz & Rincon [5]).

Warranty policies define responsibilities for both parties: the OEM and hospital managers. Indeed, the OEM is responsible for repairing the devices' eventual failures related to problems of equipment design, manufacturing and/or quality. The customers in turn should make proper use of equipment, i.e., they must comply to the specifications defined by the OEM (Rahman & Chattopadhyay [32]).

Given that, a new trend has been intensified by manufacturers, selling an additional, optional coverage, which begins after the expiration of the base warranty, called Extended Warranty (EW) (Murthy

& Djamaludin [26]). Thus, the customer decides whether to pay an extra value at the purchase epoch (Murthy & Jack [27]), whereas the OEM will correctively maintain equipment for a given period even after the ordinary warranty expires.

It is important to emphasize that EW ends up creating a conflict of interests between the owner of the equipment and the OEM. Specifically, the customer needs a high availability of its equipment at a reasonable servicing cost, whereas the manufacturer aims to maximize their profit with the addition of post-selling services. Consequently, EW affects both buyers' and manufacturer's outcomes (Ye & Murthy [41]), once the actions of one interferes in the results of the other. Due to this situation, Game Theory provides an appropriate approach to solve this problem (Forgó et al. [10]).

Among different games that can be used to model the interaction between agents, the leader-follower Stackelberg Game (SG) is a good option for drawing maintenance service contracts of medical equipment. In this context, the OEM is commonly the only party able to perform maintenance, since it has the technical knowledge, expertise, technology and spare parts for the repair execution (Rinsaka & Sandoh [33]). The hospital in turn needs its medical device available in suitable condition to provide a good service. From these two perspectives, an uneven power relationship is noted, which can be modeled via SG. The leader role is assigned to the OEM, which determines the terms of the EW, while the health institution acts as the follower, responding to actions taken by the OEM.

Quantitative studies about warranty, maintenance outsourcing and maintenance contracts are present by Kim et al. [18], Bouguerra et al. [3], Husniah et al. [14], Huang et al. [13], Moura et al. [23] and Darghouth et al. [6]. However, such studies have simplifying assumptions with respect to the state of the system after a corrective maintenance (CM) intervention. In fact, those papers considered that the system returns to either an "as good as new" condition (perfect repair) or an "as bad as old" condition (minimal repair); these two situations are modeled respectively according to a Renewal Process (RP) and a Non-Homogeneous Poisson Process (NHPP); Ross [34] describes RP and NHPP in details. The use of these assumptions may yield inadequate managerial decisions, which can result in significant losses in company profits because of incorrect definition of warranty policies.

In practical terms, maintenance actions typically return the equipment to an intermediate condition between the perfect and minimal repairs, which is called imperfect repair (Kijima & Sumita [17]; Wang & Pham [37]). Kijima & Sumita [17] proposed two methods to tackle imperfect repairs: Kijima type I and Kijima type II, which gave rise to the Generalized Renewal Process (GRP) and introduced the concept of "virtual age". Furthermore, these situations generalize RP and NHPP; other approaches may be seen in Pham & Wang [31].

In this paper, we aim to join SG and GRP to model imperfect repairs – a more realistic and general assumption. This approach considers the interaction between OEM and multiple customers. The proposed model will be characterized as follows. First, we considered the OEM offers to the hospital managers two maintenance options for the period after the ordinary warranty expires: (i) an extended warranty or (ii) on-call service. EW states that for a fixed price P, the OEM should repair all failures without any additional cost over the period of the contract; if a failed unit does not get repaired before a set time τ, a penalty, which increases over time, will be incurred. For the on-call service, failures will be repaired at a fixed cost C_s each, with no penalty incurred in case of delays. Given a number of hospitals that buy the equipment, this model estimates P^* and C_s^*, which are the maximal prices hospitals accept to pay for each maintenance option. Finally, we find the optimal number of customers, i.e., number of hospitals that maximize OEM's profit.

To that end, we develop a Discrete Event Simulation (DES) based method to reproduce the GRP-queue system and obtain the performance indicators of interest. Simulation models allow for analysis of systems with complex behavior, require fewer assumptions when compared with analytical models and are used as tools to perform experimentations with systems (Marsaro & Cavalcante [21]). Due to present model's complexity, an analytical solution is unfeasible. In this sense, simulation is useful for describing equipment behavior (Ding & Kamaruddin [7]). Thus, a simulation approach is employed for obtaining a solution for the present model.

The remainder of this paper unfolds as follows. In Section 2, the theoretical background is provided, containing the adaption of the SG for the context of maintenance contracts, as well as the characteristics of GRP, emphasizing the method proposed by Yañez et al. [39]. Section 3 presents the proposed model, the players' optimal strategies and the equilibrium of the game. In Section 4, a numerical example is presented, using real data from an angiograph, which is a device used for blood vessels mapping and diagnosis of organ diseases. Finally, Section 5 concludes remarks.

2. Theoretical background

2.1. Generalized renewal process

RP and NHPP may be adopted to model perfect and minimal repairs respectively. According to (Lins & Droguett [19]), such methods have simplifying assumptions that may be unreal in many practical situations such as the healthcare technology intensive environment. To overcome the limitations of RP and NHPP, Kijima & Sumita [17] developed a probabilistic virtual age based model, known as the Generalized Renewal Process (GRP) that deals with all classes of maintenance actions. According to this model, q (rejuvenation parameter) may generally assume values between 0 and 1:

- $q = 0$ represents a perfect repair (as good as new);
- $q = 1$ corresponds to a minimal repair (as bad as old);
- $0 < q < 1$ indicates imperfect repair (better than old, worse than new).

Cases, where $q < 0$ and $q > 1$, are also possible corresponding to *better than new* and *worse than old* conditions respectively. Generally, GRP may be classified into two types (Kijima Type I and II), according to the method used to calculate the virtual age. These types can be seen in details in Moura et al. [22] and other virtual age-based representations could be found in Guo et al. [12], Tanwar et al. [35], Ferreira et al. [9], Oliveira et al. [29] and Wang & Yang [38].

This paper uses Kijima type I so that equipment virtual age follows Eq. (1), according to which maintenance actions only act on the degradation incurred during x_i, which is the time between $(i-1)^{\text{th}}$ repair and the i^{th} failure:

$$v_i = v_{i-1} + qx_i = q\sum_{j=1}^{i} x_j \tag{1}$$

The Cumulative Distribution Function (CDF) for the time between the $(i-1)^{th}$ and i^{th} failures can be determined from the CDF of the time until a failure conditioned on the virtual age v_{i-1} as seen in Eq. (2):

$$F(x_i|v_{i-1}) = P(X \le x_i|X > v_{i-1}) = \frac{F(v_{i-1} + x_i) - F(v_{i-1})}{1 - F(v_{i-1})} \tag{2}$$

For our analysis, we consider the time to failure follows a conditioned Weibull distribution because of its flexibility to fit various degradation stages. Then, Eq. (2) can be rewritten as Eq. (3). Note that, for $i = 1$, we have the Weibull distribution itself because $v_0 = 0$. When there are reasonably sufficient failure data available, Maximum Likelihood Estimators (MLE) can be used to estimate GRP parameters α, β, and q. To that end, the procedure described in Yañez et al. [39] can be followed.

$$F\left(x_i|v_{i-1}\right) = 1 - exp\left[\left(\frac{v_{i-1}}{\alpha}\right)^{\beta} - \left(\frac{v_{i-1}-x_i}{\alpha}\right)^{\beta}\right] \tag{3}$$

In the proposed model, we consider a case with multiple clients and m service channels, where maintenance teams may serve a total of M clients. If the number of failed units is greater than the service capability, a queue is formed. The waiting time in queue is decisive to the interests of both hospitals and OEM, since the former aims high availability of equipment and the latter wants to serve a higher number of customers. Therefore, a queuing formulation is employed to incorporate system behavior due to the interaction between service level and number of clients to be served.

Then, GRP governs the equipment failure process and the queue discipline follows a FCFS (first come first served) logic. We also consider times to repair follow an exponential distribution. Thus, using the conventional notation, the queue can be described as GRP/Markovian/ m /∞/M/FCFS. For this situation, analytical solutions are not available. Therefore, we proposed a DES-based algorithm is adopted to obtain the GRP-queue system measures; the DES formulation is described in Section 3.8.

2.2. Stackelberg game

SG is a non-cooperative sequential game developed by Heinrich von Stackelberg. It was originally proposed to evaluate the equilibrium of a duopoly, where competing companies decide the optimal quantity to be produced (Gibbons [11]) in a leader-follower interaction.

At the best of authors' knowledge, Murthy & Yeung [28] were the first authors to introduce SG as a tool to model maintenance service contracts. Murthy & Asgharizadeh [25] expanded the problem by creating a game between a customer and a manufacturer by assuming perfect repairs. Ashgarizadeh & Murthy [2] and Murthy & Asgharizadeh [24] incorporated multiple customers and service channels. Esmaeili et al. [8] considered a three-level service contract between a manufacturer, a customer and an independent third agent. Moura et al. [23] used priority queues to analyze the interaction among OEM and two priority classes of hospitals. All aforementioned models intend to maximize the clients' expected utility, considering parameters like risk aversion, revenue generated by the system, maintenance costs and times to repair. In such papers, the leader (OEM) provides the maintenance options for the follower (hospitals) and obtains the highest payoff, since it charges the prices that maximize profits.

Generally, in the context of complex medical equipment, the OEM has a well-trained staff, spare parts and dominates the equipment technology. Thus, OEM behaves as a leader, acts first (by defining services and respective prices) and is the only maintenance service provider. Hospitals need to guarantee minimum levels of availability for their equipment. However, they do not have expertise in the maintenance of complex equipment. Therefore, hospitals can be considered as followers, since they react to the OEM's action when choosing a service type to hire.

The SG's solution is obtained through backwards induction (Osborne and Rubinstein [30]) corresponds to a sub-game perfect Nash equilibrium of a two-stage game, with perfect information and players with different profiles (Amir [1]). Finally, it is noteworthy to say that the papers cited earlier in this Section, which adapted SG to the field of maintenance outsourcing, make simplifying assumptions about the repair structure employed. At the best of authors' knowledge, none of them adopted the imperfect repair assumption, a more realistic hypothesis, especially for complex systems. Thus, this paper aims to incorporate the GRP to model this situation and attempts to make the model more suitable to the medical environment.

3. Model description

3.1. Game formulation

This paper aims to determine the optimal strategies for the problem of EW for the medical context. Thus, we employ a SG formulation to model the interaction between the OEM and each customer, and the game's equilibrium will be reached by finding the reservation prices, i.e., maximum prices that customers accept to pay.

This decision problem extends the model developed by Asharizadeh & Murthy [2] by considering imperfect repairs. Additionally, the problem is stochastic due to the uncertainty inherent to the presence of random variables in the model. Finally, decision makers (healthcare institutions and OEM) have their own objective function that will define the respective payoffs. Both OEM and healthcare institutions are aware about their alternatives, acting rationally, and choosing strategies that maximize their respective payoffs (Osborne & Rubinstein [30]).

3.2. Notation list

A_k : Decision variable of the hospital;

C_b : Purchase price of the equipment;

C_r : OEM's mean cost to repair a failed unit;

C_s : Price charged by the OEM per repair;

C_s^{max} : Hospital's reservation price for the on-call service;

M : Number of customers;

M^* : Optimal number of customers;

n_{rep} : Number of Monte Carlo replications;

$N_{j,1}$, $N_{j,2}$: Number of failures occurred over the intervals W_1 and W_2 , respectively, for the j^{th} device (j^{th} device is held by the j^{th} customer (hospital));

N_j : Total number of failures during W for the j^{th} device; $N_j = N_{j,1} + N_{j,2}$;

P : Price of the extended warranty;

$P^{\max}$: Hospital's reservation price for the extended warranty;

R : Revenue per hour;

T_0 : Time of the purchase of the equipment;

T_a : Time of next failure among available customers;

T_d : Time of next completion of a maintenance intervention;

$\mathrm{T_j^{dt}}$: Total downtime for j^{th} system.

$T_{j,1}^{op}$, $T_{j,2}^{op}$: Period that the j^{th} system is on operational state during W_1 and W_2 respectively;

T_j^{op} : Total period that the j^{th} system is on operational state;

$T_{j,1}^{ov}$, $T_{j,2}^{ov}$: Overtime for the j^{th} device during W_1 and W_2 respectively;

T_j^{ov} : Total overtime for the j^{th} system;

U : Expected utility of the customer;

W_1 : Base (ordinary) warranty period;

W_2 : Extended warranty period;

$W = (W_1 + W_2)$: Model analysis period;

Y_{ij} : Time between failure and completion of repairing the j^{th} equipment after the i^{th} failure.

τ_1, τ_2 : Maximum time that the manufacturer has to return a failed device to an operational state without incurring penalty, when the failure occurs during coverage period W_1 and W_2, respectively;

θ_1 : Penalty per hour for the base warranty period;

θ_2 : Penalty per hour for the EW period;

Π_{OEM} : OEM's expected profit;

δ: Hospital's risk-aversion parameter;

Π_H : Hospital's expected profit.

3.3. Problem description

OEM sells a technology-intensive medical device to multiple customers (hospitals) for a cost of C_b per unit. Each device, when in operational state, generates a revenue of R monetary units per time. Along with the purchase of the device, OEM provides a base warranty, during which OEM is responsible for all repairs required with no charge for the client. If a failed equipment is not returned to operational state within a period τ_1 after a failure occurs, OEM is charged with a penalty proportional to the overtime in repairing the equipment, which is the period from τ_1 to the time when the equipment returns to operation. Therefore, the penalty is $\theta_1(Y_{ij} - \tau_1)$ when $Y_{ij} > \tau_1$, and zero, otherwise; θ_1 is the penalty per time during overtime; Y_{ij} is the time between the occurrence of the i^{th} failure of equipment j and the completion of its respective repair. This penalty exists because medical equipment is vital for patients' treatment and for the hospitals' profit, and then unavailability affects their payoff and reputation.

After the expiration of the base warranty, each hospital may choose a type of repair service: i) EW or ii) on-call services. These options are described as follows:

i) A_1 : EW – begins after the base warranty expires and has duration W_2. The customer pays a fixed price P and the OEM repairs all failed units over W_2 at no additional cost. If a failed device is not returned to operational state within a period τ_2 after a failure occurs, the OEM is charged a penalty. Analogously to the base warranty, the penalty is $\theta_2(Y_{ij} - \tau_2)$, when $Y_{ij} > \tau_2$;

ii) A_2 : on-call services – the OEM executes each repair at a cost of C_s per intervention; no penalty is here incurred.

We also assume that option A_0 means the hospital chooses not to buy the equipment.

3.4. Hospital's decision problem

Considering the options presented by the OEM, the hospital manager decides whether to opt for either EW (A_1), or on-call services (A_2) or not to purchase the equipment (A_0), where the latter occurs if its expected utility is negative over the period W. Each strategy has consequences to the payoffs. Indeed, each customer's return, Π, depends on the option A_k, and thus the hospitals profit can be seen in Eqs. (4), (5) and (6):

$$\Pi_{H(A_0)} = 0 \tag{4}$$

$$\Pi_{H(A_1)} = RT_j^{op} + \theta_1 T_{j,1}^{ov} + \theta_2 T_{j,2}^{ov} - C_b - P \tag{5}$$

$$\Pi_{H(A_2)} = RT_j^{op} + \theta_1 T_{j,1}^{ov} - C_b - N_{j,2} C_s \tag{6}$$

where:

$$T_j^{op} = \sum_{i=1}^{N_j} X_{ij} + \widetilde{X_j}$$

$$T_{j,1}^{op} = \left[\sum_{i=1}^{N_{j,1}} \max\left\{0, (Y_{ij} - \tau_1)\right\} \right]$$

$$T_{j,2}^{op} = \left[\sum_{i=1}^{N_{j,2}} \max\left\{0, (Y_{ij} - \tau_2)\right\} \right]$$

For hospital (equipment) j, ($1 \le j \le M$), N_j is the total number of failures over the mission time W; X_{ij} ($0 \le i \le N_j$) is the time between the i^{th} repair and the $(i+1)^{th}$ failure; $\tilde{X}_j$ is the time between the last failure and W; $N_{j,1}$ and $N_{j,2}$ are the number of failures occurred over W_1 and W_2 respectively; Y_{ij} ($0 \le i \le N_j$) is the total time to finish repairing the j^{th} equipment since the occurrence of the i^{th} failure, i.e., Y_{ij} includes the waiting time in queue and repair time; T_j^{op} is the total operational time during W; $T_{j,1}^{op}$ and $T_{j,2}^{op}$ are the respective operational times during W_1 and W_2; T_j^{ov} is the total overtime during W; $T_{j,1}^{ov}$ and $T_{j,2}^{ov}$ are the total overtimes during W_1 and W_2 respectively.

We consider the hospital's risk is modeled according to a utility function U, which indicates how the customer chooses among distinct options; thus, the preferred options are represented by higher utilities. The utility function considered in this model has been used in Murthy & Asgharizadeh [25] and is shown in Eq. (7), where Π represents the associated wealth:

$$U(\Pi) = \left(\frac{1 - e^{-\delta \Pi}}{\delta} \right) \tag{7}$$

Thus, the choice A_k is strongly affected by equipment availability, pricing structure and the hospital's degree of risk aversion (δ). We assume all customers are homogeneous with respect to their risk aversion and all equipment units are identical regarding their reliability.

3.5. OEM's decision problem

The OEM is considered risk neutral and its expected profit is related to the customer's optimal choice. Consequently, the OEM's payoff can be denoted as $\Pi_{OEM}(P, C_s, M, A_k)$, where (P, C_s, M) are OEM's decision variables and $k = 0,1,2$. In this way, the manufacturer's profits for each of the customer's possible actions A_k are shown respectively in Eqs. (8), (9) and (10). Then, OEM may choose the combination P, C_s, M that maximizes its expected profit, taking into account the customer's optimal strategy A^*:

$$\Pi_{OEM}(P, C_s, M, A_0) = 0 \tag{8}$$

$$\Pi_{OEM}\left(P,C_s,M,A_1\right)=\sum_{j=1}^{M}P-C_rN_j-\theta_1T_{j,1}^{ov}-\theta_2T_{j,2}^{ov} \quad (9)$$

$$\Pi_{OEM}\left(P,C_s,M,A_2\right)=\sum_{j=1}^{M}N_{j,2}\left(C_s-C_r\right)-N_{j,1}C_r-\theta_1T_{j,1}^{ov} \quad (10)$$

3.6. Assumptions

In order to make the model manageable, we consider some assumptions:

I Equipment is repairable and subject to imperfect repair. The probabilistic failure modelling is handled by GRP;

II The times between failures are random variables. The time to first failure follows a Weibull distribution, as seen in Yañez et al. [39];

III The times to repair are exponentially distributed with parameter μ;

IV The OEM has m parallel service channels, i.e., in total they are capable of processing m units simultaneously (one unit per service channel);

V The equipment's failures are critical. Moreover, the OEM carries out just corrective maintenance interventions;

VI The manufacturer and the customer have complete information about the model's parameters, which implies that the leader is aware about the customer's risk parameter and the hospital knows the equipment reliability;

VII If there are more failed units than the number of servers, a queue following a FCFS is generated. This formulation describes a queuing system with finite population M.

3.7. Players' strategies

In order to find the optimal solution for the players and the game equilibrium, we determine how the hospital's and OEM's optimal strategies are defined, as well as understand their relation, and the degree of influence between them. Then, these strategies are shown as follows.

3.7.1. Hospital's optimal strategy

The customer's expected utility U is derived from two random variables (X_{ij}, Y_{ij}), the customer's decision A_k, and the pricing structure (P^{max}; C_s^{max}) imposed by the OEM. Given the assumptions of Section 3.6, the expected utilities for each decision are given in Eqs. (11), (12) and (13) obtained by using Eqs. (4), (5) and (6). For a pair (P, C_s) determined by the manufacturer, the customers analyze their expected utilities and choose the option that returns the highest payoff.

$$E\left[U\left(A_0,P,M,C_s\right)\right]=0 \quad (11)$$

$$E\left[U\left(A_1,P,M,C_s\right)\right]=\frac{1}{\delta}\left\{1-\exp\left[\delta\left(P+C_b\right)\right]E\left[\exp\left(-\delta\left(RT_j^{op}+\theta_1T_{j,1}^{ov}+\theta_2T_{j,2}^{ov}\right)\right)\right]\right\} \quad (12)$$

$$E\left[U\left(A_2,P,M,C_s\right)\right]=\frac{1}{\delta}\left\{1-\exp\left[\delta C_b\right]E\left[\exp\left(-\delta\left(RT_j^{op}+\theta_1T_{j,1}^{ov}-N_{j,2}C_s\right)\right)\right]\right\} \quad (13)$$

The reservation prices depend on equipment expected number of failures $E\left[N_j\right]$, and the number of units sold M. Then, they are fairly influenced by the repair assumption. For instance, by considering imperfect repair ($0\le q\le 1$), it is expected that devices with the same shape $\beta>1$ and scale $\alpha>0$ fail less frequently than in the context presented by Moura et al. [23], who considered minimal repairs. Thus, the change in repair hypothesis can modify the expected payoffs, and consequently a change of strategies. The mathematical approach used to reproduce the imperfect repair is the GRP, a which makes use of DES and allows for modeling imperfect, perfect and minimal repair assumptions, adapting to a broader range of scenarios. Another point to emphasize corresponds to the reservation prices (P^{max}, C_s^{max}), which are defined by Varian [36] as the highest prices a consumer is willing to pay. These prices affect the decision for what strategy A_k is chosen. Thus, determine the hospital's reservation prices is essential to find the OEM's expected profit, once the pricing structure has strong influence on the healthcare's institutions decision. Indeed, given the hospital's reservation prices and the pricing structure imposed by the OEM, the EW model can be seen on its extensive form, as a sequential game tree in Figure 1a, which shows all possible decisions for the players.

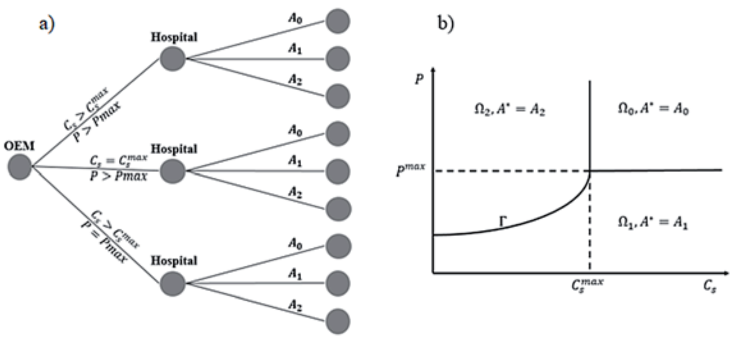


Fig. 1. a) The game tree – P^{max} and C_s^{max} represent the customer's maximum willingness to pay for the EW and on-call maintenance interventions respectively; b) Customer's optimal options adapted from Murthy & Asgharizadeh [25]

In the case the prices charged by the OEM are higher than the reservation prices ($P>P^{max}$; $C_s>C_s^{max}$), then the equipment is not purchased (A_0). Otherwise, if the EW price is superior than P^{max}, while C_s^{max} is not reached ($P>P^{max}$; $C_s\le C_s^{max}$), the hospital manager should choose A_2. Analogously, when $P\le P^{max}$; $C_s>C_s^{max}$, A_1 should be chosen. Figure 1b presents the $P-C_s$ plan, and shows that the customer's optimal choice is characterized by three regions Ω_0, Ω_1 and Ω_2. In Ω_0, $A_k^*=A_0$; in Ω_1, $A_k^*=A_1$; and in Ω_2, $A_k^*=A_2$. The curve Γ is obtained by equalizing the expected utilities for options A_1 and A_2, and represents indifference between the maintenance options since the same payoff is returned to the hospital.

3.7.2. OEM's optimal strategy

Since the OEM is considered to be risk neutral, its optimal strategy corresponds to the pricing structure that maximizes its expected profit (P^{max}, C_s^{max}). To choose its optimal strategy, OEM compares the expected profit for each type of service provided, varying the number of customers M, and then the optimal number of customers is determined.

Considering assumption VI in Section 3.6 (complete information), we conclude the OEM knows the hospital's reservation prices. Thus, OEM builds a structure that captures all the consumer surplus, which implies in the maximization of the producer profit.

Given that, we use Eqs. (12) and (13) to obtain the reservation prices. Indeed, we equalize Eq. (12) to zero, and then isolate P in order to determine $P^{max}\mathrm{P}^{\max}$. Thus, the reservation price of the EW can be given in (14):

$$P^{max} = -C_b - \frac{1}{\delta}\ln E\left[\exp\left(-\delta\left(RT_j^{op} + \theta_1 T_{j,1}^{ov} + \theta_2 T_{j,2}^{ov}\right)\right)\right] \quad (14)$$

Now, we calculate C_s^{max} by using Eq. (13). However, since it's impossible to isolate C_s, we use a numerical method to find its expected value. Eq. (15) shows C_s^{max} equilibrium equation:

$$\delta C_b + \ln E\left[\exp\left(-\delta\left(RT_j^{op} + \theta_1 T_{j,1}^{ov} - N_{j,2} C_s^{max}\right)\right)\right] = 0 \quad (15)$$

Notice that in Eqs. (14) and (15), P^{max} and C_s^{max} can be obtained provided the values of T_j^{op}, $T_{j,1}^{ov}$, $T_{j,2}^{ov}$ and $N_{j,2}$, which are random variables. Therefore, we developed a DES-based algorithm, which is explained in more detail in the following Section, to obtain these values.

3.8. GRP-queue model simulation

As seen in Section 2.1, the presented formulation describes a queuing system with finite population (M), where times until failures follow a GRP. According to assumption III in Section 3.6, times to repair are exponentially distributed. If the number of failed units is greater than the number of servers (m), a queue is formed, and follows a FCFS rule.

In order to find the optimal reservation prices P^{max} and C_s^{max}, it is necessary to simulate the alternating failure-repair process considering a GRP/Markovian/ m / ∞/M//FCFS queuing system. We adopted a DES-based approach to represent the behavior of the system of interest, allowing us for modeling and solving problems that would otherwise be considered intractable or too complex (Zio [42]).

The DES algorithm we developed for queue simulation is shown in Figure 2. This proposed algorithm is similar to what was presented by Moura et al. [23], with two main changes: (i) our simulation model covers both base and extended periods, while the aforementioned paper only simulates the extended period; (ii) our model is formulated for a single customer class, while the aforementioned work allows for two priority classes, which are not implemented here. Our formulation is explained as follows.

At $t = 0$, all M devices are considered new, the time to first failure for each equipment follows a Weibull distribution, queue is empty, and the first failure immediately begins its respective repair. Next, we track if the future events are failures or repairs based on the min (t_a, t_d), where t_a is the next time of arrival (occurrence of a failure) and t_d is the time of departure (next repair completion). Note that t_d is initially set to infinity in step 1.2, since there is no equipment being repaired; as a consequence, it will never be smaller than t_a. If a failure occurs when all service crews are busy, i.e., there are at least m failed units, the just failed device waits in queue. If there is at least one available service crew when a failure occurs, this failed device will immediately begin its repair.

This procedure will continue over the total period (W). We also consider that if a failure occurs during W_1, but its respective repair is completed during W_2 with overtime, the penalty charged is θ_1. During simulation, all events of interest (times of failures and repairs, number of failures for each device, downtime and overtime of each device) are logged so that they can be used to obtain information about system availability, number of failures (N_j), downtime of equipment (T_j^{dt}) and overtime (T_j^{ov}). These are the main outputs of the simulation (step 3), used later to obtain the prices of EWs and on-call services.

After the queue is simulated, customer's optimal strategy is defined. To that end, we estimate the reservation prices for each option offered by the OEM, which are given by Eqs. (14) and (15), using data from many replications (n_{rep}) of the simulation. The complete process used to find the optimal number of customers and their respective reservation prices is given in Figure 3, and is more detailed in Figure 4. Note that the prices that the customers are willing to pay for each option depends on their risk aversion δ.

Initially, the population size (number of devices) is $M = 1$, and then the DES is repeated for the defined number of replications, resulting in estimates for P^{max} and C_s^{max}. Since the OEM has limited service capability (the number of service crews), M cannot be increased indefinitely because high values of M result in longer waiting times in queue, thus reducing operational time and increasing the occurrence of penalties. Therefore, M is increased and this process is repeated until the OEM's profits decrease for both repair service strategies (so that optimal number of hospitals served is guaranteed to be found); this is possible since, increasing M beyond its optimal value results in considerable increases in queue waiting times and OEM penalties (Ashgarizadeh & Murthy [2]). Finally, the optimal number of hospitals and the respective service prices are found by choosing the number of devices that result in the highest values for OEM's profit.

Inputs: $\alpha, \beta, q, \mu, m, \tau_1, \tau_2, W_1, W_2$

1. **Initialization**
   - 1.1. Generate first failure time for each equipment and set t_a equal to the earliest failure time
   - 1.2. For convenience, set $t_d = \infty$
   - 1.3. Set initial time: $t = 0$
2. **Simulation**
   - 2.1. Is the next event a failure ($t_a \le t_d$) or a completion of repair ($t_d < t_a$)?
     - 2.1.1. Failure of device j (only if $t_a < W$)
       - Update current time $t = t_a$
       - Store failure information
         - Increment number of failures for device j for current warranty period ($N_{j,w}$)
         - Increment number of failed devices
         - Store failure time
       - If there is at least one available service crew
         - Allocate failed device to a service crew, and mark the crew as busy
         - Generate repair duration r and set $y = r$
         - $t_d = t + y$
         - Increase downtime by y for device j (T_j^{dt})
         - Store repair duration and departure time
         - If $y > \tau_i$, register overtime of $y - \tau_i$ for device j for current warranty period i ($T_{j,w}^{ov}$)
       - Set t_a equal to the earliest failure time among the remaining operational equipment
     - 2.1.2. Completion of repair on device j
       - Update current time $t = t_d$
       - Store repair information
         - Decrement number of failed devices
         - Generate a failure time t'_a for this device
         - If $t'_a < t_a$, then set $t_a = t'_a$
       - Are there any remaining failed devices queued?
         - Yes
           - Let u = time spent in queue
           - Generate repair duration r and set $y = u + r$
           - $t_d = t + y$
           - Increase downtime by y for device j (T_j^{dt})
           - Store repair duration and departure time
           - If $y > \tau_i$, register overtime of $y - \tau_i$ for device j for current warranty period i ($T_{j,w}^{ov}$)
         - No
           - Set service crew as free
           - Set $t_d = \infty$
   - 2.2. Is $t_a > W$ and the queue empty?
     - 2.2.1. Yes: go to step 3.
     - 2.2.2. No: repeat step 2.
3. **Output generation**: for each device, during each warranty period, return these measures:
   - 3.1. Number of failures ($N_{j,1}, N_{j,2}$)
   - 3.2. Downtime (T_j^{dt})
   - 3.3. Overtime ($T_{j,1}^{ov}, T_{j,2}^{ov}$)

Fig. 2. DES algorithm used to simulate the queue system

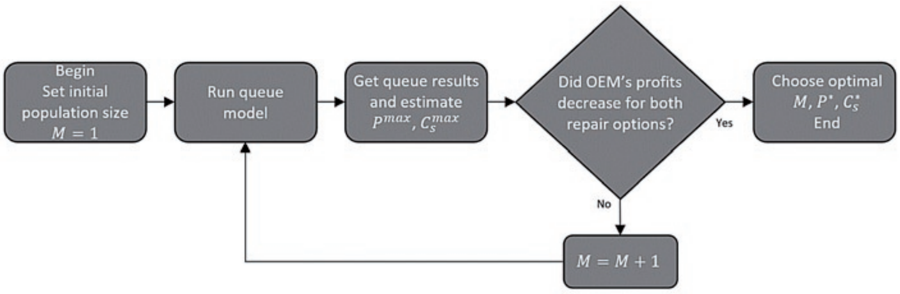


Fig. 3. Diagram of the optimization process

Inputs: $\alpha, \beta, \mu, q, \theta_1, \theta_2, \delta, \tau_1, \tau_2, R, C_b, C_r, W_1, W_2, n_{rep}$
1. **Initialization**: set initial population size $M = 1$
2. **Model execution**
   2.1. Run queue simulation for the current population M for the required n_{rep}
   2.2. Get queue measures from simulation
      2.2.1. Estimate C_s^{max} and P^{max}, using Eqs. (14) and (15)
      2.2.2. Choose best strategy
      2.2.3. Store OEM's profit for each strategy
   2.3. Did OEM's profits decrease for both repair service options?
      2.3.1. No: increment population size $M = M + 1$ and repeat step 2.
      2.3.2. Yes: go to step 3.
3. **Results definition**
   3.1. Choose population size and strategy which resulted in highest OEM profit (M^*)
   3.2. Define C_s and P accordingly

Fig. 4. Monte Carlo based algorithm for the optimization process

4. Application example

4.1. Model's parameters estimation

An application example is here presented by using a failure database of an angiography device, which is technology-intensive, and supports the treatment and diagnosis of cardiovascular diseases. The angiogram, used for visualization of arteries based on x-rays, begins with introduction of an iodine contrast material injection into blood vessels though a catheter. X-ray angiogram provides anatomical information about blood vessels (Çimen et al. [4]). By watching the flow of the contrast fluid, the doctor can identify obstructions and narrowing, proceeding with treatment.

Angiography failures can result in incorrect diagnosis and inappropriate patient treatment, having negative consequences in the patient's health and hospital's reputation. Thus, angiographies are fundamental for hospitals profitability, and their unavailability represents a great loss of revenue, resulting in a negative economic impact.

Table 1 shows 38 times between critical failures and their respective times to repair. Considering this device is subject to imperfect repairs, we need to use the simulation-based solution proposed in the previous Section to reproduce the GRP-queue system. To that end, we first obtain the MLEs for the GRP parameters by using the procedure described in Yañez et al. [39].

We also obtained the MLE's for parameters α and β when the assumptions of perfect $(q=0)$ and minimal $(q=1)$ repairs are adopted. By restricting q to fixed values, it is expected that MLE's result in inferior likelihood, or at most as good as that of obtained by imperfect repair assumption, since parameter search space is restricted. The MLE's for α, β and q for each repair hypothesis are given in Table 2, which also shows the mean squared error (MSE) for simulation data with each repair assumption in relation to observed data. Note that the lowest MSE is obtained when we considered imperfect repairs, which attests that imperfect repair is the most suitable assumption for this case.

Next, we estimate the expected number of failures $E\left[N_j\right]$ by using the procedure described by Yañez et al. [39], and we compared the results against the observed failure data in Table 1. Figure 5 shows this comparison under assumptions of imperfect repairs.

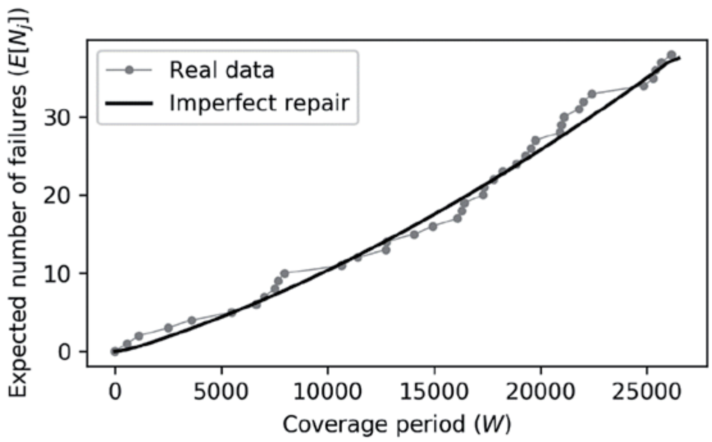


Fig. 5. Comparison between real and simulated times to failure under different repair assumptions

Table 1. Failure dataset of an Angiography device

i^{th} failure	X_i (hours)	Y_i (hours)	i^{th} failure	X_i (hours)	Y_i (hours)	i^{th} failure	X_i (hours)	Y_i (hours)
1	576	4.25	14	24	1.5	27	192	1.45
2	552	2.25	15	1,320	1.5	28	1,152	1.25
3	1,368	4.75	16	864	1.5	29	72	4.75
4	1,104	6	17	1,152	1	30	120	2
5	1,872	3.25	18	216	2.5	31	720	4
6	1,152	2.25	19	120	4	32	192	1.25
7	384	2.25	20	888	4	33	408	2.5
8	504	1.5	21	48	1	34	2,424	5
9	144	2.25	22	432	7.25	35	456	0.25
10	312	3.5	23	432	2	36	96	0.5
11	2,688	4.75	24	648	2.25	37	264	3.5
12	744	3.5	25	432	1	38	480	2.5
13	1,320	1.25	26	264	1.5	–	–	–

For the application example, we will use the parameters shown in Table 3. The inputs given above were used to feed the simulation algorithm described in Figure 2. In order to find the results, the procedures given in Figure 3 and Figure 4 are executed using $n_{rep} = 1{,}000{,}000$ replications and varying the number of equipment (M), and then P^{max} and C_s^{max} are estimated for each M .

By using those parameters and GRP MLE's, the optimal number of equipment, and hospital's reservation prices obtained were respectively: $M^* = 47$, $P^{max} = \$\,183{,}991$ and C_s^{max} = \$ 8,550. As it can be

Table 2. Parameter estimates

Repair hypothesis	$\hat{\alpha}$	$\hat{\beta}$	$\hat{q}$	MSE
Minimal Repair	1,622.55	1.309	1	1.62
Imperfect Repair	1,351.83	1.658	0.097	1.54
Perfect Repair	717.83	1.120	0	9.18

Table 3. Parameters for the application example

Angiograph sale price (C_b)	\$ 1,476.49 (10^3)
Revenue per time (R)	\$ 0.094 (10^3) h^{-1}
Cost of the repair (C_r)	\$ 2.5 (10^3)
Mean repair rate (μ)	0.2 h^{-1}
Penalty per time (base warranty) (θ_1)	\$1 (10^3) h^{-1}
Penalty per time (under extended warranty) (θ_2)	\$3 (10^3) h^{-1}
Hospital's risk aversion (δ)	0.1
Period of the game (W)	2 years = 17,520 h
Base warranty period (W_1)	1st year = 8,760 h
Extended warranty period (W_2)	2nd year = 8,760 h
Maximal time to repair the equipment under EW ($\tau_1 = \tau_2 = \tau$)	12 h

seen, the price of hiring an EW is near the range between 5% and 12% of the purchase price C_b, which is consistent to what happens in the clinical environment (Murthy & Djamaludin [26]). The OEM's expected profit for the options (A_1, A_2) are respectively $E\left[\Pi_{OEM}\left(A_1\right)\right]$ = \$ 2,979,171 and $E\left[\Pi_{OEM}\left(A_2\right)\right]$ = \$ 2,155,601. Then, the Nash Equilibrium occurs when the OEM sets $P = P^{max}$ and $C_s > C_s^{max}$, returning an expected profit of \$ 2,979,171. Finally, the optimal choice for the client is A_1 and the expected utility for the customer is zero, as the OEM extracts all the consumer's surplus.

4.2. Sensitivity analysis

For better understanding on how sensitive the model is regarding parameters and repair assumptions, we present a sensitivity analysis. Note that in the tables presented throughout this analysis, base results are highlighted in grey.

4.2.1. Effect of *q* variation

First, in order to identify possible output changes, we disregard the MLE value found for q in the previous Section, and vary the rejuvenation parameter q as shown in Table 4. As seen in Section 2.1, the rejuvenation parameter measures the quality of repair. When q approaches zero, the quality of repair increases, returning the failed unit almost to "as good as new". Thus, the wear of equipment is reduced, and it is reasonable to expect a lower failure rate. As q increases in turn the quality of repair decreases, and equipment suffers higher wear over time. Queue length, overtime ($T_j^{dt} W_i$) and amount of penalty incurred rise under option A_1, making the OEM serves fewer clients, which reduces their profit. For $q \geq 0.16$, the increased number of failures and amount of penalties are so high so that result in a change of strategy.

Table 4. Optimal solution changes due to q variations

q	M	P^{max} \$	C_s^{max} \$	A^*	$E[\pi]$ \$	$E\left[N_j\right]$	$E\left[T_j^{dt}\right]$	$E\left[T_j^{ov}\right]$
0.0	108	185,591	15,785	A_1	9,281,009	14.09	124.11	31.64
0.1	47	183,991	8,550	A_1	2,979,171	21.51	153.24	28.47
0.15	34	183,407	7,222	A_1	1,816,625	25.08	168.29	28.19
0.16	57	203,265	6,887	A_2	1,679,855	25.60	232.35	62.65
0.2	49	203,694	6,233	A_2	1,219,007	28.11	245.85	63.04
0.3	29	194,319	5,189	A_2	500,567	34.12	246.42	46.99

4.2.2. Variation on the model parameters

Generally, consumers with high risk-aversion are less willing to pay for the services, reducing OEM's profit. In this situation, the OEM tends to sell equipment to fewer customers. Table 5 shows how the players' optimal strategies change for different risk-aversion parameter values. For $\delta \leq 0.08$, strategy A_2 is chosen since customers become more tolerant to risk, thus not choosing EW anymore. For this case, OEM can perform maintenance for a greater number of hospitals, thus increasing waiting time in queue and downtime. However, profit is still increased, since customers are not as risk-averse as in the other tested cases, accepting to pay more for services. For $\delta > 0.08$, the hospital's optimal strategy is to hire the EW (A_1).

The variations on results due to changes in the device characteristic life α are given in Table 6. Higher values of α result in longer times to failures, which in turn decreases the expected number of failures $E\left[N_j\right]$. Consequently, devices have increased availability and generated revenue, also increasing the OEM's profits. Yet, the number of sold units M increases along with α, and due to the occurrence of fewer failures, customers pay considerably more for on-call repairs, since the total value they are willing to pay for repair services is now spread across fewer payments. P^{max} also increases with α increment, however at a lowest rate, because while devices fail less frequently, the OEM serves more devices as α increases, which causes more time spent in queues. For $\alpha \leq 1{,}140$, strategy A_2 is selected, due to increased failure frequency and, consequently, higher amount of penalties to the OEM.

Changes in optimal strategies due to variations in the shape parameter β can be seen in Table 7. For higher values of β, equipment wears

Table 5. Optimal solution changes due to δ variations

δ	M	P^{max} \$	C_s^{max} \$	A^*	$E[\pi]$ \$	$E\left[N_j\right]$	$E\left[T_j^{dt}\right]$	$E\left[T_j^{ov}\right]$
05	90	247,750	10,143	A_2	4,655,854	21.34	254.98	94.56
0.08	82	215,350	8,969	A_2	3,349,710	21.39	226.87	73.95
0.09	48	185,928	8,796	A_1	3,047,552	21.51	154.68	29.21
0.10	47	183,991	8,550	A_1	2,979,171	21.51	153.24	28.47
0.15	45	178,207	7,543	A_1	2,746,489	21.51	150.48	27.09
0.20	44	174,883	7,003	A_1	2,612,681	21.51	149.12	26.42

Table 6. Optimal solution changes due to α variations

α (h)	M	P^{max} \$	C_s^{max} \$	A^*	$E[\pi]$ \$	$E[N_j]$	$E[T_j^{dt}]$	$E[T_j^{ov}]$
1,100	49	201,301	6,648	A_2	1,244,617	28.29	241.34	59.40
1,140	53	201,513	6,951	A_2	1,485,647	26.95	235.11	59.75
1,150	31	183,076	7,101	A_1	1,531,656	26.75	176.50	28.61
1,200	35	183,493	7,463	A_1	1,857,136	25.25	170.36	28.77
1,300	43	183,963	8,168	A_1	2,574,340	22.66	158.96	28.75
1,351.83	47	183,991	8,550	A_1	2,979,171	21.51	153.24	28.47
1,400	51	184,093	8,854	A_1	3,373,132	20.53	148.59	28.37
1,500	60	184,386	9,545	A_1	4,244,186	18.73	140.40	28.40

Table 7. Optimal solution changes due to β variations

β	M	P^{max} \$	C_s^{max} \$	A^*	$E[\pi]$ \$	$E[N_j]$	$E[T_j^{dt}]$	$E[T_j^{ov}]$
1.5	58	184,520	9,247	A_1	3,945,706	19.60	146.41	29.37
1.6	51	184,329	8,791	A_1	3,296,711	20.82	151.21	29.01
1.658	47	183,991	8,550	A_1	2,979,171	21.51	153.24	28.47
1.7	45	184,073	8,369	A_1	2,770,587	22.00	155.66	28.60
1.74	43	184,033	8,188	A_1	2,588,444	22.46	157.59	28.54
1.75	72	204,856	8,019	A_2	2,569,419	22.46	221.40	66.84
1.8	69	205,154	7,808	A_2	2,454,541	23.03	224.62	66.48

Table 8. Optimal solution changes due to μ variations

μ (h^{-1})	M	P^{max} \$	C_s^{max} \$	A^*	$E[\pi]$ \$	$E[N_j]$	$E[T_j^{dt}]$	$E[T_j^{ov}]$
0.15	46	209,088	8,484	A_2	1,350,270	21.38	232.33	77.10
0.175	60	204,565	8,450	A_2	2,024,136	21.40	217.29	66.91
0.19	42	185.088	8.536	A_1	2540.077	21.50	157.04	30.39
0.2	47	183,991	8,550	A_1	2,979,171	21.51	153.24	28.47
0.225	60	181,767	8,508	A_1	4,149,651	21.52	145.25	24.65
0.25	73	179,841	8,525	A_1	5,408,147	21.53	138.11	21.44

out faster, resulting in higher number of failures $E[N_j]$, which in turn causes greater unavailability. Thus, the probability of penalty being incurred also increases, and the OEM chooses to serve fewer customers. In fact, as $\beta \geq \beta \geq 1.75$, EW is no longer advantageous, and A_2 becomes the optimal strategy; notice that the significant increase in M for $\beta \geq 1.75$ occurs because the OEM pays no penalty at all for strategy A_2, which allows it for serving more customers. The increase in P^{max} when $\beta \geq 1.75$ is due to the number of customers being served, which increases the likelihood of overtime, and the customers' willingness to pay for services. Notice in Eq. (14) that P^{max} increases along with the expected overtime in repairing the device $E[T_j^{ov}]$, which is higher as more customers are served, resulting in longer waiting times in queue.

Table 8 presents how optimal strategies change due to variation in the service rate μ. With higher values of μ, the service crew can repair a greater number of failed devices per time unit, decreasing the queue size. Therefore, the OEM can deal with a greater number of customers, increasing its own profit. On the other hand, when μ decreases, the service crew can repair fewer items per time unit. Notice that for $\mu \leq 0.175$, the service crew is too slow to repair the equipment by τ, resulting in a large amount of penalty under the option A_1. Then, the OEM's optimal strategy changes to set $P > P^{max}$ and $C_s = C_s^{max}$, which induces the customer to choose A_2, i.e., on-call services.

Table 9 shows how the optimal solution changes due to the maximum times to repair τ, which influences how long equipment waits (in queue or being repaired) without penalizing the OEM. The higher the value of τ, less penalty is incurred, because the frequency of those delays in returning the equipment to operational state will be lower. In other words, the OEM's payoff increases as τ increases. When the acceptable tolerance τ is smaller, overtime and penalty increase under EW option. Then, for values of $\tau \leq 10$, it is best for the manufacturer to sell on-call repairs.

Table 10 presents the behavior of optimal solution due to revenue R variations. Notice that with the increase of monetary incomes to the hospital, they are willing to pay considerably more for each

Table 9. Optimal solution changes due to τ variations

τ (h)	M	P^{max} $	C_s^{max} $	A^*	$E[\pi]$ $	$E[N_j]$	$E[T_j^{dt}]$	$E[T_j^{ov}]$
8	66	218,036	8,729	A_2	2,252,960	21.45	185.89	74.04
10	74	211,672	8,553	A_2	2,535,263	21.42	204.35	72.04
11	45	187,130	8,590	A_1	2,721,277	21.51	150.48	31.25
12	47	183,991	8,550	A_1	2,979,171	21.51	153.24	28.47
14	51	178,860	8,461	A_1	3,461,705	21.50	159.11	24.06
16	55	174,856	8,369	A_1	3,900,469	21.49	165.46	20.81

Table 10. Optimal solution changes due to R variations

R ($ 10^3/h)	M	P^{max} $	C_s^{max} $	A^*	$E[\pi]$ $	$E[N_j]$	$E[T_j^{dt}]$	$E[T_j^{ov}]$
0.093	43	164,882	7,846	A_1	2,191,878	21.52	147.80	25.77
0.0935	45	174,420	8,197	A_1	2,576,855	21.51	150.48	27.09
0.094	47	183,991	8,550	A_1	2,979,171	21.51	153.24	28.47
0.0945	49	193,620	8,885	A_1	3,397,473	21.50	156.13	29.95
0.095	51	203,276	9,206	A_1	3,830,324	21.50	159.11	31.51

option. Since hospitals generate more revenue with operation of their devices, they become willing to pay significantly more for maintenance services. Notice, however, that hospital's decision did not change with variations in R.

Finally, Table 11 and Table 12 show the changes in optimal solution due to variations on penalty parameters θ_1 and θ_2 respectively. Even though θ and τ both influence the total amount of penalties, their effects are distinct. High values for θ_1 increases the importance of repairing the equipment by τ, since delays will be severely penalized. For smaller values of θ_1, the OEM is able to serve a greater number of hospitals due to the decrease in penalties. However, when the number of hospitals increases, the penalty during W_2 also increases (since the reduction in penalty rate is only for W_1, in this case), which may cause strategy A_2 to be selected. This happens for $\theta_1 \leq 0.7$. In the case of θ_2, when it rises too much, strategy A_2 is chosen, since penalties during W_2 increase considerably. This occurs for $\theta_2 \geq 3.3$.

Table 11. Optimal solution changes due to θ_1 variations

θ_1($ 10^3 / h)	M	P^{max} $	C_s^{max} $	A^*	$E[\pi]$ $	$E[N_j]$	$E[T_j^{dt}]$	$E[T_j^{ov}]$
0.5	87	208,257	8,130	A_2	3,354,136	21.36	243.67	86.11
0.7	85	208,439	8,227	A_2	3,079,027	21.37	236.67	80.98
0.75	48	183,288	8,471	A_1	3,034,574	21.51	154.68	29.21
1	47	183,991	8,550	A_1	2,979,171	21.51	153.24	28.47
1.25	46	184,484	8,553	A_1	2,914,144	21.51	151.84	27.77
1.5	46	185,253	8,593	A_1	2,844,846	21.51	151.85	27.77

Table 12. Optimal solution changes due to θ_2 variations

θ_2 ($ 10^3 / h)	M	P^{max} $	C_s^{max} $	A^*	$E[\pi]$ $	$E[N_j]$	$E[T_j^{dt}]$	$E[T_j^{ov}]$
1	74	187,862	8,434	A_1	5,604,879	21.42	204.35	58.55
2	58	186,192	8,464	A_1	3,972,368	21.48	170.56	37.78
3	47	183,991	8,550	A_1	2,979,171	21.51	153.24	28.47
3.2	46	183,978	8,516	A_1	2,823,874	21.51	151.84	27.77
3.3	74	201,541	8,444	A_2	2,766,223	21.42	204.36	58.40
4	74	203,169	8,436	A_2	2,758,993	21.42	204.36	58.56
5	74	204,831	8,436	A_2	2,758,993	21.42	204.36	58.56

5. Concluding remarks

In this paper, a decision model for an Extended Warranty involving hospitals and OEM was proposed. For modelling this situation and determining the players' optimal strategies, a SG formulation was employed, with the OEM being the leader and the hospital the follower. This situation is commonly found in the market of technology-intensive equipment, which is characterized by a greater bargaining power for the manufacturer, which is the only part capable of performing maintenance interventions adequately.

In order to approximate the problem to a more realistic context, we considered the equipment is subject to imperfect repairs, and to model this issue, two approaches were joined: GRP and queueing theory. Additionally, an application example was presented with real failure data of an angiograph to determine the optimal strategies for each player and demonstrate applicability of the model. Furthermore, we perform a series of sensitivity analyses by showing how model results and players' strategies behave under different scenarios.

Some limitations of the presented approach may also be pointed out. In real-world situations, customers do not present homogeneous risk behavior, consequently, different customers often choose different strategies. Also, different agents commonly have access to different levels of information (asymmetric information), so that it is difficult to predict the actions of other players. Based on these limitations, and also intending to extend the present model, the following features could be implemented:

- Consideration of information asymmetry by employing a principal-agent formulation (Jiang et al. [15], Jin et al. [16]).
- Analysis of consumer usage rate, along with the definition of a two-dimensional warranty policy (Yang et al. [40]).
- A dynamic SG with a greater time horizon to analyze the possibilities for renewal of the extended warranty, analyzing the behavior of the players during longer periods.
- Incorporation of a heterogeneous market, with different profiles of customers, represented by different risk-aversion parameters.

Acknowledgments

The authors would like to thank the National Council for Scientific and Technological Development – CNPq and the Foundation for the Support of Science and Technology of the State of Pernambuco – FACEPE for the financial support through research grants.

References

1. Amir R. Stackelberg versus Cournot Equilibrium 1999; 21: 1–21.
2. Ashgarizadeh E, Murthy D N P. Service contracts: A stochastic model. Mathematical and Computer Modelling 2000; 31(10–12): 11–20, https://doi.org/10.1016/S0895-7177(00)00068-6.
3. Bouguerra S, Chelbi A, Rezg N. A decision model for adopting an extended warranty under different maintenance policies. International Journal of Production Economics 2012, https://doi.org/10.1016/j.ijpe.2011.10.022.
4. Çimen S, Gooya A, Grass M, Frangi A F. Reconstruction of coronary arteries from X-ray angiography: A review. Medical Image Analysis 2016; 32, https://doi.org/10.1016/j.media.2016.02.007.
5. Cruz A M, Rincon A M R. Medical device maintenance outsourcing: Have operation management research and management theories forgotten the medical engineering community? A mapping review. European Journal of Operational Research 2012; 221(1): 186–197, https://doi.org/10.1016/j.ejor.2012.02.043.
6. Darghouth M N, Ait-kadi D, Chelbi A. Joint optimization of design, warranty and price for products sold with maintenance service contracts. Reliability Engineering and System Safety 2017; 165: 197–208, https://doi.org/10.1016/j.ress.2017.03.033.
7. Ding S H, Kamaruddin S. Maintenance policy optimization—literature review and directions. International Journal of Advanced Manufacturing Technology 2014; 76 (5–8): 1263–1283, https://doi.org/10.1007/s00170-014-6341-2.
8. Esmaeili M, Shamsi Gamchi N, Asgharizadeh E. Three-level warranty service contract among manufacturer, agent and customer: A game-theoretical approach. European Journal of Operational Research 2014; 239 (1): 177–186, https://doi.org/10.1016/j.ejor.2014.04.045.
9. Ferreira R J, Firmino P R A, Cristino C T. A mixed Kijima model using the Weibull-based Generalized Renewal Processes. PLoS ONE 2015; 10(7): 1–17, https://doi.org/10.1371/journal.pone.0133772.
10. Forgó F, Szép J, Szidarovszky F. Introduction to the Theory of Games: Concepts, Methods, Applications. Springer 1999.
11. Gibbons R. A Primer in Game Theory. A Primer in Game Theory 1992, https://doi.org/10.1017/CBO9780511791307.017.
12. Guo R, Ascher H, Love E. Toward Pratical and Synthetical Modelling of Repairable Systems. Economic Quality Control 2001; 16(2): 147–182.
13. Huang Y, Huang C, Ho J. A customized two-dimensional extended warranty with preventive maintenance. European Journal of Operational Research 2017; 257(3): 971–978, https://doi.org/10.1016/j.ejor.2016.07.034.
14. Husniah H, Pasaribu U S, Iskandar B P. Service contract management with availability improvement and cost reduction. ARPN Journal of Engineering and Applied Sciences 2015; 10(1): 146–151.
15. Jiang H, Pang Z, Savin S. Performance-Based Contracts for Outpatient Medical Services. Manufacturing & Service Operations Management 2012; 14(4): 654–669, https://doi.org/10.1287/msom.1120.0402.
16. Jin T, Tian Z, Xie M. A game-theoretical approach for optimizing maintenance, spares and service capacity in performance contracting. International Journal of Production Economics 2015; 161: 31–43, https://doi.org/10.1016/j.ijpe.2014.11.010.
17. Kijima M, Sumita U. A Useful Generalization of Renewal Theory : Counting Processes Governed by Non-Negative Markovian Increments. Journal of Applied Probability 1986; 23(1): 71–88, https://doi.org/10.2307/3214117.
18. Kim S-H, Cohen M a., Netessine S, Veeraraghavan S. Contracting for Infrequent Restoration and Recovery of Mission-Critical Systems. Management Science 2010; 56(9): 1551–1567, https://doi.org/10.1287/mnsc.1100.1193.
19. Lins I D, Droguett E L. Redundancy allocation problems considering systems with imperfect repairs using multi-objective genetic algorithms and discrete event simulation. Simulation Modelling Practice and Theory 2011; 19(1): 362–381, https://doi.org/10.1016/j.simpat.2010.07.010.
20. Markets and Markets. Medical Equipment Maintenance Market by Device (Imaging (MRI, CT, PET, SPECT, X-ray, Ultrasound, C Arm), Endoscopy, ECG, EMG, Defibrillator, Dental, Ventilator), Service Provider (OEM, ISO, In house), End User (Public, Private) - Global Forecast to 2022. 2017. Retrieved from http://www.marketsandmarkets.com/Market-Reports/medical-equipment-maintenance-market-69695102.html

21. Marsaro M F, Cavalcante C A V. Random Preventive Maintenance Policy Based on Inspection for a Multicomponent System Using Simulation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19(4): 552–559, https://doi.org/http://dx.doi.org/10.17531/ein.2017.4.8.
22. Moura M C, Droguett E L, Alves Firmino P R, Ferreira R J. A competing risk model for dependent and imperfect condition-based preventive and corrective maintenances. Proceedings of the Institution of Mechanical Engineers Part O-Journal of Risk and Reliability 2014; 228(6): 590–605, https://doi.org/10.1177/1748006X14540878.
23. Moura M C, Santana J M, Droguett E L, Lins I D, Guedes B N. Analysis of extended warranties for medical equipment: A Stackelberg game model using priority queues. Reliability Engineering & System Safety 2017; 168: 338–354, https://doi.org/10.1016/j.ress.2017.05.040.
24. Murthy D N P, Asgharizadeh E. Optimal decision making in a maintenance service operation. European Journal of Operational Research 1999; 116(2): 259–273,https://doi.org/10.1016/S0377-2217(98)90202-8.
25. Murthy D N P, Asgharizadeh E. a Stochastic Model for Service Contract. International Journal of Reliability, Quality and Safety Engineering 1998; 5(1): 29–45.
26. Murthy D N P, Djamaludin I. New product warranty: A literature review. International Journal of Production Economics 2002; 79(3): 231–260, https://doi.org/10.1016/S0925-5273(02)00153-6.
27. Murthy D N P, Jack N. Extended Warranties, Maintenance Service and Lease Contracts, https://doi.org/10.1007/978-1-4471-6440-1
28. Murthy D N P, Yeung V. Modelling and analysis of maintenance service contracts. Mathematical and Computer Modelling 1995; 22(10–12): 219–225, https://doi.org/10.1016/0895-7177(95)00199-C.
29. Oliveira C C F de, Cristino C T, Firmino P R A. In the kernel of modelling repairable systems: a goodness of fit test for Weibull-based generalized renewal processes. Journal of Cleaner Production 2016; 133: 358–367, https://doi.org/10.1016/j.jclepro.2016.05.123.
30. Osborne M J, Rubinstein A. A Course in Game Theory. New York: The MIT Press, 1994.
31. Pham H, Wang H. Imperfect maintenance. European Journal of Operational Research 1996; 94(3): 425-438, https://doi.org/10.1016/S0377-2217(96)00099-9.
32. Rahman A, Chattopadhyay G. Review of long-term warranty policies. Asia-Pacific Journal of Operational Research 2006; 23(4): 453–472, https://doi.org/10.1142/S021759590600108X.
33. Rinsaka K, Sandoh H. A stochastic model on an additional warranty service contract. Computers and Mathematics with Applications 2006; 51(2): 179–188, https://doi.org/10.1016/j.camwa.2005.11.024.
34. Ross S M. Introduction to Probability Models. Academic Press 2014.
35. Tanwar M, Rai R N, Bolia N. Imperfect repair modeling using Kijima type generalized renewal process. Reliability Engineering and System Safety 2014; 124: 24–31, https://doi.org/10.1016/j.ress.2013.10.007.
36. Varian H R. Microeconomic Analysis. New York: W. W. Norton & Company, Inc, 1992.
37. Wang H, Pham H. Reliability and Optimal Maintenance. Risk Management. Springer-Verlag London 1996, https://doi.org/10.1007/b138077
38. Wang Z M, Yang J G. Numerical method for Weibull generalized renewal process and its applications in reliability analysis of NC machine tools. Computers and Industrial Engineering 2012; 63(4): 1128-1134, https://doi.org/10.1016/j.cie.2012.06.019.
39. Yañez M, Joglar F, Modarres M. Generalized renewal process for analysis of repairable systems with limited failure experience. Reliability Engineering and System Safety 2002; 77(2): 167–180, https://doi.org/10.1016/S0951-8320(02)00044-3.
40. Yang D, He Z, He S. Warranty claims forecasting based on a general imperfect repair model considering usage rate. Reliability Engineering & System Safety 2016; 145: 147–154, https://doi.org/10.1016/j.ress.2015.09.012.
41. Ye Z S, Murthy D N P. Warranty menu design for a two-dimensional warranty. Reliability Engineering and System Safety 2016; 155: 21–29, https://doi.org/10.1016/j.ress.2016.05.013.
42. Zio E. The Monte Carlo Simulation Method for System Reliability and Risk Analysis. Principles of Loads and Failure Mechanisms-Applications in Maintanence, Reliabilty and Design 2013, https://doi.org/10.1007/978-1-4471-4588-2.

João Mateus Marques de SANTANA
Rafael Luiz Velôzo SANTIAGO
Márcio das Chagas MOURA
Isis Didier LINS
Center for Risk Analysis and Environmental Modeling – CEERMA
Department of Production Engineering
Universidade Federal de Pernambuco – UFPE
Av. Prof. Moraes Rego, 1235 – University City
Recife – PE – Brazil – 50670-901

Emails: joaomateusmsantana@gmail.com, rafa.velozo@hotmail.com, marcio@ceerma.org, isis.lins@ceerma.org

Article citation info:
ZHANG Y, MA Y, OUYANG L, LIU L. A novel reliability model for multi-component systems subject to multiple dependent competing risks with degradation rate acceleration. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 579–589, http://dx.doi.org/10.17531/ein.2018.4.9.

Yanjing ZHANG
Yizhong MA
Linhan OUYANG
Lijun LIU

A NOVEL RELIABILITY MODEL FOR MULTI-COMPONENT SYSTEMS SUBJECT TO MULTIPLE DEPENDENT COMPETING RISKS WITH DEGRADATION RATE ACCELERATION

NOWATORSKI MODEL NIEZAWODNOŚCI DLA SYSTEMÓW WIELOELEMENTOWYCH NARAŻONYCH NA LICZNE ZALEŻNE RYZYKA KONKURUJĄCE UWZGLĘDNIAJĄCY PRZYSPIESZENIE TEMPA DEGRADACJI

The purpose of this paper is to establish a new reliability model of the system subject to multiple dependent competing risks. For a system subject to multiple dependent competing risks, the total degradation consists of natural degradation amount and sudden degradation increments (SDIs) caused by random shocks arriving at the system. Most researchers on this topic only focus on the SDIs. However, the impact of random shocks on degradation rate is ignored. In this paper, a novel reliability model considering degradation rate acceleration (DRA) caused by random shocks is proposed, in which the degradation model is based on the degradation path. The dependence relationship between multiple degradation processes is dealt with by copula method, and the arrival time of shocks is assumed to follow a non-homogeneous Poisson process (NHPP). Finally, the effectiveness of the proposed reliability model is demonstrated by an example of a series system. Moreover, the effect of model parameters is evaluated through sensitivity analysis.

Keywords: *dependent competing risks; degradation rate acceleration; reliability model; copula method; sensitivity analysis.*

Celem niniejszej pracy było stworzenie nowego modelu niezawodności systemu narażonego na liczne zależne ryzyka konkurujące. W przypadku systemu eksponowanego na wiele zależnych ryzyk konkurujących, na wartość całkowitą degradacji składa się wartość degradacji naturalnej oraz wartość nagłych przyrostów degradacji (sudden degradation increments, SDI) powodowanych przez losowe zaburzenia systemu. Większość badaczy tej tematyki koncentruje się wyłącznie na SDI, ignorując tym samym wpływ zaburzeń losowych na tempo degradacji. W niniejszym artykule zaproponowano nowy model niezawodności uwzględniający przyspieszenie tempa degradacji powodowane zaburzeniami losowymi, w którym model degradacji opiera się na krzywej degradacji. Zależność między mnogimi procesami degradacji rozpatrywano za pomocą metody funkcji kopuły przy założeniu, że czas wystąpienia zaburzenia odpowiada niejednorodnemu procesowi Poissona. Skuteczność proponowanego modelu niezawodności zademonstrowano na przykładzie systemu szeregowego. Ponadto, wykorzystano analizę czułości do oceny wpływu parametrów modelu na niezawodność systemu.

Słowa kluczowe: *zależne ryzyka konkurujące; przyspieszenie tempa degradacji; model niezawodności; metoda funkcji kopuły; analiza czułości.*

1. Introduction

The failure of a system is usually caused by internal degradation or external random shocks. The failure caused by internal degradation is called soft failure, such as erosion, fatigue, wear, etc. And the hard failure is caused by external random shocks, such as device breakdown, short circuit, etc. The degradation processes for components in a system and a shock process arriving at the system compete with each other. The occurrence of any failure mode may lead to the failure of systems. A system deteriorates with its use and age, which is a continuous accumulation of degradation. However, the hard failure may or may not happen in its life cycle.

In many studies [10, 14, 28, 32], the competing risks were treated as independent. However, the natural degradation processes of a system are usually affected by the shock loads. That is, the same shock arriving at a system will influence the degradation process of each component. Hence, the assumption of independence between competing risks is not reasonable, and it may cause underestimation or overestimation of the system reliability. It is very necessary to consider the dependence relationship between multiple degradation processes and a shock process when establishing the reliability model for a system.

Recently, some researchers [5, 15] have taken the dependence relationship into account to develop the reliability model of a system. Peng and Feng [20] built a reliability model for the system subject to multiple dependent competing risks, where dependent competing risks referred to soft failure and hard failure. Soft failure was caused by continuous natural degradation and additional SDI due to random shocks. Hard failure was induced by fatal shock loads from the shock process. Jiang and Feng [11] proposed a reliability model for a sys-

Table 1. Comparisons with existing reliability models

Models	DBDS[1]	SDIs[2]	CL[3]	DRA[4]	NDP[5]	ATS[6]
Peng, Feng and Coit (2010)	Y	Y			1	HPP
Guo, Wang and Guo (2013)	Y	Y			2	HPP
Song, Coit, and Feng (2014)	Y	Y			*m*	HPP
Jiang, Feng, and Coit (2015)	Y	Y	Y		1	HPP
Wang and Pham (2012)	Y	Y		Y	*m*	HPP
An and Sun (2017)	Y	Y	Y		*m*	HPP
Huynh, Castro, and Barros (2012)	N				1	NHPP
Bocchetti, Giorgio, and Guida (2009)	N				1	NHPP
Proposed model	Y	Y	Y	Y	*m*	NHPP

1 Dependence between degradation and shocks.
2 Sudden degradation increments.
3 Certain level.
4 Degradation rate acceleration.
5 Number of degradation processes.
6 Arrival time of shocks.

tem subject to multiple dependent competing risks. In their research, the shock threshold may shift due to exposure to various shock patterns. Guo and Wang [8] developed a joint copula reliability model for systems subject to two degradation processes and a random shock process. Song and Coit [24] developed a reliability model for a multi-component system. They thought a system may fail due to any soft failure process or hard failure process. Meanwhile, the soft failure processes of different components in a system were mutually competing. Besides, Song and Coit [25] also established a reliability model for multi-component systems subject to dependent competing risks of natural degradation and random shocks, in which the shocks were categorized according to their sizes, function, etc.

Previous researches have mainly investigated the dependence relationship between multiple degradation processes and a shock process based on an assumption that the arrival of each shock only causes SDIs. Nevertheless, the assumption is not always reasonable for systems with high reliability and long life because they have the ability to resist small shock loads. The research by Jiang and Feng [12] manifested that small shock loads had no effect on the degradation process, which was supposed to be a gamma process. In addition, Wang and Pham [29] considered two types of shocks when evaluating the system reliability: fatal shock loads leading to the hard failure of systems and general shock loads increasing the system degradation level. Moreover, a novel reliability model was proposed by An and Sun [1] for highly reliable systems experiencing multiple dependent competing risks. They assumed that only shock loads above a certain level can affect degradation processes. In their study, shock loads were separated into three parts by the shock threshold and the certain level. The first part of shocks, which were above the hard failure threshold, were fatal shocks causing sudden failure. The shocks in the second part were general shocks, which were between the certain level and shock threshold. Only general shocks can cause SDIs. The rest of shocks were small shocks, which were supposed to have no effect on degradation processes. However, almost all the researchers were used to considering the SDIs of degradation processes and ignoring the impact of general shock loads on the degradation rate. What is more, in most previous studies about the random shock process, the arrival of shocks was assumed to be a homogeneous Poisson process (HPP) [30]. But this assumption is not always appropriate for systems which work in the convoluted environment. Under a complex circumstance, the occurrence rate of shocks is a variable rather than a constant. Therefore, the assumption that the arrival of shocks is a NHPP may be more reasonable [2, 9].

Motivated by the above, we propose a new reliability model to evaluate the reliability of the system subject to multiple dependent competing risks. In the model, the general shock loads can bring about DRA in degradation processes, besides SDIs. In addition, only shock loads above a certain level can impact the degradation processes. The dependence between a shock process and multiple degradation processes is dealt with by conditional probability. In addition, the dependence relationship between multiple degradation processes is solved by a copula method. The arrival of shocks is assumed to be a NHPP. A summary of the comparison with existing reliability models is presented in Table 1. The remainder of this paper is organized as follows: a system subject to multiple dependent competing risks is described in Section 2. In Section 3, we develop a new reliability model in which DRA is considered into the degradation path. In Section 4, a numerical example is presented to validate the proposed reliability model. Finally, some concluding remarks are given in Section 5.

2. System description

In fact, systems with multiple components used in industrial applications always experience two types of failures: soft failure and hard failure. Either of them may lead to the failure of systems. Assume that there are i components in a system and every component experiences a degradation process. Once a shock arriving at the system, it will affect all the degradation processes. If there are no random shocks, the ith natural degradation process corresponding to the ith component of the system is shown in Figure 1, which is a continuous increasing process [3]. $D_i(t)$ represents the wear volume of the ith component of the system at t. The threshold of soft failure of the ith component is denoted by $l^{(i)}$.

As shown in Figure 2, any random shock arriving at the system, whose load is above the threshold of hard failure W_U, may cause the failure of components in the system. Meanwhile, general shocks, whose loads are between the failure threshold W_U and a certain level W_L, could give rise to the SDI of every component once they arrive at the system. Actually, the sudden degradation increments on different components caused by the same shock may be different due to the different material properties of components. In addition, small shocks, whose loads are below the certain level W_L, have no effect on degradation processes of components due to the system's good performance. It indicates that the degradation process and the shock process are dependent because every degradation process is affected by the same shock process.

From Figure 3, we can see that the total degradation of the ith component in a system is composed of natural degradation and SDIs, which is denoted by Y_{ij} (i=1, 2, …, m; j=1, 2, …, ∞). Y_{ij} represents the SDIs of the ith component of a system caused by the jth shock arriving at the system. The ith component occurs soft failure once the cumulative degradation of any component exceeds its soft failure threshold $l^{(i)}$. The total degradation $M^{(i)}(t)$ (i=1, 2, …, m) of the ith component exceeds the degradation threshold $l^{(i)}$ or the magnitude of a shock exceeds the shock threshold W_U, which will result in the failure of the component.

With the development of science and technology, the reliability of modern products has been greatly improved. The failure occurs rarely in a long working time. However, many products degrade over time before they fail or break down. Thus, in many engineering reliability experiments, the measure of degradation can be observed over a period of time before failure occurs to provide additional information of systems. There are many studies to overcome this kind of difficulty. For highly-reliable or long-life modern products, it often takes much more time to obtain lifetime and degradation data under usual use conditions. One solution is to use accelerated degradation tests to collect the performance degradation data at greater environmental stress levels so that the degradation data can be analyzed earlier before any specimens "fail" [6, 16, 17, 19, 21, 22, and 31]. For the convenience to exhibit reliability model, a summary of the main notations used in this paper is presented in Table 2.

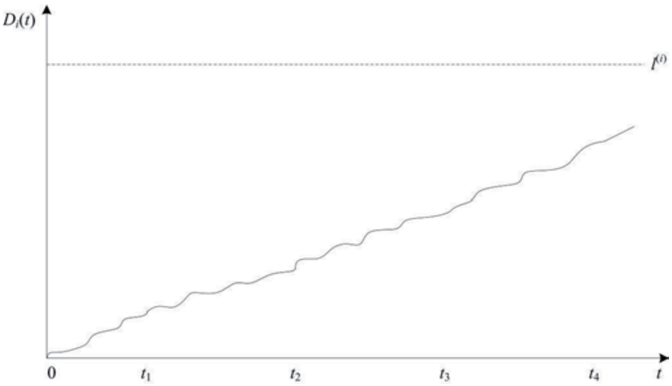


Fig. 1. Continuous degradation of the ith component

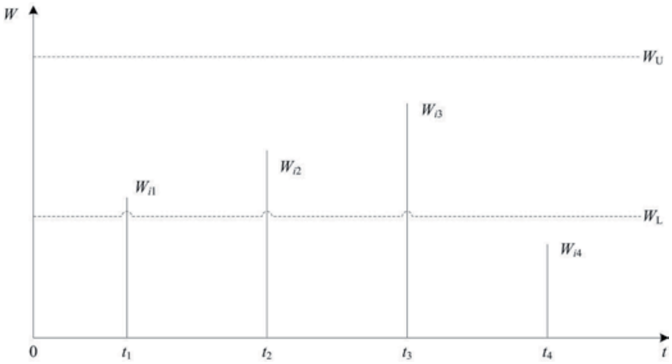


Fig. 2. Magnitude of every shock arriving at the system imposed on the ith component

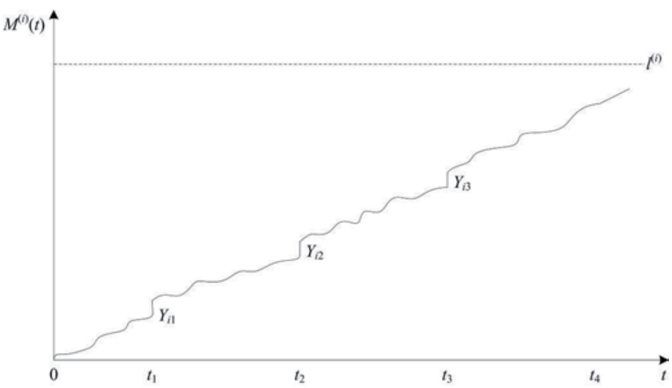


Fig. 3. Total degradation of the ith component

Table 2. List of symbols and definitions

Symbol	Definition
$N_1(t)$	Number of general shocks at time t
$N_2(t)$	Number of fatal shocks at time t
W_U	Hard failure threshold for the shock process
W_L	A certain level only above which the sudden degradation increments occurs
W_{ij}	Magnitude of the jth general shock on the ith component
$\lambda(t)$	Intensity function of non-homogeneous Poisson process at time t
p_1	Probability of general shock
p_2	Probability of fatal shock
$D_i(t)$	Natural degradation at time t for the ith component
$S_i(t)$	Cumulative sudden degradation increments of the ith component caused by general shocks at time t
$M^{(i)}(t)$	Cumulative degradation of the ith component at time t
Y_{ij}	Sudden degradation increments of the ith component caused by the jth general shock
$\gamma_1^{(i)}$	Coefficient of general shock number on the ith component
$\gamma_2^{(i)}$	Coefficient of cumulative sudden degradation increments on the ith component
$l^{(i)}$	Soft failure threshold of the ith component
$T^{(i)}$	Failure time of the ith component
$R_M^{(i)}(t)$	Marginal reliability function of the ith component at time t
$R(t)$	System reliability function at time t
C	Copula function

3. Proposed reliability model with degradation rate acceleration

Recently, many reliability models have been developed to estimate the reliability of systems which are subject to multiple dependent competing risks. Song and David [24-25] considered that any shock load on the system increases the amount of degradation. However, the research by Tanner and Walraven [27] showed that no SDIs in micro-engine when shock loads are below a certain level. Thus, for systems with high reliability and long life, only those shocks, whose loads are above a certain level, can increase systems' degradation. However, general shocks may result in an increase in degradation rate as well as SDIs. Based on the analysis above, a new reliability model is proposed to contribute to the improvement of reliability evaluation. In the reliability model, DRA caused by general shock loads is paid attention to. Random shocks are divided into three parts: fatal shocks, general shocks and small shocks. In addition, the arrival time of random shocks is assumed to follow a NHPP instead of HPP.

3.1. Reliability analysis for shock process

Random shocks are likely to be introduced from the external environment. Most shocks are harmful to systems, decreasing residual useful life of systems. Even, some shocks may immediately cause the failure of the system.

As shown in Figure 2, for a system with m components, any shock arriving at the system will affect all the m components. Take the ith component for example, the magnitude of shock load imposed on it is denoted by W_{ij} (i=1, 2, …, m; j=1, 2, …, ∞), which is caused by jth shock reaching the system. The component fails once W_{ij} exceeds the shock threshold W_U. Assume W_{ij} imposed on the ith component of the system is independent and identically distributed ($i.i.d$) random variable. Then the cumulative distribution function of W_{ij} can be denoted by $F_{W_{ij}}(w)$, then the probability that the hard failure of the ith component does not occur under the jth shock is:

$$P(W_{ij} < W_U) = F_{W_{ij}}(W_U),\ i = 1,2,\ldots,m;\ j = 1,2,\ldots,\infty \tag{1}$$

To simplify the calculation process, W_{ij} for the ith component is assumed to follow a normal distribution $N(\mu_{W_i}, \sigma_{W_i}^2)$. Then Equation (1) can be expressed by:

$$P(W_{ij} < W_U) = \phi\left(\frac{W_U - \mu_{W_i}}{\sigma_{W_i}}\right),\ i = 1,2,\ldots,m;\ j = 1,2,\ldots,\infty \tag{2}$$

3.2. Reliability analysis for degradation process

As is shown in Figure 3, the soft failure of the ith degradation process happens when total degradation $M^{(i)}(t)$ exceeds its degradation threshold $l^{(i)}$. The total degradation $M^{(i)}(t)$ includes continuous natural degradation amount and SDIs.

3.2.1. Analysis of SDI

It is assumed that the probabilities of shocks occurring at different time intervals are independent. Simultaneously, these random shocks are supposed to occur in a NHPP with an intensity function:

$$\lambda(t) = re^{ct},\ r \in (0,\infty),\ c \in (-\infty,+\infty) \tag{3}$$

Let $N(t)$ represent the number of random shocks until t. Then the numbers of general shocks and fatal shocks can be denoted by $N_1(t)$ and $N_2(t)$, respectively. The probability of general shocks is $p_1=P(W_L<W_{ij}<W_U)$. And the probability of fatal shocks can be calculated as $p_2=P(W_{ij}>W_U)$. According to the decomposition method of Poisson process, the arrival time of general shocks follows a NHPP with an intensity function:

$$\lambda_1(t) = p_1\lambda(t) = p_1 re^{ct},\ r \in (0,\infty),\ c \in (-\infty,+\infty) \tag{4}$$

Similarly, the arrival time of fatal shocks also follows a NHPP with an intensity function:

$$\lambda_2(t) = p_2\lambda(t) = p_2 re^{ct},\ r \in (0,\infty),\ c \in (-\infty,+\infty) \tag{5}$$

Then the numbers of arrivals for general shocks and fatal shocks at time t are given by:

$$W_1(t) = E\left[N_1(t)\right] = \int_0^t p_1 re^{cs}ds = \begin{cases} p_1 \cdot \dfrac{r}{c}(e^{ct}-1),\ c \neq 0 \\ p_1 \cdot rt,\ c = 0 \end{cases} \tag{6}$$

$$W_2(t) = E\left[N_2(t)\right] = \int_0^t p_2 re^{cs}ds = \begin{cases} p_2 \cdot \dfrac{r}{c}(e^{ct}-1),\ c \neq 0 \\ p_2 \cdot rt,\ c = 0 \end{cases} \tag{7}$$

Therefore, the probability of n general shocks and n fatal shocks reaching the system are calculated by Equation (8) and Equation (9), respectively:

$$P(N_1(t) = n) = \frac{(W_1(t))^n}{n!}e^{-W_1(t)} \tag{8}$$

$$P(N_2(t) = n) = \frac{(W_2(t))^n}{n!}e^{-W_2(t)} \tag{9}$$

General shocks usually cause additional damages to degradation processes. Here, we utilize Y_{ij} (i=1, 2, …, m; j=1, 2, …, ∞) to represent the SDIs in ith degradation process caused by the jth general shock arriving at the system. What is more, Y_{ij} greatly depends on the magnitude of shock load W_{ij} (i=1, 2, …, m; j=1, 2, …, ∞). Y_{ij} is described by a linear function, which reflects the difference between the magnitude of shock load W_{ij} and the certain level W_L, namely:

$$Y_{ij} = b(W_{ij} - W_L) \tag{10}$$

where b is a constant, which indicates the SDIs in degradation process caused by a unit change in the magnitude of shock load. Then, the cumulative degradation caused by general shocks at time t can be written as:

$$S_i(t) = \begin{cases} \sum_{j=1}^{N_1(t)} Y_{ij},\ N_1(t) > 0 \\ 0,\ N_1(t) = 0 \end{cases} \tag{11}$$

3.2.2. Analysis of continuous natural degradation

Consider a system with unknown soft failure threshold, which is a fixed quantity. The system is placed under an environment with accelerated stress levels, which is steadily increased until the failure of the system. Assume the increasing loads are converted to discrete values so the stress is incremented by small, discrete amounts until the system fails. And each small increment of stress causes a nonnegative damage amount, which is a random variable and denoted by X. According to [7, 13, 23], the cumulative damage after n+1 increments of stress is denoted by:

$$D_{n+1} = D_n + X_n \cdot \eta(D_n) \tag{12}$$

where $\eta(\mu)$ is the damage model function. For example, $\eta(\mu)$=1 gives an additive damage model, whereas $\eta(\mu)=\mu$ gives a multiplicative damage model. In fact, the stress is incremented by continuous amounts, so the cumulative damage of the system at t should be:

$$D(t) = X \cdot \eta(t) \tag{13}$$

In this paper, a multiplicative path function is used to reflect the item-to-item variation. Then the degradation model of the ith degradation process for a series system with m components is given by:

$$D_i(t) = X_i \cdot \eta_i(t) \tag{14}$$

where X_i is a random variable, $\eta_i(t)$ represents the mean degradation path in the ith degradation process, which is either monotonically decreasing or monotonically increasing.

The cumulative degradation $M^{(i)}(t)$ at time t for the ith degradation process consists of natural degradation and the SDIs caused by the general shocks, which is written as:

$$M^{(i)}(t) = D_i(t) + S_i(t) \tag{15}$$

The first term $D_i(t)$ refers to the continuous natural degradation of the ith degradation process. And the second term $S_i(t)$ shown in Equation (11) reflects the SDIs in the ith degradation process.

In this section, DRA is considered into the degradation model because the general shocks may accelerate the internal clock of systems. To embody the effect of DRA, a time-scaled covariate factor is used. A new term $G(t,\gamma^{(i)})$ is introduced into $D_i(t)$ for the ith degradation process through the time-scaled model of accelerated life testing. Here, the ith degradation path $D_i(t)$ is scaled by an accelerated factor from t to $te^{G(t,\gamma(i))}$. Thus, Equation (15) can be rewritten as:

$$M^{(i)}(t) = X_i \eta_i(te^{G(t,\gamma^{(i)})}) + \sum_{j=1}^{N_1(t)} Y_{ij} \tag{16}$$

where $G(t,\gamma^{(i)}) = \gamma_1^{(i)} N_1(t) + \gamma_2^{(i)} \sum_{j=1}^{N_1(t)} Y_{ij}$, and the vector parameters $\gamma^{(i)}$ are unknown. Note that the first term in function $G(t,\gamma^{(i)})$ embodies the effect from the number of general shocks towards the ith degradation process. In general, we have $\gamma^{(i)} \geq 0$, and the first term reflects the fact that the degradation rate likely increases with the increase of the number of general shocks. If $\gamma_1^{(i)} = 0$, it signifies that the degradation rate does not be affected by the number of general shocks. The second term is developed to present the situation that the cumulative SDIs may give rise to an accelerated degradation rate of systems. Likewise, it indicates that the cumulative SDIs have no effect on the degradation rate of systems if $\gamma_2^{(i)} = 0$.

3.3. Reliability modeling for systems with a shock process and multiple degradation processes

3.3.1. System reliability model

For a series system with m components, which is subject to a shock process and m degradation processes, the measurements of m degradation processes at time t are denoted by $M(t)=\{M^{(1)}(t), M^{(2)}(t), \ldots, M^{(m)}(t)\}$. The system fails once any degradation process reaches its soft failure threshold. The soft failure thresholds corresponding to the m degradation processes are denoted by $\boldsymbol{L}=\{l^{(1)}, l^{(2)}, \ldots, l^{(m)}\}$. Meanwhile, the hard failure occurs once any fatal shock arrives at the system. Hence, only when there is no fatal shock and the degradation amount of each degradation process keeps below its soft failure threshold, the system is in the working state. Let $T^{(i)}$ be the time to failure for the ith degradation process. Then, the reliability of the series system subject to a shock process and m degradation processes at time t can be expressed by:

$$\begin{aligned} R(t) &= P[T^{(1)} > t, T^{(2)} > t, \ldots, T^{(m)} > t] \times P[N_2(t) = 0] \\ &= P[M^{(1)} < l^{(1)}, M^{(2)} < l^{(2)}, \ldots, M^{(m)} < l^{(m)}] \times P[N_2(t) = 0] \end{aligned} \tag{17}$$

If the m degradation processes are assumed to be independent, the reliability of the system presented in Equation (17) can be rewritten as:

$$R(t) = R_M^{(1)}(t) \times R_M^{(2)}(t) \times \ldots \times R_M^{(m)}(t) \times P[N_2(t) = 0] \tag{18}$$

where $R_M^{(m)}(t)$ indicates the marginal reliability of the mth degradation process at t. However, Equation (18) can't provide precise estimation of system reliability because these degradation processes are not independent with each other. Therefore, the dependence relationship between any two degradation processes should be paid attention to.

Similarly, for a parallel system with m components, which is subject to m degradation processes and a shock process, the reliability of this system is calculated as:

$$\begin{aligned} R(t) &= 1 - [1 - P(T^{(1)} > t)] \times [1 - P(T^{(2)} > t)] \times \ldots \times [1 - P(T^{(m)} > t)] \times P[N_2(t) = 0] \\ &= 1 - [1 - P(M^{(1)} < l^{(1)})] \times [1 - P(M^{(2)} < l^{(2)})] \times \ldots \times [1 - P(M^{(m)} < l^{(m)})] \times P[N_2(t) = 0] \end{aligned} \tag{19}$$

From Equation (19), it is found that there is a dependence relationship between multiple degradation processes. Obviously, the reliabilities of other systems of any structure functions can be easily calculated according to their structures through the definition of reliability.

As we know, copula method is a powerful statistical tool to specify joint distribution if the known marginal distributions are complex. Thus, a copula method is utilized to develop the dependent structure among multiple degradation processes. In Sections 3.3.2 and 3.3.3, we use a series system as a representative system to represent the copula method.

3.3.2. Copula method for marginal reliability function of degradation processes

In probability theory and statistic, a copula is a multivariate probability distribution for which the marginal probability distribution of each variable is uniform. Copulas are used to describe the dependence structure between random variables. Any multivariate joint distribution can be written in terms of univariate marginal distribution functions [18]. Consider a random vector $(X_1, X_2, \ldots, X_m)$. Suppose its marginals are continuous, and the marginal CDFs $F_i(x)=P(X_i \leq x)$ are continuous functions. Sklar's theorem states that every multivariate cumulative distribution function $H(x_1, x_2, \ldots, x_m)=P(X_1 \leq x_1, \ldots, X_m \leq x_m)$ of the random vector $(X_1, X_2, \ldots, X_m)$ can be expressed in terms of its marginals $F_i(x)=P(X_i \leq x_i)$ and a copula C:

$$H(x_1, x_2, \ldots, x_m) = C(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)) \tag{20}$$

The copula C contains all information on the dependence structure between the components of $(X_1, X_2, \ldots, X_m)$, whereas the marginal cumulative distribution function $F_i(x_i)$ contains all information of the marginal distribution.

The CDF of the time to failure for the m degradation processes in a series system can be expressed as $F_i(t)=1-\mathrm{R}_i(t)$ $(i=1, 2, \ldots, m)$. The joint CDF of $T_1, T_2, \ldots, T_m$ is written as:

$$P(T_1 \leq t_1, T_2 \leq t_2, \ldots, T_m \leq t_m) = H(t_1, t_2, \ldots, t_m) = C(F_1(t_1), F_2(t_2), \ldots, F_m(t_m)) \tag{21}$$

Correspondingly, the marginal reliability for the system is expressed as:

$$P(T_1 > t_1, T_2 > t_2, \ldots, T_m > t_m) = \overline{H}(t_1, t_2, \ldots, t_m) = C(R_1(t_1), R_2(t_2), \ldots, R_m(t_m)) \tag{22}$$

The relationship between $C(F_1(t_1), F_2(t_2), \ldots, F_m(t_m))$ and $C(R_1(t_1), R_2(t_2), \ldots, R_m(t_m))$ is represented as:

$$C(R_1(t_1), R_2(t_2), \ldots, R_m(t_m)) = 1 - \sum_{i=1}^{m} F_i(t_i) + \sum_{1 \le i < h \le m} C(F_i(t_i), F_h(t_h), \ldots)$$
$$- \sum_{1 \le i \le h < k \le m} C(F_i(t_i), F_h(t_h), F_k(t_k), \ldots) + (-1)^m C(F_1(t_1), F_2(t_2), \ldots, F_m(t_m)) \tag{23}$$

Therefore, the marginal reliability of the series system subject to m degradation processes at time t is expressed as:

$$\begin{aligned} R(t) &= P(T_1 > t, T_2 > t, \ldots, T_m > t) \\ &= P(M^{(1)}(t) < l^{(1)}, M^{(2)}(t) < l^{(2)}, \ldots, M^{(m)}(t) < l^{(m)}) \\ &= C(R_M^{(1)}(t), R_M^{(2)}(t), \ldots, R_M^{(i)}(t)) \end{aligned} \tag{24}$$

where $R_M^{(i)}(t)$ denotes the marginal reliability function for the ith degradation process at time t.

In particular, based on Equation (23) and Equation (24), the marginal reliability function for a system subject to two degradation processes is computed by:

$$\begin{aligned} P(M^{(1)}(t) < l^{(1)}, M^{(2)}(t) < l^{(2)}) &= C(R_M^{(1)}(t), R_M^{(2)}(t)) \\ &= 1 - F_M^{(1)}(t) - F_M^{(2)}(t) + C(F_M^{(1)}(t), F_M^{(2)}(t)) \\ &= R_M^{(1)}(t) + R_M^{(2)}(t) - 1 + C(u, v) \end{aligned} \tag{25}$$

where $u = F_M^{(1)}(t) = 1 - R_M^{(1)}(t),\ v = F_M^{(2)}(t) = 1 - R_M^{(2)}(t)$.

3.3.3. Calculation steps for the system reliability

In this paper, a maximum likelihood estimation (MLE) is introduced to complete the statistical inference for copula. In stage 1, we should calculate marginal reliability function for the ith degradation process. The system does not fail due to the ith degradation process until t only if the cumulative degradation is below its soft failure threshold $l^{(i)}$ conditioned on the event that there is no fatal shock. Hence, the marginal reliability function for the ith degradation process at time t is given by:

$$R_M^{(i)}(t) = P(M^{(i)}(t) < l^{(i)}) = \sum_{n=0}^{\infty} P(X_i \cdot \eta_i(te^{G(t,\gamma^{(i)})}) + S_i(t) < l^{(i)} \mid N_1(t) = n) P(N_1(t) = n) \tag{26}$$

Suppose that random variables W_{ij} (i=1, 2, …, m; j=1, 2, …, ∞) are $i.i.$d. Based on Equation (10) and (11), these random variables Y_{ij} (i=1, 2, …, m; j=1, 2, …, ∞) are also $i.i.$d. In this section, Y_{ij} (i=1, 2, …, m; j=1, 2, …, ∞) are assumed to follow a common distribution $Q_i(x)$ (i=1, 2, …, m) in the ith degradation process. Let $Q_i^{(j)}(z) = P(Y_{i1} + Y_{i2} + \ldots + Y_{ij} \le z)$ for j=1, 2, …, ∞. The marginal reliability function for the ith degradation process at time t can be rewritten as:

$$R_M^{(i)}(t) = P(X_i \cdot \eta_i(t) < l^{(i)}) P(N_1(t) = 0) + \sum_{n=1}^{\infty} P(N_1(t) = n) \times \int_{z=0}^{l^{(i)}} P(X_i \cdot \eta_i(te^{\gamma_1^{(i)} n + \gamma_2^{(i)} z}) + z < l^{(i)}) dQ_i^{(n)}(z)$$
$$= \exp(-\int_0^t \lambda_1(s)ds) \times F_{X_i}\left(\frac{l^{(i)}}{\eta_i(t)}\right) + \sum_{n=1}^{\infty} \frac{\exp(-\int_0^t \lambda_1(s)ds) \times (\int_0^t \lambda_1(s)ds)^n}{n!} \times \int_{z=0}^{l^{(i)}} F_{X_i}\left(\frac{l^{(i)} - z}{\eta_1(te^{\gamma_1^{(i)} n + \gamma_2^{(i)} z})}\right) dQ_i^{(n)}(z) \tag{27}$$

where the number of general shocks is denoted by n, and $S_i(t)$ is replaced by z. Assume $Q_i(x)$ follows a normal distribution with mean μ_{Yi} and variance $\sigma_{Y_i}^2$. Then, $Q_i^{(n)}(z)$ also follows a normal distribution $N(n\mu_{Y_i}, n\sigma_{Y_i}^2)$. Then, the marginal reliability probability $R_M^{(i)}(t)$ of the ith degradation process at discrete time points can be calculated by Equation (27). When m=2, we can obtain a group of data of u and v at the corresponding time t.

In stage 2, we firstly estimate parameters in the copula $C(F_1(t_1), F_2(t_2), \ldots, F_m(t_m))$ using the values of u and v through the Copulafit function provided by MATLAB. A summary of the common copulas for two variates is presented in Table 3. Then, a likelihood criterion is used to check the goodness of fitting and specify the optimal copula. The likelihood function of the bivariate copula is given by:

$$L = \sum_{k=1}^{m_0} \ln c(F_M^{(1)}(t_k), F_M^{(2)}(t_k) \mid \alpha) \tag{28}$$

where m_0 represents the number of discrete time points, and α is the parameter in copulas. The copula with the biggest likelihood value is the most suitable copula, which can be used to describe the dependence between any two degradation processes.

Finally, the probability that the system will not fail due to degradation processes is calculated according to Equation (24). Thus, the specific formulation of the system reliability becomes:

$$R(t) = C(R_M^{(1)}(t), R_M^{(2)}(t), \ldots, R_M^{(m)}(t)) P[N_2(t) = 0] \tag{29}$$

If m=2, the reliability of the system at time t is derived as:

$$R(t) = [R_M^{(1)}(t) + R_M^{(2)}(t) - 1 + C(u, v)] \exp(-\int_0^t \lambda_2(s)ds) \tag{30}$$

where $\lambda_2(t)$ is shown in Equation (5). Thus, the simplified steps to calculate the system reliability with copula approach are shown in Table 4.

4. Numerical example

Microelectromechanical systems (MEMS) oscillators are timing devices that generate highly stable reference frequencies to sequence electronic systems, manage data transfer and measure elapsed time. MEMS oscillators vibrate at their natural resonant frequency. Due to the working loss of operation, the mass of MEMS oscillators decreases after a period of time. The decrease of mass may cause an increase in the frequency of vibration, which is an obvious common phenomenon that exists in MEMS oscillators. On the other hand, thermal shock, jitter and other vibration from the environment can bring about the change of system frequency. If these shocks or vibration are large enough, hard failure will occur to the MEMS oscillators. MEMS oscillators are particularly interesting and typical systems which are subject to multiple degradation processes and a random shock process [26]. Besides, these processes are dependent and compete with each other.

4.1. Reliability modeling

In this example, the reliability model of microelectromechanical systems (MEMS) will be developed. Assume MEMS oscillators subject to two degradation processes and a shock process operate in unstable environment. The arrival time of shocks follows a NHPP with an intensity function $\lambda(t)$. Meanwhile, for the first degradation path, we set $D_1(t; X_1) = X_1 t$, in which X_1 is Weibull distributed with a CDF $F_{X_1}(x) = 1 - \exp[-(x/\mu)^k]$, μ, k>0. For the second degradation path, suppose that $D_2(t; X_2)=X_2\log[\theta_1+t]$, θ_1>0, in which X_2 is gamma distributed with a pdf $f_{X_2}(x) = x^{\alpha-1}e^{-(x/\beta)} / \beta^{\alpha}\Gamma(\alpha)$, $x \geq 0$. The soft failure threshold of the two degradation processes are denoted by $l^{(1)}$ and $l^{(2)}$. The system fails once the total degradation of any degradation process reaches its soft failure threshold. In addition, the shock loads towards the two degradation processes are assumed to follow a normal distribution with mean μ_{W_i} and variance $\sigma_{W_i}^2$. The model parameters are presented in Table 5.

Substituting these parameters listed in Table 5 into Equation (27), we can obtain the marginal reliability functions of two degradation processes. Then, the two functions are plotted in Figure 4.

As shown in Figure 4, in the early stage of system operation, the marginal reliability of the system under the second degradation process decreases faster compared to the first degradation process. In the later period of the system operation, the marginal reliability of the system under the first degradation process is slowly approaching 0, while the marginal reliability of the system under the second degradation process drops rapidly to 0. The marginal reliability for the first degradation process becomes close to zero when the running time of the system is about 150 hours, while the marginal reliability for the second degradation process is almost 0 when the system works for 75 hours. The values of marginal reliability for two degradation processes at some discrete time points are given in Table 6.

According to the properties of copula function, we can utilize the marginal reliability function to obtain the joint reliability function for degradation processes by using Copulafit function provided by

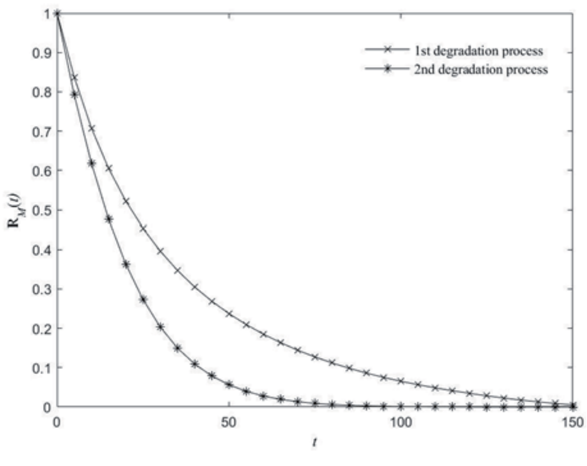


Fig. 4. Marginal reliability function for two degradation processes

Table 3. Commonly used copulas

Copula	C (u, v|α)	α∈Ω
Gauss	$\int_{-\infty}^{\phi^{-1}(u)}\int_{-\infty}^{\phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-a^2}}\exp\left(\frac{2asw-s^2-w^2}{2(1-a^2)}\right)dsdw$	α∈[-1,1]
t	$\int_{-\infty}^{T_v^{-1}(u)}\int_{-\infty}^{T_v^{-1}(v)}\frac{1}{2\pi\sqrt{1-a^2}}\left(1+\frac{s^2+w^2-2asw}{v(1-a^2)}\right)^{\left(-\frac{v+2}{2}\right)}dsdw$	α∈[-1,1]
Frank	$-\frac{1}{\alpha}\ln\left(1+\frac{(e^{-au}-1)(e^{-av}-1)}{e^{-a}-1}\right)$	α∈(-∞, ∞)/{0}
Gumbel	$\exp\left\{-[(-\ln u)^{1/\alpha}+(-\ln v)^{1/\alpha}]^{\alpha}\right\}$	α∈(0, ∞)
Clayton	$(u^{-a}+v^{-a}-1)^{-1/\alpha}$	α∈(0, ∞)

Table 4. The calculation steps of system reliability

I Marginal Reliability Function for Degradation Processes

1. Calculate the marginal reliability function for each degradation process.
2. Calculate the marginal reliability probability for each degradation process at discrete time points.

II Joint Reliability Function for Degradation Processes

3. Estimate the parameters in different copulas based on the marginal reliability probability.
4. Find out the most suitable copula through MLE.
5. Calculate the joint copula $C(u, v)$, and then calculate the reliability of the system subject to two degradation processes.

Table 5. Parameters for MEMS oscillators subject to a shock process and two degradation processes

Parameter	Value	Source
μ_{Wi}	5	Wang et al. (2012)
σ_{Wi}	3	Wang et al. (2012)
$\gamma_1^{(1)}$	0.05	Wang et al. (2012)
$\gamma_2^{(1)}$	0.008	Wang et al. (2012)
$\gamma_1^{(2)}$	0.01	Wang et al. (2012)
$\gamma_2^{(2)}$	0.005	Wang et al. (2012)
$l^{(1)}$	100	Wang et al. (2012)
$l^{(2)}$	3.5	Wang et al. (2012)
W_U	8	Wang et al. (2012)
W_L	2	Assumption
b	0.6	Assumption
c	0.01	Guo et al. (2013)
r	0.1	Guo et al. (2013)
μ	0.8	Wang et al. (2012)
k	1	Wang et al. (2012)
θ	4.8	Wang et al. (2012)
α	20	Assumption
β	0.01	Wang et al. (2012)

Table 6. Simulated marginal reliability and values of u and v for two degradation processes

t	0	5	10	15	20	25	30	35	40
$R_M^{(1)}(t)$	1	0.8373	0.7078	0.6050	0.5222	0.4540	0.3965	0.3475	0.3054
$R_M^{(2)}(t)$	1	0.7923	0.6185	0.4767	0.3631	0.2734	0.2036	0.1500	0.1093
u	0	0.1627	0.2922	0.3950	0.4778	0.5460	0.6035	0.6525	0.6946
v	0	0.2077	0.3815	0.5233	0.6369	0.7266	0.7964	0.8500	0.8907

Table 7. Likelihood values and correlated parameters under five copulas

Copula	Likelihood value	Parameter
t	4.0099	α=0.9934, v=1.9724
Gaussian	5.9222	α=0.9930
Gumbel	1.1569	α=17.9540
Clayton	7.4402	α=13.7277
Frank	29.28	α=54.2398

Table 8. Values of C(u, v) at discrete time points under five copulas

Copula	$C(u, v)$ at t								
	0	5	10	15	20	25	30	35	40
t	0	0.1611	0.2907	0.3939	0.4771	0.5455	0.6032	0.6523	0.6945
Gaussian	0	0.1616	0.2919	0.3950	0.4778	0.5460	0.6035	0.6525	0.6946
Gumbel	0	0.1615	0.2920	0.3950	0.4778	0.5460	0.6035	0.6525	0.6946
Clayton	0	0.1623	0.2917	0.3944	0.4771	0.5452	0.6026	0.6514	0.6933
Frank	0	0.1612	0.2921	0.3950	0.4778	0.5460	0.6035	0.6525	0.6946

Table 9. Comparisons of system reliability under five copulas and independence assumption

Copula	$R(t)$								
	0	5	10	15	20	25	30	35	40
t	1	0.7289	0.5222	0.3679	0.2550	0.1739	0.1167	0.0770	0.0500
Gaussian	1	0.7294	0.5232	0.3687	0.2555	0.1742	0.1168	0.0771	0.0501
Gumbel	1	0.7293	0.5233	0.3687	0.2555	0.1742	0.1168	0.0771	0.0501
Clayton	1	0.7301	0.5230	0.3683	0.2550	0.1737	0.1163	0.0766	0.0495
Frank	1	0.7290	0.5234	0.3687	0.2555	0.1742	0.1168	0.0771	0.0501
Independence	1	0.6634	0.4378	0.2884	0.1896	0.1241	0.0807	0.0521	0.0334

MATLAB. In this article, t-copula, Gaussian copula, Gumbel copula, Clayton copula and Frank copula are used to fit the joint reliability function for two degradation processes. MLE is employed to estimate the parameters of different copulas based on the values of marginal failure functions, namely, u and v listed in Table 6. Thus, the likelihood values and parameter values under five kinds of copulas are listed in Table 7.

Through MATLAB, the values of five copula functions based on u and v in Table 6 are obtained and listed in Table 8. Then the reliabilities of the system shown in Table 9 at different discrete time points are calculated according to Equation (30). And the reliabilities of the system at some discrete time points when the two degradation processes are assumed to be independent are also listed in Table 9. From Table 9, it is noted that the system reliabilities at different time points under those five copulas are similar, while they are bigger than that when the two degradation processes are assumed to be independent. It indicates that the reliability of the series system will be underestimated if not considering the independence between multiple degradation processes.

We find from the five likelihood values in Table 7 that Frank copula is the most appropriate copula for fitting joint reliability for degradation processes of the system because it has the maximal likelihood estimation value 29.28. Therefore, the reliability of the system should be calculated under Frank copula. The system reliability functions based on dependence and independence assumption are plotted in Figure 5. From Figure 5, the system reliability decreases quickly when two degradation processes are assumed to be independent. An interesting finding is that the system reliability is relatively high under the assumption that the two degradation processes are dependent. It also indicates that the reliability of the series system will be underestimated when the two degradation processes are assumed to be independent. The finding has been clearly stated and explained in [4].

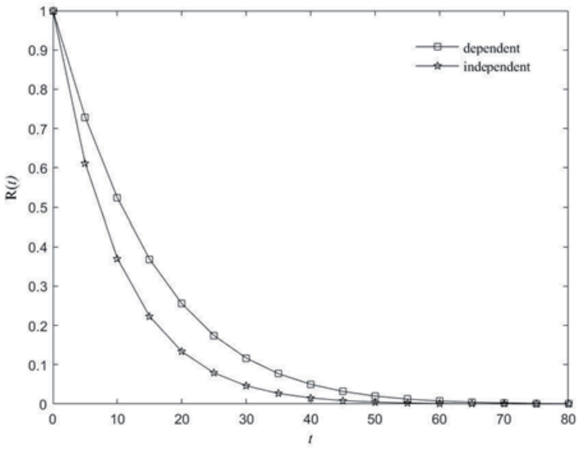


Fig. 5. System reliability function R(t)

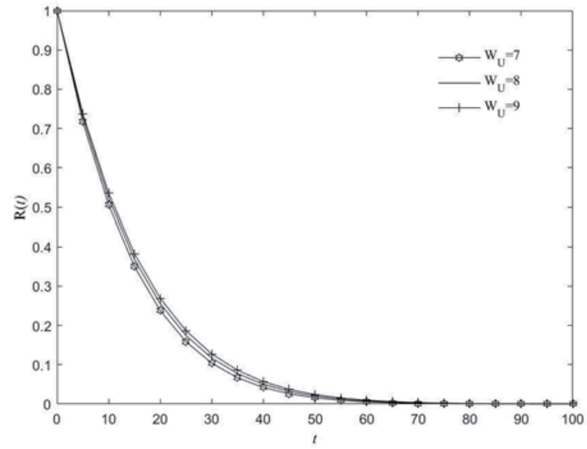


Fig. 6. Sensitivity analysis of R(t) on WU

Therefore, it is very essential and important for reliability engineers to take the dependence among multiple degradation processes into account.

4.2. Sensitivity analysis

In this section, sensitivity analysis is performed to estimate the effects of the model parameters W_U and W_L on the system reliability $R(t)$. The results are exhibited in Figure 6 and Figure 7. As shown in Figure 6, we can find that $R(t)$ is sensitive to the hard failure threshold W_U. When the hard failure threshold W_U increases from 7 to 9, the system reliability $R(t)$ shifts to right, which embodies a better reliability performance for a larger value of W_U. It indicates that the system has the capacity to resist external shocks as the improvement of system reliability. Moreover, as shown in Figure 7, the system reliability $R(t)$ shifts to the right when the certain level W_L increases from 1 to 3. It indicates that the bigger the certain level W_L, the better the performance of systems. The reason is that if the performance of the system is good enough, small shocks will not have any impact on the system. It is obvious that a system with a certain level 3 is more reliable than systems with a certain level 1 or 2. From the viewpoint of sensitivity analysis, we know that the accuracy of the hard failure threshold and the certain level is very important for the reliability of systems.

5. Conclusions

A new reliability model for systems subject to multiple dependent competing risks has been proposed. The DRA caused by random shocks into each degradation process is considered in the novel reliability model. The dependence relationship among different degradation processes has been dealt with by a copula method. Meanwhile, the shock process is assumed to be a NHPP instead of HPP. A numerical example is presented to verify the feasibility of the proposed model. Furthermore, a detailed analysis about model parameters has been done through sensitivity analysis. Based on the results of the numerical example, it can be concluded that the proposed reliability model is very suitable for complex systems subject to multiple dependent competing risks, and it can assess the system reliability more factually. In addition, three interesting research findings are obtained through the analysis of the reliability model:

(1) The reliabilities of series systems will be underestimated when multiple degradation processes are assumed to be independent. Therefore, it is very necessary to consider the dependence among different degradation processes.

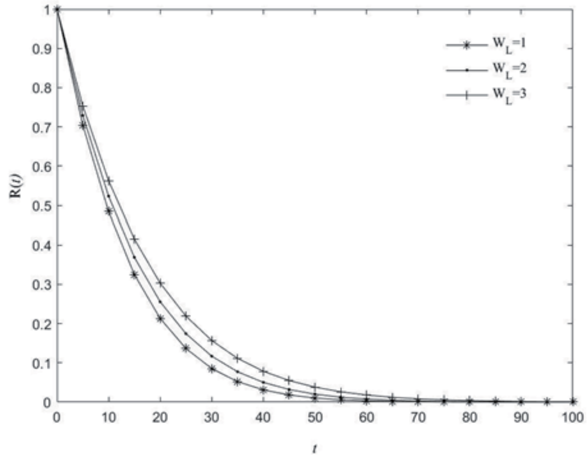


Fig. 7. Sensitivity analysis of R(t) on WL

(2) The assumption that the arrival time of random shocks follows a NHPP is more reasonable when a system works in complex environment.

(3) The series system reliability changes with the values of the hard failure threshold W_U and the certain level W_L. The greater the values of W_U and W_L, the higher the system reliability.

In the numerical example, for convenience, multiple degradation processes is simplified to two degradation processes. However, a system always experiences multiple different degradation processes when it is in a working state. Therefore, a challenging work is to calculate the reliability of systems subject to multiple degradation processes and a shock process. Furthermore, the aim of reliability assessment is to monitor systems real-timely and ensure the system reliability at a predetermined level by adopting some maintenance policies. Hence, the future work can be extended to develop a condition-based maintenance policy for systems subject to multiple competing risks.

Acknowledgements

The research work was supported by grants from the China Scholarship Council (201706840122), the National Natural Science Foundation of China (NSFC 71471088, 71702072, 71871119), and the Natural Science Foundation for Jiangsu Institutions (BK20170810), which are gratefully acknowledged.

References

1. An Z, Sun D. Reliability modeling for systems subject to multiple dependent competing failure processes with shock loads above a certain level. Reliability Engineering and System Safety 2017; 157: 129-138, https://doi.org/10.1016/j.ress.2016.08.025.
2. Bocchetti D, Giorgio M, Guida M. A competing risk model for the reliability of cylinder liners in marine diesel engines. Reliability Engineering and System Safety 2009; 94(8): 1299-1307, https://doi.org/10.1016/j.ress.2009.01.010.
3. Bae S J, Kuo W, Kvam P H. Degradation models and implied lifetime distributions. Reliability Engineering and System Safety 2007; 92(5), 601-608, https://doi.org/10.1016/j.ress.2006.02.002.
4. Barlow R E, Proschan F. Statistical theory of reliability and life testing: probability models. Florida: Florida State University Tallahassee, 1975.
5. Cha J H, Pulcini G. A dependent competing risks model for technological units subject to degradation phenomena and catastrophic failures. Quality and Reliability Engineering International 2016; 32(2), 505-517, https://doi.org/10.1002/qre.1767.
6. Durham S D, Padgett W J. Cumulative damage models for system failure with application to carbon fibers and composites. Technometrics 1997; 39(1): 34-44, https://doi.org/10.1080/00401706.1997.10485437.
7. Desmond A. Stochastic models of failure in random environments. Canadian Journal of Statistics 1985; 13(3): 171-183, https://doi.org/10.2307/3315148.
8. Guo C, Wang W, Guo B. Maintenance optimization for systems with dependent competing risks using a copula function. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2013; 15(1): 9-17.
9. Huynh K T, Castro I T, Barros A, Berenguer C. Modeling age-based maintenance strategies with minimal repairs for systems subject to competing failure modes due to degradation and shocks. European Journal of Operational Research 2012; 218(1): 140-151, https://doi.org/10.1016/j.ejor.2011.10.025.
10. Jiang R. Discrete competing risk model with application to modeling bus-motor failure data. Reliability Engineering and System Safety 2010; 95(9): 981-988, https://doi.org/10.1016/j.ress.2010.04.009.
11. Jiang L, Feng Q, Coit D W. Reliability and maintenance modeling for dependent competing failure processes with shifting failure thresholds. IEEE Transactions on Reliability 2012; 61(4): 932-948, https://doi.org/10.1109/TR.2012.2221016.
12. Jiang L, Feng Q, Coit D W. Modeling zoned shock effects on stochastic degradation in dependent failure processes. IIE Transactions 2015; 47(5): 460-470, https://doi.org/10.1080/0740817X.2014.955152.
13. Jiang C, Wang D. A sort of multiple life model under dependent causes of decrement. Statistical and Application 2015; 4(3): 169-175, https://doi.org/10.12677/SA.2015.43019.
14. Li W, Pham H. An inspection-maintenance model for systems with multiple competing processes. IEEE Transactions on Reliability 2005; 54(2): 318-327, https://doi.org/10.1109/TR.2005.847264.
15. Liu Z, Ma X, Shen L, Zhao Y. Degradation-shock-based reliability models for fault-tolerant systems. Quality and Reliability Engineering International 2016; 32(3), 949-955, https://doi.org/10.1002/qre.1805.
16. Mann N R, Schafer R E, and Singpurwalla N D. Methods for Statistical Analysis of Reliability and Life Data. New York: John Wiley & Sones, 1974.
17. Meeker W Q, Escobar L A. Statistical Methods for Reliability Data. New York: John Wiley & Sons, 1998.
18. Nelsen R B. An introduction to copulas. New York: Springer Science & Business Media, 2006.
19. Onar A, Padgett W J. Inverse Gaussian accelerated test models based on cumulative damage. Journal of Statistical Computation and Simulation 2000; 66(3): 233-247, https://doi.org/10.1080/00949650008812024.
20. Peng H, Feng Q, Coit D W. Reliability and maintenance modeling for systems subject to multiple dependent competing failure processes. IIE Transactions 2010; 43(1): 12-22, https://doi.org/10.1080/0740817X.2010.491502.
21. Padgett W J, Durham S D, and Mason A M. Weibull analysis of the strength of carbon fibers using linear and power law models for the length effect. Journal of Composite Materials 1995; 29(14): 1873-1884, https://doi.org/10.1177/002199839502901405.
22. Park C, Padgett W J. Stochastic degradation models with several accelerating variables. IEEE Transactions on Reliability 2006; 55(2): 379-390, https://doi.org/10.1109/TR.2006.874937.
23. Park C, Padgett W J. New cumulative damage models for failure using stochastic processes as initial damage. IEEE Transactions on Reliability 2005; 54(3): 530-540, https://doi.org/10.1109/TR.2005.853278.
24. Song S, Coit D W, Feng Q, Peng H. Reliability analysis for multi-component systems subject to multiple dependent competing failure processes. IEEE Transactions on Reliability 2014; 63(1): 331-345, https://doi.org/10.1109/TR.2014.2299693.
25. Song S, Coit D W, Feng Q. Reliability for systems of degrading components with distinct component shock sets. Reliability Engineering and System Safety 2014; 132: 115-124, https://doi.org/10.1016/j.ress.2014.06.020.
26. Song S, Coit D W, Feng Q. Reliability analysis of multiple-component series systems subject to hard and soft failures with dependent shock effects. IIE Transactions 2016; 48(8), 720-735, https://doi.org/10.1080/0740817X.2016.1140922.
27. Tanner D M, Walraven J A, Helgesen K, Irwin L W, Brown F, Smith N F, Masters N. MEMS reliability in shock environments. Proceedings of IEEE International Reliability Physics Symposium 2000; 129-138, https://doi.org/10.1109/RELPHY.2000.843903.
28. Wang Z, Huang H Z, Li Y. An approach to reliability assessment under degradation and shock process. IEEE Transactions on Reliability 2011; 60(4): 852-863, https://doi.org/10.1109/TR.2011.2170254.
29. Wang Y, Pham H. A multi-objective optimization of imperfect preventive maintenance policy for dependent competing risk systems with hidden failure. IEEE Transactions on Reliability 2011; 60(4): 770-781, https://doi.org/10.1109/TR.2011.2167779.
30. Wang Y, Pham H. Modeling the dependent competing risks with multiple degradation processes and random shock using time-varying copulas. IEEE Transactions on Reliability 2012; 61(1):13-22, https://doi.org/10.1109/TR.2011.2170253.
31. W. Nelson. Accelerated Testing: Statistical Models, Test Plans, and Data Analyses. New York: John Wiley & Sons, 1990, https://doi.org/10.1002/9780470316795.
32. Zhu Y, Elsayed E A, Liao H. Availability optimization of systems subject to competing risk. European Journal of Operational Research 2010; 202(3): 781-788, https://doi.org/10.1016/j.ejor.2009.06.008.

Yanjing ZHANG
School of Economics and Management
Nanjing University of Science and Technology
No. 200 Xiaolingwei, Xuanwu District, Nanjing, Jiangsu Province, China

Yizhong MA
School of Economics and Management
Nanjing University of Science and Technology
No. 200 Xiaolingwei, Xuanwu District, Nanjing, Jiangsu Province, China

Linhan OUYANG
College of Economics and Management
Nanjing University of Aeronautics and Astronautics
No. 29 General Avenue, Jiangning District, Nanjing, Jiangsu Province, China

Lijun LIU
School of Economics and Management
Nanjing University of Science and Technology
No. 200 Xiaolingwei, Xuanwu District, Nanjing, Jiangsu Province, China

E-mails: zhangyanjing_1211@163.com,yzma-2004@163.com, oylh871109@163.com, lijun_liu06@163.com

Article citation info:
TANG D, SHENG W, YU J. Dynamic condition-based maintenance policy for degrading systems described by a random-coefficient autoregressive model: A comparative study. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 590–601, http://dx.doi.org/10.17531/ein.2018.4.10.

Diyin TANG
Wubin SHENG
Jinsong YU

DYNAMIC CONDITION-BASED MAINTENANCE POLICY FOR DEGRADING SYSTEMS DESCRIBED BY A RANDOM-COEFFICIENT AUTOREGRESSIVE MODEL: A COMPARATIVE STUDY

DYNAMICZNA STRATEGIA UTRZYMANIA RUCHU NA PODSTAWIE STANU TECHNICZNEGO DLA ULEGAJĄCYCH DEGRADACJI SYSTEMÓW OPISANYCH MODELEM AUTOREGRESYJNYM Z PARAMETRAMI LOSOWYMI – STUDIUM PORÓWNAWCZE

In this paper, we optimize a dynamic condition-based maintenance policy for a slowly degrading system subject to soft failure and condition monitoring at equidistant, discrete time epochs. A random-coefficient autoregressive model with time effect is developed to describe the system degradation. The system age, previous state observations, and the item-to-item variability of the degradation are jointly combined in the proposed degradation model. Stochastic behavior for both the age-dependent and the state-dependent term are considered, and a Bayesian approach for periodically updating the estimates of the stochastic coefficients is developed to combine information from a degradation database with real-time condition-monitoring information. Based on this degradation model, the dynamic maintenance policy is formulated and solved in a semi-Markov decision process framework. Incorporated with the same semi-Markov decision process framework is a novel approach for mean residual life estimation, which enables simultaneous residual life estimation with the optimization procedure. The effectiveness of using the proposed random-coefficient autoregressive model with time effect rather than the existing fixed-coefficient ones to describe system degradation is demonstrated through a comparative study based on a real degradation dataset. The advantages of using a dynamic maintenance policy are also revealed.

Keywords: *degradation modeling, autoregressive model, Bayesian method, residual life estimation, semi-Markov decision process, condition-based maintenance.*

W prezentowanej pracy dokonano optymalizacji dynamicznej, uwzględniającej stan techniczny obiektu strategii utrzymania ruchu dla wolno ulegającego degradacji systemu monitorowanego w równoodległych dyskretnych chwilach czasu (epokach) pod względem uszkodzeń parametrycznych oraz stanu technicznego. Do opisu degradacji systemu opracowano model autoregresyjny z parametrami losowymi uwzględniający wpływ czasu. Proponowany model degradacji bierze pod uwagę zarówno wiek systemu jak i wcześniejsze obserwacje stanu oraz zmienność degradacji pomiędzy obiektami. Rozważano zachowanie stochastyczne zarówno składnika zależnego od wieku jak i składnika zależnego od stanu; opracowano bayesowską metodę okresowej aktualizacji oszacowań współczynników stochastycznych, która pozwala łączyć informacje z bazy danych o degradacji z informacjami z monitorowania stanu w czasie rzeczywistym. W oparciu o otrzymany model degradacji, sformułowano dynamiczną politykę utrzymania ruchu; problem optymalizacji tej polityki rozwiązywano w ramach procesu decyzyjnego semi-Markowa. Do procesu decyzyjnego włączono nowatorską metodę obliczania trwałości resztkowej, co umożliwiło ocenę trwałości resztkowej jednocześnie z przeprowadzeniem procedury optymalizacyjnej. Skuteczność wykorzystania proponowanego modelu autoregresyjnego do opisu degradacji systemu porównywano ze skutecznością dotychczasowych modeli z parametrami stałymi w badaniu opartym na rzeczywistym zbiorze danych o degradacji. Wskazano również zalety stosowania proponowanej dynamicznej strategii utrzymania ruchu.

Słowa kluczowe: *modelowanie degradacji, model autoregresyjny, metoda bayesowska, ocena trwałości resztkowej, semi-markowski proces decyzyjny, utrzymanie na podstawie stanu technicznego.*

1. Introduction

To sustain high reliability is the goal of every system and product. However, no matter how good the system design is, the performance of every system and product will ultimately deteriorate due to wear, fatigue, environmental conditions and other causes. When the deterioration becomes too severe, it may cause system malfunction or failure, which may result in significantly high maintenance costs and worst of all, safety hazards. Therefore, both operators and maintainers tend to adopt a preventive maintenance strategy to prevent system breakdowns, in that the preventive maintenance strategy usually performs before failures and thus it has a higher economic and safety significance than corrective maintenance which only takes place when the failure is observed.

Condition-based maintenance (CBM) is a kind of preventive maintenance strategy. It recommends maintenance actions based on the health status of the operating system. In a typical CBM policy, the health status of the system monitored throughout its operating life determines whether a preventive maintenance should be performed. Compared to traditional time-based preventive maintenance strategy, which sets a periodic interval to perform preventive maintenance regardless of the health status of a system, CBM is more reliable and cost-effective. Some successful examples of implementing CBM in real systems have demonstrated its efficiency in preventing catastrophic failures and improving maintenance performance (e.g., [2, 3, 15]).

To analyze and optimize a CBM policy for a specific system, the essential procedure is to develop a degradation model to describe the deterioration behavior of the operating system. The degradation model can be developed based on discrete-state or continuous-state stochastic processes, or their combinations. [13] and [11] considered a three-state continuous-time discrete-state Markov chain to model wear process of the diesel engine in locomotives and the gear shaft in gearboxes, respectively. To relax the Markovian assumption, [17] developed a degradation model based on a non-homogeneous semi-Markov process to de- scribe the deterioration of wear process in the turbofan engines. Using these Markovian degradation models, the CBM optimization problem can be trans- formed into determining maintenance actions for all system states with different maintenance objectives considered, e.g., [4, 14, 18, 19]. When the evolution of degradation state is continuous over time, continuous-state stochastic degradation models are more suitable. Gamma process (e.g., [26]), Wiener process(e.g.,[12, 21]), inverse Gaussian process (e.g.,[5]) are very popular models of this kind, which often allow simple and even elegant solutions to inference, hypothesis testing, goodness of fit, and prediction problems.

However, most previously developed continuous-state stochastic degradation models assume the degradation trend is only driven by system age and not by its previous states. Nevertheless, even if this assumption is appropriate for many types of degradation processes, in a general situation it is more realistic and appropriate to assume that the degradation trend can depend on either the system state or its age, or both. For example, the crack propagation rate might be higher if the current crack length is larger and a longer time has passed since the crack started propagation. In recent years, some researchers have noticed the need of effective modeling approaches for describing age- and state-dependent degradation processes. The first worth-noting contribution comes from [9]. In their proposed age- and state-dependent degradation model, the degradation increment over an elementary time interval has a discretized gamma distribution which depends on both the current degradation state and the operating age. Recently, [8,10] proposed a new class of Markovian age- and state-dependent degradation models, the transformed gamma process, by which the conditional distribution of the degradation growth over a generic time interval can be formulated in an analytical closed form.

The above models work very well in real applications (see examples in [9, 8, 10]), but they are only suitable to represent strictly monotonic degradation processes. It is not difficult to discover that in many industrial cases, effective description of non-monotonic degradation process, e.g., due to minimal repair, reduced load, or self-recovering mechanism [28], is also needed. Examples can be found in rotating machines [7], batteries [29], electronic devices [22, 23], etc. Therefore, [27] proposed an age- and state-dependent degradation model based on Wiener process to describe non-monotonic degradation processes. They obtained an analytical approximated residual life distribution to facilitate the residual life estimation of an operating system. However, the model did not consider the presence of observed heterogeneity among different individuals. Another age- and state-dependent degradation model capable of describing non-monotonic degradation processes was proposed by [23]. [23] took advantage of the state-dependency characteristic of autoregressive models and added an age-dependent term to the autoregressive model to include the influence of time. This model is easy to implement in real applications and mathematically tractable. However, the model formulation in [23] is also not able to describe the heterogeneity among different individuals.

Therefore, in this paper, we will improve the autoregressive model with time effect proposed in [23], in order to extend its capability of describing degradation processes. To do this, we will (i) consider a more general formation of the autoregressive model with time effect by assuming both the age-dependent term and the state-dependent term have stochastic behaviors; and (ii) derive a Bayesian updating method to update the model coefficients during system operation, which combines the information across the population and the information coming from the real-time condition monitoring(CM). The model coefficient updates have explicit formulas to allow fast computation in each update, which is an advantage of this model, and currently cannot be achieved by other existing age- and state-dependent degradation models. Using this model, the procedure of estimating the mean residual life in [23] is no longer applicable, due to the fact that the explicit form for the failure time distribution is quite complicated to obtain in mathematical point of view. Thus, this estimation task will be achieved via a Monte Carlo simulation procedure. We will demonstrate through a comparative study using the same dataset in [23] that the proposed model formation is superior and the Bayesian updating procedure indeed improves the accuracy of residual life estimation.

The proposed random-coefficient autoregressive model with time effect will then be applied to the optimization problem of a dynamic CBM policy. This maintenance policy is a commonly used control-limit policy in many industrial applications (e.g.,[13, 23]). Since we update model coefficients during system operation, we will consider the control-limit as a dynamic one, which is up-dated when new CM data becomes available. The optimization of this dynamic control-limit policy is achieved using a semi-Markov decision process(SMDP) framework. In this framework, we discover that the mean residual life of an operating system can be estimated simultaneously with the searching of the optimal control-limit. Therefore, it provides a novel idea of estimating the mean residual life for age- and state-dependent degradation models, and it also extends the application of the SMDP framework which is commonly considered as an approach only for policy decision problems. We will compare the mean residual life estimation results obtained by the SMDP-based approach with that by the Monte Carlo simulation approach, to reveal the advantage and disadvantage of the SMDP-based approach.

The rest of the paper is organized as follows. Section 2 introduces the general formation of the random-coefficient autoregressive model with time effect and the procedure to calculate the prior estimates of model coefficients. Section 3 develops a Bayesian updating framework to update the model coefficients for operating system. Section 4 describes the algorithms for optimizing the CBM policy and calculating the mean residual life for operating system. Section 5 gives numerical analysis. Conclusions and future research are given in Section 6.

2. The random-coefficient autoregressive model with time effect

Recall that the autoregressive model with time effect proposed in [23] deals with the degradation process whose degradation state can only be known at discrete inspection times. Suppose the degradation process starts from a known initial state Y_0, and is monitored through regular periodic inspections with inspection interval h. Let Y_1 denote the degradation state observed at inspection time $\{t\_n\} = nh, (n = 1, 2, 3, ...)$, then the autoregressive model with time effect has the following form defined as:

$$Y_n - \delta_0 = \beta t_n + \sum_{v=1}^{b} \varphi_v \left(Y_{n-v} - \delta_0\right) + \varepsilon_n$$
$$b = 1,2,3,...,n = 1,2,3,... \tag{1}$$

where b is the model order; δ_0 is a constant model coefficient; β and φ_v are also model coefficients; $\{\varepsilon_n\}$ are i.i.d. error terms and follow normal distribution $N\left(0,\sigma^2\right)$. To account for the heterogeneity among degradation paths of individual units, we consider a random-coefficient autoregressive model with time effect by supposing some (or all) of β and φ_v are possibly random.

[23] presented the procedure of choosing an appropriate model order b given historical data. In this paper, we will illustrate our method by considering the situation of $b = 1$. The other situations can be derived by the same procedure. For illustration purpose, we rewrite the random-coefficient autoregressive model with time effect as the following form:

$$Y_n - \delta_0 = \beta t_n + \varphi\left(Y_{n-1} - \delta_0\right) + \varepsilon_n,\ n = 1,2,3,... \tag{2}$$

From the model, we can observe that when $\varphi \neq 1$, the current state Y_n depends on previous state Y_{n-1} and age t_n. To account for the heterogeneity among different individuals, assume β follows normal distribution with mean μ_β and variance σ_β^2, φ follows normal distribution with mean μ_φ and variance σ_β^2, and they have mutual covariance ρ. Therefore, the model of Eq.(2) has the model coefficients $\gamma = \left(\delta_0, \mu_\beta, \sigma_\beta^2, \mu_\varphi, \sigma_\varphi^2, \rho, \sigma^2\right)$, which are unknown and need to be estimated.

To estimate the constant model coefficient δ_0, we set $\delta = \delta_0 - \varphi_1\delta_0$, then the model of Eq.(2) can be rewritten as:

$$Y_n = \delta + \beta t_n + \varphi Y_{n-1} + \varepsilon_n, n = 1,2,3,... \tag{3}$$

Assume that we have M histories of the system. For the sth data history, we denote the number of inspections by m_s, the observed system states by $\left\{y_0, y_1^s, y_2^s, ..., y_{m_s}^s\right\}$ where y_0 is the same for all histories, and the inspection times by $\left\{t_1^s, t_2^s, ..., t_{m_s}^s\right\}$, $s = 1,2,...,M$. So that for the M observed data histories, we have the regression representation $W = VA + E$, where:

$$\begin{aligned}
W' &= \left[y_{m_M}^M ... y_1^M ... y_{m_1}^1 ... y_1^1\right] \\
E' &= \left[\varepsilon_{m_M}^N ... \varepsilon_1^N ... \varepsilon_{m_1}^1 ... \varepsilon_1^1\right] \\
A' &= \left[\delta \quad \beta \quad \varphi\right] \\
V' &= \begin{bmatrix} 1 & ... & 1 & ... & 1 & ... & 1 \\ t_{m_M}^M & .... & t_1^M & ... & t_{m_M}^1 & ... & t_1^1 \\ y_{m_M-1}^M & ... & y_0 & ... & y_{m_M-1}^1 & ... & y_0 \end{bmatrix}
\end{aligned} \tag{4}$$

The least squares estimates for A is given by:

$$\hat{A} = \left(V'V\right)^{-1} V'W \tag{5}$$

Therefore,

$$\hat{\delta}_0 = \frac{\hat{\delta}}{1-\hat{\varphi}} \tag{6}$$

Let $X_n = Y_n - \delta_0$, $n = 1,2,3,...$, the model of Eq.(2) can be transformed into the following form:

$$\begin{aligned} X_0 &= Y_0 - \delta_0, \\ X_n &= \beta t_n + \varphi X_{n-1} + \varepsilon_n, n = 1,2,3,... \end{aligned} \tag{7}$$

Next, we have to estimate the distribution of β and φ. Let $\hat{\beta}_s$ and $\hat{\varphi}_s$ be the least squares estimates of parameter β and φ for each history s, $s = 1,2,...,M$, then we have $\mu_\beta, \sigma_\beta^2, \mu_\varphi, \sigma_\varphi^2, \rho$ estimated by:

$$\begin{aligned}
\hat{\mu}_\beta &= \frac{1}{M}\sum_{s=1}^{M}\hat{\beta}_s,\ \hat{\mu}_\varphi = \frac{1}{M}\sum_{s=1}^{M}\hat{\varphi}_s, \\
\hat{\sigma}_\beta^2 &= \frac{1}{M-1}\sum_{s=1}^{M}\left(\hat{\beta}_s - \hat{\mu}_\beta\right)^2, \hat{\sigma}_\varphi^2 = \frac{1}{M-1}\sum_{s=1}^{M}\left(\hat{\varphi}_s - \hat{\mu}_\varphi\right)^2, \\
\hat{\rho} &= \frac{1}{M-1}\sum_{s=1}^{M}\left(\hat{\beta}_s - \hat{\mu}_\beta\right)\left(\hat{\varphi}_s - \hat{\mu}_\beta\right).
\end{aligned} \tag{8}$$

Using the least squares estimates $\hat{\beta}_s$ and $\hat{\varphi}_s$ of parameter β and φ for each history s, $s = 1,2,...,M$, we are able to estimate the expected mean and stan- dard deviation of X_n^s conditioning on the initial degradation state X_0 for each data history, which are given by:

$$\begin{aligned}
E(X_n^s \mid X_0) &= \hat{\beta}_s \sum_{r=1}^{n} \hat{\varphi}^{n-r} t_r + \hat{\varphi}_s^n X_0, \\
Std(X_n^s \mid X_0) &= \sqrt{\sigma^2 \sum_{r=0}^{n-1} \hat{\varphi}_s^{2r}}.
\end{aligned} \tag{9}$$

Thus, let $x_n = y_n - \hat{\delta}_0$, the estimate of σ^2 is calculated using the following equation:

$$\hat{\sigma}^2 = \left(Q-1\right)^{-1} C'C, \tag{10}$$

where $Q = \sum_{s=1}^{M}\left(m_s - 1\right)$ is the total number of available observations of the degradation state, and:

$$C' = \sigma\left[\frac{x_{m_M}^M - E(X_{m_M}^M \mid X_0)}{Std(X_{m_M}^M \mid X_0)} \quad ... \quad \frac{x_1^M - E(X_1^M \mid X_0)}{Std(X_1^M \mid X_0)} \quad ... \quad \frac{x_1^1 - E(X_1^1 \mid X_0)}{Std(X_1^1 \mid X_0)}\right] \tag{11}$$

The least squares estimates $\hat{\beta}_s$ and $\hat{\varphi}_s$ for each history s, $s = 1,2,...,M$ are given by:

$$\hat{A}_s = \left(V_s'V_s\right)^{-1} V_s'W_s \tag{12}$$

where:

$$\begin{aligned}
W_s' &= \left[x_{m_s}^s ... x_1^s\right],\ E_s' = \left[\varepsilon_{m_s}^s ... \varepsilon_1^s\right],\ A_s' = \left[\beta_s \quad \varphi_s\right] \\
V_s' &= \begin{bmatrix} t_{m_s}^s & ... & t_1^s \\ x_{m_s-1}^s & ... & x_0 \end{bmatrix}
\end{aligned} \tag{13}$$

Using Eq.(4)-Eq.(13), the estimates of model coefficients $\hat{\gamma} = \left(\hat{\mu}_\beta, \hat{\sigma}_\beta^2, \hat{\mu}_\varphi, \hat{\sigma}_\varphi^2, \hat{\rho}, \hat{\sigma}^2\right)$ are obtained.

3. The Bayesian framework for adaptive model parameters via real-time CM data

The estimation procedure presented in Section 2 obtains the value of model coefficients for the whole population given historical data. However, to estimate the model coefficients for a specific system, it is more desirable to use the real-time CM observations collected during the system operation. In this section, we will develop a Bayesian framework for the update of model coefficients. According to the model in Section 2, the stochastic parameter $\{\beta, \varphi\}$ has the prior distribution of $\beta, \varphi \propto N\left(\mu_\beta, \sigma_\beta^2, \mu_\varphi, \sigma_\varphi^2, \rho\right)$. The estimates $\left\{\hat{\mu}_\beta, \hat{\sigma}_\beta^2, \hat{\mu}_\varphi, \hat{\sigma}_\varphi^2, \hat{\rho}, \hat{\sigma}^2\right\}$ calculated by the procedure presented in Section 2 using historical data can be the prior estimates of $\left\{\mu_\beta, \sigma_\beta^2, \mu_\varphi, \sigma_\varphi^2, \rho, \sigma^2\right\}$.

Let $X_{1:r} = \{x_1, x_2, ..., x_r\}$ where $x_r = y_r - \delta_0$, then given β and φ, the sampling distribution of $X_{1:r}$ is multi-variable normal as:

$$p(X_{1:r} \mid \beta, \varphi) = \frac{1}{\prod_{j=1}^{r} \sqrt{2\pi\sigma^2}} \times \exp\left[-\sum_{j=1}^{r} \frac{\left(x_j - \beta t_j - \varphi x_{j-1}\right)^2}{2\sigma^2}\right]. \quad (14)$$

Then the joint posterior estimate of β and φ conditional on $X_{1:r}$ is still normal resulted from the fact of the normal distribution assumption of β and φ. In other words, $\beta, \varphi \mid X_{1:r} \sim N\left(\mu_{\beta,r}, \sigma_{\beta,r}^2, \mu_{\varphi,r}, \sigma_{\varphi,r}^2, \rho_r\right)$. To be more precise, we have:

$$\begin{aligned}
p\left(\beta,\varphi \mid X_{1:r}\right) &\propto p\left(X_{1:r} \mid \beta,\varphi\right) \cdot p\left(\beta,\varphi\right) \\
&\propto exp\left[-\sum_{j=1}^{r} \frac{\left(x_j - \beta t_j - \varphi x_{j-1}\right)^2}{2\sigma^2}\right] \\
&\cdot exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left[\frac{\left(\beta-\mu_\beta\right)^2}{\sigma_\beta^2} - 2\rho\frac{\left(\beta-\mu_\beta\right)\left(\varphi-\mu_\varphi\right)}{\sigma_\beta\sigma_\varphi} + \frac{\left(\varphi-\mu_\varphi\right)^2}{\sigma_\varphi^2}\right]\right\} \\
&\propto exp\{-\frac{1}{2\sigma^2\sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)}[\beta^2\left(\sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)\sum_{j=1}^{r} t_i^2 + \sigma^2\sigma_\varphi^2\right) \\
&\quad +\varphi^2\left(\sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)\sum_{j=1}^{r} x_{i-1}^2 + \sigma^2\sigma_\beta^2\right) \\
&\quad -2\beta\left(\sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)\sum_{j=1}^{r} t_i x_i + \sigma^2\mu_\beta\sigma_\varphi^2 - \sigma^2\mu_\varphi\sigma_\beta\sigma_\varphi\rho\right) \\
&\quad -2\varphi\left(\sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)\sum_{j=1}^{r} x_i x_{i-1} + \sigma^2\mu_\varphi\sigma_\beta^2 - \sigma^2\mu_\beta\sigma_\beta\sigma_\varphi\rho\right) \\
&\quad +2\beta\varphi\left(\sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)\sum_{j=1}^{r} t_i x_{i-1} - \sigma^2\sigma_\beta\sigma_\varphi\rho\right)]\} \\
&\propto \frac{1}{2\pi\sigma_{\beta,r}\sigma_{\varphi,r}\sqrt{1-\rho_r^2}} \exp\{-\frac{1}{2\left(1-\rho_r^2\right)}[\frac{\left(\beta-\mu_{\beta,r}\right)^2}{\sigma_{\beta,r}^2} \\
&\quad -2\rho_r\frac{\left(\beta-\mu_{\beta,r}\right)\left(\varphi-\mu_{\varphi,r}\right)}{\sigma_{\beta,r}\sigma_{\varphi,r}} + \frac{\left(\varphi-\mu_{\varphi,r}\right)^2}{\sigma_{\varphi,r}^2}]\}
\end{aligned} \quad (15)$$

with:

$$\begin{aligned}
\mu_{\beta,r} &= \frac{A_2B_1 - A_1C}{B_1B_2 - C^2}, \quad \mu_{\varphi,r} = \frac{A_1B_2 - A_2C}{B_1B_2 - C^2}, \\
\sigma_{\beta,r}^2 &= \frac{B_1D}{B_1B_2 - C^2}, \quad \sigma_{\varphi,r}^2 = \frac{B_2D}{B_1B_2 - C^2}, \\
\rho_r &= \frac{-C}{\sqrt{B_1B_2}}.
\end{aligned} \quad (16)$$

where:

$$\begin{aligned}
A_1 &= \sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)\sum_{j=1}^{r} x_{i-1}x_i + \sigma^2\mu_\varphi\sigma_\beta^2 - \sigma^2\mu_\beta\sigma_\beta\sigma_\varphi\rho, \\
A_2 &= \sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)\sum_{j=1}^{r} x_i t_i + \sigma^2\mu_\beta\sigma_\varphi^2 - \sigma^2\mu_\varphi\sigma_\beta\sigma_\varphi\rho, \\
B_1 &= \sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)\sum_{j=1}^{r} x_{i-1}^2 + \sigma^2\sigma_\beta^2, \\
B_2 &= \sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)\sum_{j=1}^{r} t_i^2 + \sigma^2\sigma_\varphi^2, \\
C &= \sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right)\sum_{j=1}^{r} x_{i-1}t_i - \sigma^2\sigma_\beta\sigma_\varphi\rho, \\
D &= \sigma^2\sigma_\beta^2\sigma_\varphi^2\left(1-\rho^2\right).
\end{aligned} \quad (17)$$

Using the conditional joint posterior distribution of β and φ, we are able to calculate the distribution of X_n given $X_{n\text{-}1}$. In fact, X_n follows a normal distribution with mean and variance given by:

$$\begin{aligned}
E\left(X_n \mid X_{n-1}\right) &= \mu_{\beta,n-1}t_n + \mu_{\varphi,n-1}X_{n-1}, \\
Var(X_n \mid X_{n-1}) &= \sigma_{\beta,n-1}^2 t_n^2 + \sigma_{\varphi,n-1}^2 X_{n-1}^2 + \sigma^2 \\
&\quad + 2\rho t_n X_{n-1}\sigma_{\beta,n-1}\sigma_{\varphi,n-1}.
\end{aligned} \quad (18)$$

4. Residual life estimation and maintenance policy optimization

Our next goal is to determine the residual life of the operating system and the optimal control-limit to initialize preventive replacement. Since φ is a random variable, the derivation of an analytical form for the mean residual life and the failure time distribution, is encountering a large amount of difficulties. Therefore, we hereby consider at first a Monte Carlo simulation-based approach which generates a large sample of deterministic degradation paths to approximate the residual life distribution. On the other hand, for the maintenance optimization problem, we consider the commonly-used control-limit policy (e.g., [15, 12, 23, 6]), by which the system will be preventively replaced if its observed degradation state exceeds a control-limit (optimized) and it is left operational until next inspection if its degradation level is below the control-limit. After preventive replacement or corrective replacement (perform when the system fails), the system will go back to as-good-as-new state Y_0. Since the model coefficients update as real-time condition monitoring data becomes available, the optimal control-limit to initialize preventive replacement may have to change as well. Therefore, the optimal control-limit is dynamic.

We develop an optimization algorithm based on SMDP framework to obtain the dynamic control-limit, which is based on the algorithm proposed in [23] for fixed control-limit situation. Moreover, we discover that using the SMDP framework, the mean residual life and the dynamic optimal control-limit can be obtained simultaneously. To be specific, without any extra effort, the mean residual life can be obtained while calculating the optimal control-limit. It is interesting to find that the SMDP-based approach provides a novel way of calculating mean residual life, which is commonly considered as an approach only for policy decision problems. We will introduce this approach and discuss its advantages and disadvantages.

For both problems, we assume that failure occurs when the degradation signal reaches some given failure threshold ξ. When the degradation state Y_n reaches the threshold ξ, the system is no longer assumed to be able to function satisfactorily or safely and it should be correctively replaced, although no physical failure is observed. In this paper, we take the threshold value ξ as fixed and known, and assume the failure should be discovered by equidistant inspections.

4.1. Revisited: residual life estimation in the situation of fixed model coefficients

Before describing our approaches, we first revisit the residual life estimation approach under the situation of fixed β and φ for comparison purposes (refer to [23] for more details). For fixed β and φ, the conditional expected mean and variance of X_n can be obtained by conditioning on previous observations $X_0, X_1, ..., X_j$ for some integer $j < n$:

$$\begin{aligned} E_f(X_n \mid X_j) &= \beta \sum_{r=j+1}^{n} \varphi^{n-r} t_r + \varphi^{n-j} X_j \\ Var_f(X_n \mid X_j) &= \sigma^2 \sum_{r=0}^{n-j-1} \varphi^{2r}. \end{aligned} \tag{19}$$

Let $F_T(t_n \mid X_j)$ be the failure time distribution given the current observation X_j, and it is calculated by:

$$\begin{aligned} F_T(t_n \mid X_j) &= Pr\left(T \langle t_n | X_j\right) \\ &= Pr(X_n \geq \xi - \delta_0 \mid X_j) \\ &= 1 - \Phi\left(\frac{\xi - \delta_0 - E_f(X_n \mid X_j)}{\sqrt{Var_f(X_n \mid X_j)}}\right) \end{aligned} \tag{20}$$

Thus, the mean residual life at a given time $t_j = jh$ can be calculated by:

$$E(T \mid jh) = \sum_{n=j+1}^{+\infty} t_n [F_T(t_n \mid X_j) - F_T(t_{n-1} \mid X_j)] - jh. \tag{21}$$

4.2. Monte Carlo simulation-based approach for residual life estimation

When both β and φ are random variables, it is difficult to derive analytically the failure time distribution $F_T(t_n \mid X_j)$, the conditional expected mean $E(X_n \mid X_j)$ and variance $Var(X_n \mid X_j)$. When there is no closed-form expression for the above distributions, one can evaluate their estimates to any desired degree of precision using Monte Carlo simulation. This is done by generating a sufficiently large number of random sample paths from the assumed degradation model with the estimated coefficients. We use the following procedure:

1. Generate U simulated realizations of $\tilde{\beta}$ and $\tilde{\varphi}$ from $\beta, \varphi \mid X_{1:j} \sim N\left(\mu_{\beta,j}, \sigma^2_{\beta,j}, \mu_{\varphi,j}, \sigma^2_{\varphi,j}, \rho_j\right)$, where U is a large number (e.g., $U = 100{,}000$);
2. Given the current observation Y_j at inspection time $t_j = jh$, generate simulated random errors $\hat{\varepsilon}_n$ from $N\left(0, \sigma^2\right)$. For each of the U simulated paths, $X_n = Y_n - \delta_0$ for any $n > j$ is calculated by:

$$\begin{aligned} X_{j+1} &= \beta t_{j+1} + \varphi X_j + \hat{\varepsilon}_1, \\ X_{j+2} &= \beta t_{j+2} + \varphi X_{j+1} + \hat{\varepsilon}_2 \\ &= \beta t_{j+2} + \varphi\left(\beta t_{j+1} + \varphi X_j + \hat{\varepsilon}_1\right) + \hat{\varepsilon}_2, \\ X_n &= \beta t_n + \varphi X_{n-1} + \varepsilon_n,\ n = j+1, j+2, j+3, ... \end{aligned} \tag{22}$$

Then the residual life T_l for the l th simulated path is determined by:

$$\hat{T}_l = h \times \underset{n}{\operatorname{argmin}} \left\{X_n - (\xi - \delta_0) \geq 0\right\} \tag{23}$$

3. Compute the corresponding $F_T(t_n \mid X_j)$ using U simulated paths for any desired values of $t_n = h, 2h, ...$

$$F_T(t_n \mid X_j) = \frac{\text{number of } \hat{T}_l \leq t_n}{U} \tag{24}$$

4. Calculate the mean residual life $E(T \mid jh)$ at current inspection time $t_j = jh$ by Eq.(21).

4.3. SMDP-based approach for maintenance optimization and mean residual life estimation

We consider using a SMDP framework to optimize the dynamic maintenance policy based on the proposed random-coefficient autoregressive model with time effect. In this maintenance policy, three costs are required, which are, a cost of C_P for a preventive replacement, a cost of C_F for a corrective replacement and a cost of C_{Obs} for an inspection; two replacement times are included, which are preventive replacement time T_F and corrective replacement time T_P. After these quantities are assigned, the economic consequence of using this maintenance policy can be reflected by the long-run expected average cost per unit time g.

The control-limit will be optimized each time when the model coefficients are updated using new available observations. Therefore, if new observations are available at inspection time $t_n = nh$, the objective of the maintenance policy is to find the optimal control-limit $\overline{w}_n^*$ to determine whether a preventive replacement should be initialized before next update of model coefficients, by minimizing the long-run expected average cost per unit time g. By renewal theory (see e.g.[20]), the cost minimization problem is equivalent to finding an optimal control-limit $\overline{w}_n^* \in \left(y_0, \xi\right]$ such that:

$$g\left(\overline{w}_n^*\right) = \frac{E_{\overline{w}_n^*}\left(CC\right)}{E_{\overline{w}_n^*}\left(CL\right)} = \inf \frac{E_{\overline{w}_n^*}\left(CC\right)}{E_{\overline{w}_n^*}\left(CL\right)}. \tag{25}$$

where CL and CC denote the cycle length and cycle cost for the systems whole lifecycle, respectively.

We surprisingly discover that the problem of estimating the mean residual life of the operating system can be incorporated into the problem of optimizing the dynamic control-limit. That is, if the control-limit $\overline{w}_n$ equals to the failure threshold ξ, $\overline{w}_n = \xi$, then the total length of the lifecycle CL under this policy is the same as that without any preventive policy. The system will ultimately go to the failure state. Thus, $CL - jh - T_F$ equals to the residual life estimated at $t_j = jh$. Then the expected total length of the lifecycle $E\left(CL\right) - jh - T_F$ equals exactly to the mean residual life $E(T \mid jh)$.

Since the system will ultimately fail, the expected total cost $E\left(CC\right)$ can be given by:

$$E\left(CC\right) = C_F + C_{Obs}\frac{E\left(CL\right)}{h} \tag{26}$$

On the other hand, if the long-run expected average cost per unit time g is known, $E\left(CC\right)$ can also be calculated by:

$$E\left(CC\right) = gE\left(CL\right) \tag{27}$$

Using Eq.(26) and Eq.(27), the residual life estimated at time $t_j = jh$ is given by:

$$\begin{gathered} E\left(CL\right) = \frac{C_F h}{gh - C_{Obs}}, \\ E(T \mid jh) = E\left(CL\right) - jh - T_F. \end{gathered} \tag{28}$$

Note that although the values of costs C_F, C_{Obs} and replacement times T_F are very important for maintenance optimization, for the residual life estimation, the costs and replacement times are intermediate quantities to obtain results. They can be set to any value. No matter what values they have, the system will ultimately go failure (the lifecycle length won't change) and have corrective maintenance. Therefore, when they change, the average cost g changes accordingly, which is given by $E\left(CC\right)/E\left(CL\right)$.

Now the only remaining question for the residual life estimation problem is, how to calculate g without knowing the exact value of $E\left(CL\right)$? For the maintenance optimization problem, we develop a SMDP framework to calculate g for each $\overline{w}_n \in \left(Y_0, \xi\right]$ and decide the optimal $\overline{w}_n^*$ at inspection time $t_n = nh$ by the minimum g^*. The whole searching procedure contains the situation of $\overline{w}_n = \xi$. Hence, we are able to obtain g when the control-limit is $\overline{w}_n = \xi$, simultaneously with the searching procedure of optimal control-limit $\overline{w}_n^*$.

Next, we will describe in detail how to calculate g using a SMDP framework and how to find the optimal dynamic control-limit $\overline{w}_n^*$ at each inspection time $t_n = nh$. To develop this SMDP framework, firstly, the possible range of Y_n is required to be discretized into a finite set of states. We define the state space as a combination of countable time points and value intervals. Denote the state space by $\Omega = \left(\mathbf{K}, \mathbf{H}\right)$, where $\mathbf{K}$ represents the discretized states of observed Y_n, and $\mathbf{H} = \{nh; n = 0,1,2,...\}$ represents the inspection times. We set the possible smallest value of Y_n by $y_0^{'} = y_0 - 3\sigma$. Define $\left[\xi, +\infty\right)$ as the failure state $\mathbf{F}$, then we can divide the continuous state space of $\left[y_0^{'}, \xi\right]$ into L equidistant intervals with constant length $\Delta = \left(\xi - y_0^{'}\right)/L$. For the maintenance policy, we define the control-limit $\overline{w}_n$ by $\overline{k}\Delta$, for some fixed integer $0 < \overline{k} < L$, then the warning state $\mathbf{W}$ will be $\left[\overline{k}\Delta + y_0^{'}, L\Delta + y_0^{'}\right)$, and the healthy state $\mathbf{S}$ will be $\left[y_0^{'}, \overline{k}\Delta + y_0^{'}\right)$.

Secondly, the quantities in the SMDP should be determined, which are one-step transition probabilities of degradation states, one-step expected sojourn times and one-step expected costs. We also define an integer $\tilde{n}$ to be the total number of inspections for the system. If the system still operates without failure when the last inspection is performed, we will enforce a preventive replacement. The degrading system will surely be replaced before the last inspection if $\tilde{n}$ is large enough.

The one-step transition probabilities are calculated by Eq.(29), where $E(X_{n+1} \mid X_n)$ and $Std(X_{n+1} \mid X_n)$ are calculated by Eq.(18).

$$\begin{aligned}
p_{(k,n),(l,n+1)} &= \begin{cases} \Pr\left(Y_{n+1} \in \left[l\Delta + y_0^{'}, (l+1)\Delta + y_0^{'}\right], 0 \le k < L \middle| Y_0 = y_0\right), \text{for } k = 0, n = 0 \\ \Pr\left(Y_{n+1} \in \left[l\Delta + y_0^{'}, (l+1)\Delta + y_0^{'}\right], 0 \le k < L \middle| Y_n = (k+0.5)\Delta + y_0^{'}, 0 \le k < \overline{k}\right), \text{for n} \ge 1 \end{cases} \\
&= \int_{l\Delta + y_0^{'}}^{(l+1)\Delta + y_0^{'}} \frac{1}{\sqrt{2\pi}\, Std(Y_{n+1} \mid Y_n)} \exp\left\{-\frac{1}{2}\left[\frac{y - E(Y_{n+1} \mid Y_n)}{Std(Y_{n+1} \mid Y_n)}\right]^2\right\} dy \\
&= \Phi\left(\frac{(l+1)\Delta + y_0^{'} - E(Y_{n+1} \mid Y_n)}{Std(Y_{n+1} \mid Y_n)}\right) - \Phi\left(\frac{l\Delta + y_0^{'} - E(Y_{n+1} \mid Y_n)}{Std(Y_{n+1} \mid Y_n)}\right) \\
&= \Phi\left(\frac{(l+1)\Delta + y_0^{'} - \delta_0 - E(X_{n+1} \mid X_n)}{Std(X_{n+1} \mid X_n)}\right) - \Phi\left(\frac{l\Delta + y_0^{'} - \delta_0 - E(X_{n+1} \mid X_n)}{Std(X_{n+1} \mid X_n)}\right) \\
p_{(k,n),(\mathbf{W},n+1)} &= \begin{cases} Pr\left(Y_{n+1} \in \mathbf{W} \middle| Y_0 = y_0\right), \text{ for } k = 0, n = 0 \\ Pr\left(Y_{n+1} \in \mathbf{W} \middle| Y_n = (k+0.5)\Delta + y_0^{'}, 0 \le k < \overline{k}\right), \text{for n} \ge 1 \end{cases} \\
&= \sum_{l=\overline{k}}^{L} p_{(k,n),(l,n+1)} \\
p_{(k,n),(\mathbf{F},n+1)} &= \begin{cases} Pr\left(Y_{n+1} \in \mathbf{F} \middle| Y_0 = y_0\right), \text{for } k = 0, n = 0 \\ Pr\left(Y_{n+1} \in \mathbf{F} \middle| Y_n = (k+0.5)\Delta + y_0^{'}, 0 \le k < \overline{k}\right), \text{for } n \ge 1 \end{cases} \\
&= 1 - \Phi\left(\frac{\xi - E(Y_{n+1} \mid Y_n)}{Std(Y_{n+1} \mid Y_n)}\right) = 1 - \Phi\left(\frac{\xi - \delta_0 - E(X_{n+1} \mid X_n)}{Std(X_{n+1} \mid X_n)}\right).
\end{aligned} \tag{29}$$

The expected sojourn times for the SMDP are given by:

$$\begin{aligned} &\tau\left(k,n\right) = h, k = 0,1,\ldots,\overline{k}; n = 0,1,\ldots,\tilde{n} - 1 \\ &\tau\left(\mathbf{F},n\right) = T_F, n = 0,1,...,\tilde{n} \\ &\tau\left(\mathbf{W},n\right) = \tau\left(k,\tilde{n}\right) = T_P,\ k = 0,1,...,\overline{k}; n = 0,1,...,\tilde{n}. \end{aligned} \tag{30}$$

Similarly, the expected costs for the SMDP are given by:

$$\begin{aligned} &c\left(k,n\right) = C_{Obs}, k = 0,1,...,\overline{k} \text{ and } n = 0,1,...,\tilde{n} - 1 \\ &c\left(\mathbf{F},n\right) = C_F,\ n = 0,1,...,\tilde{n} \\ &c\left(\mathbf{W},n\right) = c\left(k,\tilde{n}\right) = C_P,\ k = 0,1,...,\overline{k}, \text{ and } n = 0,1,...,\tilde{n}. \end{aligned} \tag{31}$$

Therefore, at each inspection time $t_j = jh$, with new available observation Y_j and updated model parameters, the long-run expected average cost per unit time $g\left(\overline{w}_j\right)$ given the fixed control-limit $\overline{w}_j = \overline{k}\Delta$ can be obtained by solving the following system of linear equations:

$$
\begin{aligned}
& v(0,0) = jC_{Obs} - g\left(\bar{w}_j\right) jh + v(a,j) \\
& v(a,j) = c(a,j) - g\left(\bar{w}_j\right)\tau(a,j) + \sum_{l=0}^{\bar{k}} p_{(a,j),(l,j+1)} v(l,j+1) \\
& \quad + p_{(a,j),(\mathbf{W},j+1)} v(\mathbf{W},j+1) + p_{(a,j),(\mathbf{F},j+1)} v(\mathbf{F},j+1) \\
& v(k,n) = c(k,n) - g\left(\bar{w}_j\right)\tau(k,n) + \sum_{l=0}^{\bar{k}} p_{(k,n),(l,n+1)} v(l,n+1) \\
& \quad + p_{(k,n),(\mathbf{W},n+1)} v(\mathbf{W},n+1) + p_{(k,n),(\mathbf{F},n+1)} v(\mathbf{F},n+1), \\
& k = 0,1,\ldots,\bar{k} \text{ and } n = j+1, j+2, \ldots, \tilde{n}-1 \qquad (32) \\
& v(\mathbf{W},n) = c(\mathbf{W},n) - g\left(\bar{w}_j\right)\tau(\mathbf{W},n) + v(0,0), n = j+1,2,\ldots,\tilde{n} \\
& v(\mathbf{W},n) = c(\mathbf{W},n) - g\left(\bar{w}_j\right)\tau(\mathbf{W},n) + v(0,0), n = j+1,2,\ldots,\tilde{n} \\
& v(k,\tilde{n}) = v(\mathbf{W},\tilde{n}), k = 0,1,\ldots,\bar{k} \\
& v(p,q) = 0, \text{ for an arbitrarily selected single state } (p,q)
\end{aligned}
$$

where $v(\bullet,\bullet)$ is the so-called relative value function plus a constant and $a=\lceil\left(Y_j - y_0'\right)/\Delta\rceil$. The first equation in Eq.(32) indicates that after $t_j = jh$ unit of time, the system runs from the initial state $(0,0)$ to the state of (a,j). The last equation in Eq.(32) guarantees that the solution to the system of linear equations in Eq.(32) is unique (see e.g.,[25]).

So that the optimal control-limit $\bar{w}_j^*$ at inspection time $t_j = jh$ and the corresponding minimum long-run expected average cost per unit time $g\left(\bar{w}_j^*\right)$ can be found by:

$$
g\left(\bar{w}_j^*\right) = \inf_{\bar{w}_j \in \left[y_0', \xi\right]} \left\{ g\left(\bar{w}_j\right) \right\} \qquad (33)
$$

Since only a single admissible action in each state is possible for a given control-limit, it is not necessary to formally apply the whole policy iteration procedure. We are only interested in computing the long-run average cost per unit time for a given control limit policy, and we chose SMDP for the computation, because we can make efficient use of the linear equations in step 1 of the policy iteration algorithm.

5. Case study and discussions

In this section, we will conduct a comparative study to demonstrate our proposed method. To reveal the effectiveness of the proposed random-coefficient autoregressive model with time effect, we will compare it with its fixed-coefficient counterpart using the same case studied in [22]. We will also use this case to illustrate our proposed dynamic maintenance policy and the approach of estimating the mean residual life for a functioning system.

5.1. The degradation dataset

The dataset is a real laser degradation data set presented by [16] (Example 13.6). It consists of 13 degradation histories of GaAs lasers (see Fig.1). Over the life of these lasers, degradation causes a decrease in light output. However, the lasers contain a feedback mechanism that maintains nearly constant light output by increasing operating current as the laser degrades. When operating current gets too high, the laser is considered to have failed. In applications, experts consider the laser failed if the operating current increases to ξ percent of its original value ($\xi \le 10$). To track the lasers degradation, in this data set, the operating currents were measured every $h = 20$ hours up to 4000 hours.

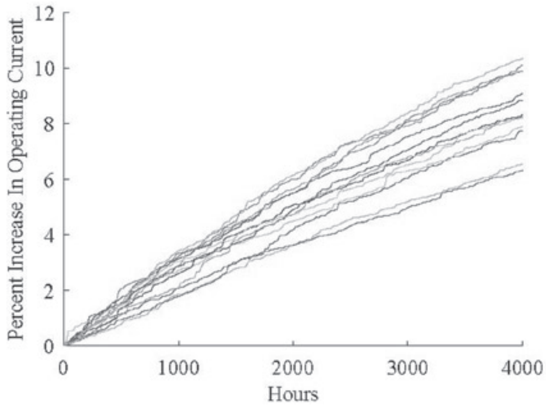


Fig. 1. Degradation paths in terms of the percent increase in operating current for the GaAs laser data set coming from [16]

5.2. Estimation of model coefficients

To demonstrate the present random-coefficient model formation is preferable to the fixed-coefficient one proposed in [23], we firstly use the 13 degradation histories as the training data to estimate the model coefficients $\gamma = \left(\delta_0, \mu_\beta, \sigma_\beta^2, \mu_\varphi, \sigma_\varphi^2, \rho, \sigma^2\right)$, using the two model formations respectively. According to the estimation procedure presented in Section 2 and in [23], we obtain the values of the model coefficients for both models, as listed in Table 1. Note that for the fixed-coefficient model, $\sigma_\beta^2 = 0$, $\sigma_\varphi^2 = 0$, and $\rho = 0$.

In Fig.2, a graphical proof is presented to directly show the superiority of the random-coefficient model over the fixed-coefficient model in capturing the lasers degradation behavior. This figure shows simulated degradation paths of 30 lasers based on the estimates in Table 1. It can be observed that the simulated paths using the random-coefficient model are very similar to the actual paths (refer to Fig.1). However, the simulated paths using the fixed-coefficient model are quite different, in that

1. The number of intersections in the actual paths is quite smaller than that in the simulated paths.
2. The decreasing phases are less pronounced than that in the simulated paths.
3. the variability of each path around its mean is quite smaller in the actual paths than that in the simulated paths.

Table 1. Parameters estimation results using the data of the training 13 lasers.

Model coefficients	Fixed-coefficient (refer to [23])	Random-coefficient
$\hat{\delta}_0$	-14.3722	-14.3722
$\hat{\mu}_\beta$	-1.4755×10^{-5}	-1.4750×10^{-5}
$\hat{\sigma}_\beta$	/	4.6099×10^{-6}
$\hat{\mu}_\varphi$	1.0038	1.0038
$\hat{\sigma}_\varphi$	/	6.6884×10^{-4}
$\hat{\rho}$	/	-0.9915
$\hat{\sigma}$	0.0636	0.0240

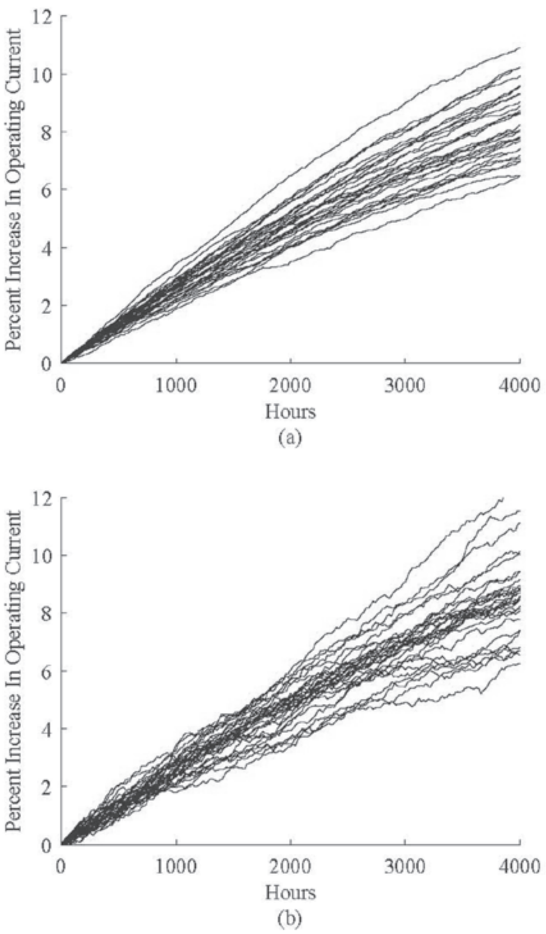


Fig. 2. Simulated degradation paths of 30 GaAs lasers based on (a) random-coefficient autoregressive model with time effect and (b) fixed-coefficient autoregressive model with time effect

All these differences indicate the heterogeneity does exist in this dataset and considering the model coefficient as random is more appropriate. The model with fixed coefficients places all uncertainties on the parameter σ, whose value we can observe from Table 1 is quite larger than the one in random-coefficient model. This is why the fluctuations in Fig.2(b) are more and larger. Moreover, in Fig.3, the estimated mean and the standard deviation of the percent increase in operating current calculated by the random-coefficient model (using mean and standard deviation of 3000 simulated paths) are compared with the corresponding empirical estimates calculated from the observed data. The results also indicate that the random-coefficient model fits well with this dataset.

5.3. Update of model coefficients using the Bayesian framework

In order for the model coefficients to better accommodate a specific functioning laser, we use the Bayesian framework proposed in Section 3 to update the model coefficients once new observations are available. To demonstrate the advantages of this update procedure, we randomly select 4 degradation histories which relatively degrade slowly in the dataset as the training units, to estimate the initial values of the model coefficients. Then we select 1 degradation history which relatively degrades quickly in the dataset as the testing unit, assuming this unit is the functioning laser. The training units and the testing unit are plotted in Fig.4.

With the joint prior distribution of model coefficients $\{\beta,\varphi\}$, we are able to find the joint posterior distribution of $\{\beta,\varphi\}$ for the functioning laser any time we obtain a new observation, i.e., at any inspection time $t_n = nh$. Fig.5 presents the evolution of the posterior means for β, and φ respectively, given the observation data from the testing unit. Given these posterior means, we can then compute the expected percent increase at the next sampling time using the following equation:

$$E_{\hat{\gamma}}(Y_n \mid Y_{n-1}) = \hat{\mu}_{\beta,n-1} t_n + \hat{\mu}_{\varphi,n-1}\left(Y_{n-1} - \hat{\delta}_0\right) + \hat{\delta}_0 \quad (34)$$

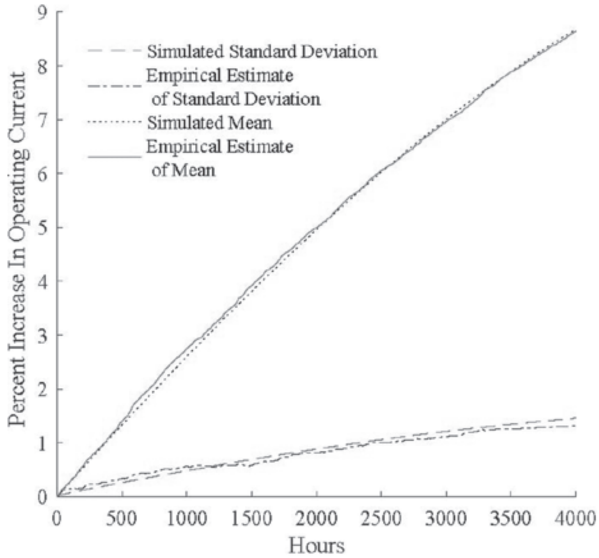


Fig. 3. Curves of the estimated mean and standard deviation of the percent increase in operating current compared with the corresponding empirical estimates

Fig.6 shows the observed and the expected percent increase in the operating current plotted against time for the testing unit. The results show that using the Bayesian updating framework, the expected degradation path follows well with the actual degradation path. To demonstrate the advantages of using the Bayesian updating framework, the root mean squared error (RMSE) for the testing unit under the random-coefficient model with and without Bayesian updating procedure are calculated respectively, which is defined by Eq.(35). For the model without Bayesian updating procedure, the RMSE is 0.0464, while for the model with Bayesian updating procedure, the RMSE is 0.0440. The results show that the Bayesian updating procedure increases the accuracy of predicting the degradation state of functioning laser.

$$RMSE_1 = \sqrt{\frac{1}{Q}\sum_{n=1}^{Q}[Y_n - E_{\hat{\gamma}}(Y_n \mid Y_{n-1})]^2} \quad (35)$$

where $Q = 200$ is the total number of observations of the testing unit.

5.4. Optimization of the dynamic maintenance policy

With the posterior estimates of model coefficients, we are able to find the dynamic control-limit to initialize preventive maintenance. To illustrate the whole optimization procedure, we use the 4 training units in Fig.4 to obtain the initial values of model coefficients. Then, the dynamic policy is optimized using the algorithm proposed in Sec-

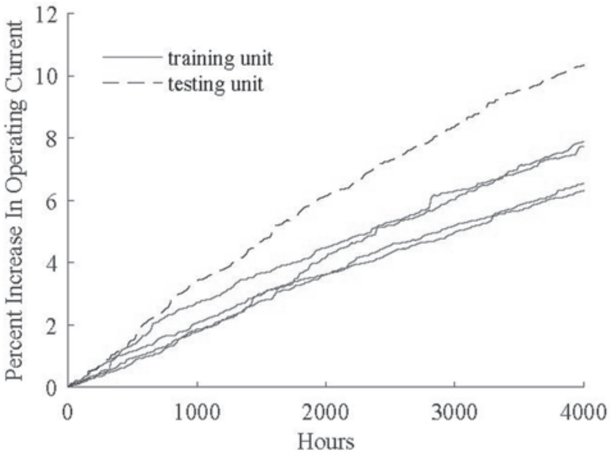


Fig. 4. Degradation paths in terms of the percent increase in operating current for the training units and the testing unit.

tion 4, with the help of real-time observations coming from the testing unit and the posterior estimates of model coefficients.

Suppose that the failure threshold is $\xi = 4$. We determine the optimal control-limit every 20 inspections with the following replacement times and cost data: $T_P = T_F = 20$ hours, $C_F = \$3000$, $C_P = \$1000$ and $C_{Obs} = \$100$. We partition the continuous degradation interval $\left[Y_0 = 0, \xi = 4\right]$ into 64 sub- intervals, and set the maximum number of inspections to $\hat{n} = 128$. For every 20 inspections,

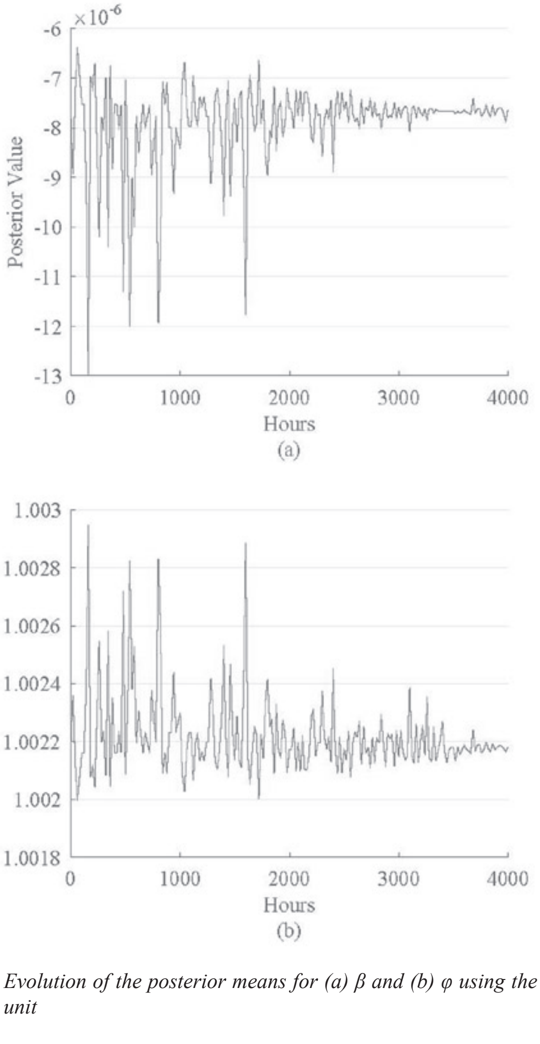


Fig. 5. Evolution of the posterior means for (a) β and (b) φ using the testing unit

that is, at inspection time $t_n = nh$, $n = 20, 40, 60, \ldots$, we obtain the minimum long-run expected average cost per hour $g\left(\bar{\omega}_n^*\right)$ with the corresponding optimal control-limit $\bar{\omega}_n^*$. The dynamic optimal control-limit for the maintenance policy is shown in Fig.7. It reveals that for this dataset, the optimal control-limit neither shows a monotone nondecreasing trend as [6] and [24], nor shows a decreasing trend as [1]. It firstly decreases then raises back to its initial value. This going-back control-limit seems to be a proof to doubt the significance of using this dynamic maintenance policy. However, from the behavior of the dynamic control-limit, we actually can see the increased urgency for preventive maintenance when degradation of the functioning laser begins to deviate fast from the historical degradation paths (refer to Fig.4 to see the faster degradation of the testing unit). The following increased behavior of the control-limit can be explained by increased accuracy (less variability) in predicting the future degradation trend. Furthermore, this result only reflects the situation under current maintenance situation (maintenance costs, maintenance times and etc.).

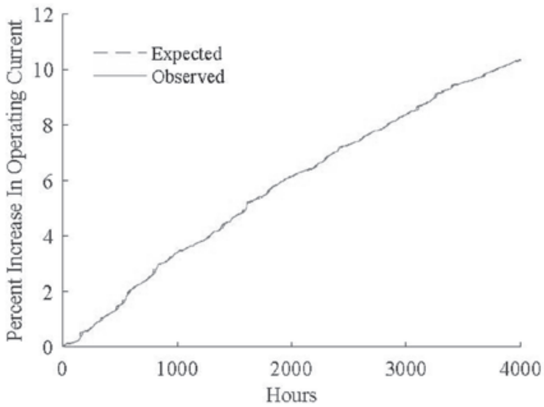


Fig. 6. Expected and observed percent increase in operating current vs. time for the testing unit

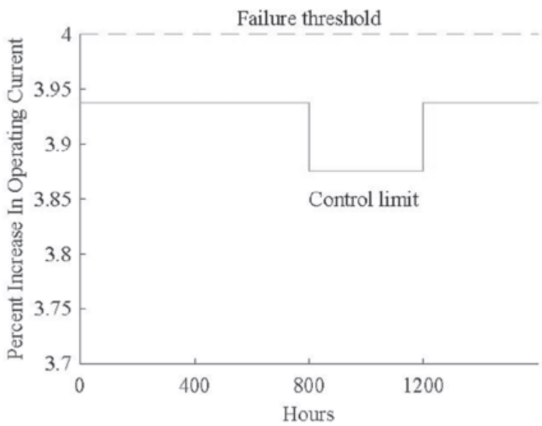


Fig. 7. Optimization results of the dynamic control-limit

5.5. Real-time estimation of residual life

We then use the posterior means and variances of model coefficients to compute the mean residual life given observations obtained up to that point in time. We firstly use the Monte Carlo simulation-based approach and make a comparison between the results obtained by the proposed random-coefficient model with Bayesian updating procedure and the fixed-coefficient model. The model coefficients are calculated using the training units presented in Fig.4 and updated by observations of the testing unit in Fig.4. We use the following RMSE to evaluate their performances:

$$RMSE_2 = \sqrt{\frac{1}{Q}\sum_{j=1}^{Q}[E_{true}(T \mid jh) - E(T \mid jh)]^2} \tag{36}$$

where $E_{true}(T \mid jh)$ and $E(T \mid jh)$ are true remaining time to observe failure at $t_j = jh$ and expected mean residual life estimated at $t_j = jh$, respectively, and Q is the total number of observations before system failure.

The results of RMSE are: for the fixed-coefficient model, $RMSE_2 = 432$, while for the random-coefficient model with Bayesian updating procedure, $RMSE_2 = 427$. The results support that the random-coefficient model with Bayesian updating procedure improves the performance of residual life estimation. To give a vivid impression of the results, we plot in Fig.8 the estimation results at each inspection time for both the fixed-coefficient model and the random-coefficient model with Bayesian updating procedure. The figure shows that although the estimates of the residual life become more accurate with more available CM observations for both models, the random-coefficient model with Bayesian up-dating procedure seems to give more estimates which are closer to the true residual life, especially in the middle stage of the degradation.

Next, we present the procedure of using our proposed SMDP-based approach to estimate the mean residual life. In this case study, with the process of optimizing the control-limit at inspection time $t_n = nh$, $n = 0, 20, 40, 60$, we are able to obtain the mean residual life estimates given the observations up until t_n, using the long-run expected average cost per hour $g(\bar{\omega}_n)$ when $\bar{\omega}_n = \xi$ and the equations Eq.(26) - Eq.(28). The results are shown in Table 2. The true residual life and the estimated mean residual life obtained by Monte Carlo simulation-based approach are also listed in Table 2. It shows that the estimation results by SMDP-based approach are very close to the results obtained by Monte Carlo simulation-based approach, indicating that the SMDP-based approach for residual life estimation is reliable. The slight difference may due to the different approximation methods used in these two approaches.

Although the SMDP-based approach is feasible for residual life estimation, we still have to point out that it is only an approximate approach for residual life estimation and its computation cost is too large. It could estimate mean residual life simultaneously and quickly with the optimization process of the dynamic maintenance policy analyzed in this paper, but it is not an efficient approach if it is used independently. However, for a system using control-limit policy to conduct condition-based maintenance, engineers are more enthusiastic to know the system remaining time to the preventive control-limit than the system remaining time to failure. To approximately estimate the system remaining time to the control-limit, our proposed SMDP-based approach provides a very direct and easy way, by using the results of the minimum long-run expected average cost per unit time .

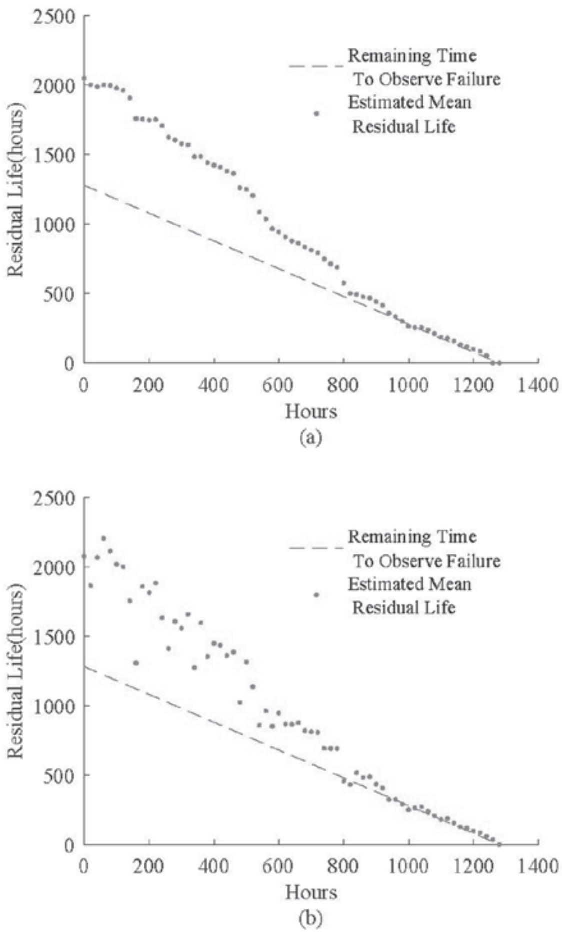


Fig. 8. The actual remaining time to observe failure and the estimated mean residual life for the testing unit using (a) the fixed-coefficient autoregressive model with time effect and (b) the random-coefficient autoregressive model with time effect and Bayesian updating procedure

Table 2. The results of mean residual life estimation using SMDP-based approach.

Inspection time (hours)	0	400	800	1200
True residual life	1280	900	500	100
Estimate in Fig.8(b)	2074	1450	456	98
Estimate using SMDP-based approach	2062	1499	464	135

6. Conclusion

In this paper, we have presented a dynamic CBM policy using random-coefficient autoregressive model with time effect. This random-coefficient autoregressive model with time effect has a more general formation than the fixed-coefficient counterpart ([23]) by considering stochastic behavior for both the age-dependent and the state-dependent term. Furthermore, a Bayesian approach for automatically updating the estimates of the stochastic coefficients is developed to combine information from a degradation database with real-time CM information. With the normal assumption for the prior distribution of those stochastic coefficients, the updates have explicit formulas. This implies that each update can be performed with a single computation, which leads to an extremely fast and simple updating procedure.

We believe that the comparison results presented in Section 5 clearly indicate the value of using the improved random-coefficient autoregressive model with time effect and the Bayesian approach to incorporate real-time CM information.

The dynamic CBM policy is optimized using a SMDP framework. This optimization approach is based on [23] but surpasses it by considering real-time CM information. We have demonstrated through a case study of GaAs lasers that the dynamic control-limit maintenance policy is more sensitive to the urgency of preventive maintenance when the functioning system exhibits distinct difference from the degradation database. Moreover, using this SMDP framework, we have also explored the possibility of using SMDP to estimate mean residual life for a functioning system, for the first time in literature. The comparison between Monte Carlo simulation-based approach and the SMDP-based approach verifies the feasibility of the latter approach. However, we note that the computation cost of the SMDP-based approach hinders its independent application in residual life estimation problems. Only when it is combined with optimization problems of dynamic maintenance policy could its efficiency be revealed. Further research topics include developing residual life distributions for this random-coefficient autoregressive model and exploring more dynamic CBM policies based on this model. Extension of the model to describe and control partially observable degrading processes with soft failures is also a suitable one.

Acknowledgement

We would like to thank the National Natural science Foundation of China (Grant No. 71701008) for supporting this research.

References

1. Benyamini Z, Yechiali U. Optimality of control limit maintenance policies under nonstationary deterioration. Probability in the Engineering and Informational Sciences 1999; 13(1): 55–70, https://doi.org/10.1017/S026996489913105X.
2. Bergquist B, Soderholm P. Data analysis for condition-based railway in- frastructure maintenance. Quality & Reliability Engineering International 2015; 31(5): 773-781, https://doi.org/10.1002/qre.1634.
3. Besnard F, Bertling L. An approach for condition-based maintenance opti mization applied to wind turbine blades. IEEE Transactions on Sustainable Energy 2010; 1(2): 77–83, https://doi.org/10.1109/TSTE.2010.2049452.
4. Chen D, Trivedi K S. Optimization for condition-based maintenance with semi-Markov decision process. Reliability Engineering & System Safety 2005; 90(1): 25–29, https://doi.org/10.1016/j.ress.2004.11.001.
5. Chen N, Ye Z S, Xiang Y, Zhang L. Condition-based maintenance using the inverse Gaussian degradation model. European Journal of Operational Research 2015; 243(1): 190–199, https://doi.org/10.1016/j.ejor.2014.11.029.
6. Elwany A H, Gebraeel N Z, Maillart L M. Structured replacement policies for components with complex degradation processes and dedicated sensors. Operations research. 2011; 59(3): 684–695, https://doi.org/10.1287/opre.1110.0912.
7. Gebraeel N Z, Pan J. Prognostic degradation models for computing and updating residual life distributions in a time-varying environment. IEEE Transactions on Reliability 2008; 57(4): 539–550, https://doi.org/10.1109/TR.2008.928245.
8. Giorgio M, Guida M, Pulcini G. A new class of Markovian processes for deteriorating units with state dependent increments and covariates. IEEE Transactions on Reliability 2015; 64(2): 562–578, https://doi.org/10.1109/TR.2015.2415891.
9. Giorgio M, Guida M, Pulcini G. An age-and state-dependent Markov model for degradation processes. IIE Transactions 2011; 43(9): 621–632, https://doi.org/10.1080/0740817X.2010.532855.
10. Giorgio M, Pulcini G. A new state-dependent degradation process and related model misidentification problems. European Journal of Operational Research 2018; 267, https://doi.org/10.1016/j.ejor.2017.12.038.
11. Jiang R, Yu J, Makis V. Optimal Bayesian estimation and control scheme for gear shaft fault detection. Computers & Industrial Engineering 2012; 63(4): 754–762, https://doi.org/10.1016/j.cie.2012.04.015.
12. Kaiser K A, Gebraeel N Z. Predictive maintenance management using sensor-based degradation models. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 2009; 39(4): 840–849, https://doi.org/10.1109/TSMCA.2009.2016429.
13. Kim M J, Jiang R, Makis V, Lee CG. Optimal Bayesian fault prediction scheme for a partially observable system subject to random failure. European Journal of Operational Research. 2011; 214(2): 331–339, https://doi.org/10.1016/j.ejor.2011.04.023.
14. Kim M J, Makis V. Optimal maintenance policy for a multi-state deteriorating system with two types of failures under general repair. Computers & Industrial Engineering 2009; 57(1): 298–303, https://doi.org/10.1016/j.cie.2008.11.023.
15. Lin D, Wiseman M, Banjevic D, Jardine A K S. An approach to signal processing and condition-based maintenance for gearboxes subject to tooth failure. Mechanical Systems and Signal Processing 2004; 18(5): 993–1007, https://doi.org/10.1016/j.ymssp.2003.10.005.
16. Meeker W Q, Escobar L A. Statistical methods for reliability data. John Wiley & Sons 2014.
17. Moghaddass R, Zuo M J. An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process. Reliability Engineering & System Safety 2014; 124: 92–104, https://doi.org/10.1016/j.ress.2013.11.006.
18. Papakonstantinou K G, Shinozuka M. Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory. Reliability Engineering & System Safety 2014; 130: 202–213, https://doi.org/10.1016/j.ress.2014.04.005.
19. Papakonstantinou K G, Shinozuka M. Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part II: POMDP implementation. Reliability Engineering & System Safety 2014; 130: 214–224, https://doi.org/10.1016/j.ress.2014.04.006.
20. Ross SM. Introduction to probability models. Academic press 2014.
21. Si X S, Wang W, Hu C H, Zhou D H, Pecht MG. Remaining useful life estimation based on a nonlinear diffusion degradation process. IEEE Transactions on Reliability 2012; 61(1): 50–67, https://doi.org/10.1109/TR.2011.2182221.
22. 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(1): 167–190, https://doi.org/10.1109/TR.2014.2299151.
23. Tang D, Makis V, Jafari L, Yu J. Optimal maintenance policy and residual life estimation for a slowly degrading system subject to condition monitoring. Reliability Engineering & System Safety 2015; 134: 198–207, https://doi.org/10.1016/j.ress.2014.10.015.
24. Tang D, Yu J. Optimal replacement policy for a periodically inspected system subject to the competing soft and sudden failures. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2015; 17 (2): 228-235, http://dx.doi.org/10.17531/ein.2015.2.9.
25. Tijms H C. Stochastic models. John Wiley and sons 1994.
26. Van Noortwijk J M. A survey of the application of gamma processes in maintenance. Reliability Engineering & System Safety 2009; 94(1): 2–21, https://doi.org/10.1016/j.ress.2007.03.019.
27. Zhang Z X, Si X S, Hu C H. An age- and state-dependent nonlinear prognostic model for degrading systems. IEEE Transactions on Reliability 2015; 64(4): 1214–1228, https://doi.org/10.1109/TR.2015.2419220.
28. Zhao X, Fouladirad Mi, Berenguer C, Bordes L. Condition-based inspection/ replacement policies for non-monotone deteriorating systems with environmental covariates. Reliability Engineering & System Safety 2010; 95(8): 921–934, https://doi.org/10.1016/j.ress.2010.04.005.

29. Zhou Y, Huang M. Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model. Microelectronics Reliability 2016; 65: 265–273, https://doi.org/10.1016/j.microrel.2016.07.151.

Diyin TANG
Wubin SHENG
Jinsong YU
School of Automation Science and Electrical Engineering
Beihang University
Beijing, China, 100191

E-mail: tangdiyin@buaa.edu.cn, shengwubin@buaa.edu.cn, yujs@buaa.edu.cn

Article citation info:
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, http://dx.doi.org/10.17531/ein.2018.4.11.

Karol ANDRZEJCZAK
Marek MŁYŃCZAK
Jarosław SELECH

POISSON-DISTRIBUTED FAILURES IN THE PREDICTING OF THE COST OF CORRECTIVE MAINTENANCE

POISSONOWSKIE STRUMIENIE USZKODZEŃ W PROGNOZOWANIU KOSZTÓW OBSŁUG KOREKCYJNYCH FLOTY POJAZDÓW*

Maintaining high efficiency of using the fleet of public mass transport vehicles puts many challenges ahead of the operator. Among them, when planning periodic operational activities, the operator should take into account the assessment of possible unexpected vehicle failures and the costs of their removal under the so-called corrective maintenance. Due to the random nature of vehicle breakdowns, knowledge about stochastic processes is necessary to maintain their efficient and safe operation. The research problem formulated in the title meets these needs. Therefore, the costs of corrective maintenance of vehicles are modelled, i.e. the costs that are not included in the scheduled maintenance costs and are not related to preventive maintenance. The costs of corrective maintenance and the costs of replacement of damaged parts are unexpectedly created at random moments of operating means of transport, usually between scheduled maintenance. Due to the variety of failure processes of individual parts of the vehicle, the methods and applications of stochastic modelling for simple failures modelled by the Poisson process are presented in this paper. The basis for the application of the presented methods is to identify those parts of the vehicle that are damaged in accordance with this process. The assessment of parameters of failure processes of individual vehicle parts is made on the basis of the operational database of a homogeneous fleet of vehicles operated for 5 years. The operational database is dynamically updated with new events. On the basis of actual data on corrective maintenance of a distinguished group of damaged parts of vehicles, the possibilities and limitations of practical applications of the Poisson process to assess the risk of incurring costs in the further process of fleet operation were shown.

Keywords: *Poisson-distributed failures, corrective maintenance costs, vehicle fleet.*

Utrzymanie wysokiej efektywności użytkowania floty pojazdów publicznego transportu masowego stawia przed operatorem wiele wyzwań. Wśród nich planując okresowe działania eksploatacyjne operator powinien uwzględniać oceny możliwości pojawiania się niespodziewanych awarii pojazdów oraz kosztów ich usuwania w ramach tak zwanych obsług korekcyjnych. Ze względu na losowy charakter awarii pojazdów dla utrzymania efektywnej i bezpiecznej ich eksploatacji niezbędna okazuje się wiedza dotycząca procesów stochastycznych. Sformułowany w tytule problem badawczy wychodzi naprzeciw tym potrzebom. Modelowane są więc koszty obsług korekcyjnych pojazdów, tj. koszty, które nie są uwzględniane w ich przeglądach planowych i nie obciążają prewencyjnych napraw. Koszty robót korekcyjnych i koszty wymiany uszkodzonych części powstają niespodziewanie w losowych chwilach użytkowania środków transportu, zwykle między okresowymi przeglądami. W niniejszej pracy, ze względu na różnorodność procesów uszkadzania poszczególnych części pojazdu, przedstawiono metody i zastosowania stochastycznego modelowania dla prostych strumieni uszkodzeń, których modelem jest proces Poissona. Podstawą zastosowania przedstawionych metod jest zidentyfikowanie tych części pojazdu, które uszkadzają się zgodnie z tym procesem. Oceny parametrów procesów uszkodzeń poszczególnych części pojazdów dokonywane są na podstawie bazy danych eksploatacyjnych jednorodnej floty pojazdów eksploatowanych przez 5 lat. Wraz z upływem czasu baza danych eksploatacyjnych jest dynamicznie uzupełniana o nowe zdarzenia. Na podstawie danych rzeczywistych o obsługach korekcyjnych wyróżnionej grupy uszkadzających się części pojazdów pokazano możliwości i ograniczenia praktycznych zastosowań procesu Poissona do oceny ryzyka poniesienia kosztów w dalszym procesie eksploatacji floty.

Słowa kluczowe: *poissonowski strumień uszkodzeń, koszty obsług korekcyjnych, flota pojazdów.*

1. Introduction

The design and construction of complex technical objects that meet high requirements in terms of energy consumption, cost-consumption, ecology, safety, availability and functionality require obtaining extensive, but also detailed knowledge concerning, among others, forecasting the frequency of failures occurring during their use [2, 5, 18, 21, 22, 26]. Due to the possible consequences, designers, manufacturers and end users of equipment and technical objects try to minimise the possibility of various types of failures appearing during their operation. In order to minimise the costs of removing failures and ensuring the safety of using a technical object, the designer needs to know why and which adverse events may occur during the operation phase of the object. This knowledge is necessary for optimisation of the life cycle cost (LCC) of the object and is made available by the constant flow of information from the various phases of the existence of a technical object understood in the Agile Systems categories [14, 21, 27]. As a result, it reduces the vulnerability to such events and makes the

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

object immune to their consequences by designing redundant systems and organising preventive actions that reduce unit operating costs [18, 24]. Despite the fact that all the requirements regarding i.a. operational availability of the technical object are taken into consideration during the designing phase, random failures are unavoidable during its operation. These failures are removed as part of unscheduled corrective maintenance (CM), which reduces availability and generates additional, unplanned operating costs.

The problem of failures is particularly important for the operator planning long-term use of the fleet of urban transport vehicles [17], [25]. For such an operator, the main cost factor in the life cycle costs, apart from the cost of purchase, are costs incurred in the vehicle operation phase [3, 4, 21]. A significant share in these costs is the cost of corrective maintenance, i.e. unplanned maintenance costs that must be incurred in order to restore the operational availability of the vehicle after incidental failures or accidents. Corrective maintenance must be carried out immediately, as the loss of operational availability of each vehicle affects the rostering of the vehicle fleet and generates costs of replacing damaged vehicles. These arguments justify the need to conduct reliability tests regarding the process of losing availability of fleet vehicles and the costs of their corrective maintenance. The present work is devoted to the development of research on these needs.

Reliability models that take into account the occurrence of random failures are usually based on stochastic processes. In particular, in the predictive problems of system reliability analysis, they constitute strong tools [1, 10, 11, 16, 19]. The classic problem is still valid: how system reliability can be improved using mathematical modelling. In this area, many interesting articles and books have already been written [7, 12, 13, 19, 20, 26]. Applications of mathematics, in particular probability theory and statistics, for problems of operation and maintenance of technical objects cover such issues as: planning of active and passive experiments, inventory management, analysis of data about failures [6, 8], preventive maintenance policy optimisation, maintenance costs analysis, management of operation and maintenance processes. In the paper [20] as part of the Performance Based Logistics (PBL) approach, the authors proposed a new stochastic model to determine the annual intensity of repairs of critical aircraft components. This model can be used to plan the base inventory level and the capacity of maintenance plants.

In [13] the authors formulated the conditions under which preventive replacements increase the availability of machine operation systems and the income per unit of their operation time. The aircraft maintenance process was analysed in [23] in accordance with the lean method. Optimisation of logistics processes is considered the most promising way to reduce maintenance time and cost of spare parts.

For complex technical objects such as vehicles it is not possible to use only one probability distribution for times between failures that cause the object to lose its operational availability. The reason for this is that the failure of complex objects is caused by hundreds of various components, which are characterised by various failure mechanisms. The estimation of parameters of vehicle failure processes caused by failures of various parts was made by the authors of this paper based on a long-term operational database for a certain fleet of urban transport vehicles. The initial reliability analysis of historical failure processes allowed to distinguish several types of them. The stochastic model of the simplest of them is the Poisson failure process. In this study, the authors present the results of research on the applicability of this stochastic process to modelling the processes of losing operational availability of vehicles and to assessment of the risk of incurring the corrective maintenance costs. The research concerns the most frequently damaged parts of those that generate high costs of corrective maintenance. Poisson processes as one of the simplest ones are a good basis for developing further research on the prediction of the costs of corrective maintenance of other failure process types. It is worth noting that the operational database was used in [3] by the authors to study the significance of the differences in the average costs of corrective maintenance in the distinguished periods of fleet use.

This study consists of six parts. Section 2 presents the research problem and the research method. The Poisson method of modelling the vehicle corrective maintenance process together with the theoretical properties of this process was developed in section 3. Section 4 is devoted to the use of this method in the reliability and cost analysis for the event recorder system. Section 5 deals with the methods of assessing the risk of incurring corrective maintenance costs for a distinguished group of parts in the further period of vehicle operation. In section 6 a case study was carried out for an identified group of parts resulting in the loss of operational availability of vehicles. The entire study ends with a summary and conclusions.

2. Research problem and research method

The subject of the research is the cost of corrective maintenance of urban rail vehicles used for passenger transport. Vehicles of the tested fleet are homogeneous and the conditions of their operation are comparable. The vehicles covered by the research are repairable technical objects with high complexity and long life cycle. Maintaining a high level of vehicle safety and availability is provided by scheduled technical inspections during which preventive and predictive maintenance is carried out [22].

Despite compliance with scheduled maintenance, vehicles during operation at random times are subject to an unexpected failures, as a result of which they usually lose their ability to perform transport functions. Vehicle failures are caused by failures of certain parts of the vehicle. During 5 years, the vehicles of the tested fleet have been damaged several thousand times due to failures of several hundred different parts. In order to restore the operational availability of the vehicle, damaged parts are replaced by new ones as part of the corrective maintenance. Thus, the costs of corrective maintenance are modelled, i.e. costs that are not included in scheduled and preventive maintenance costs.

Corrective maintenance can be classified according to different criteria. For the operator of a fleet of urban transport vehicles, the cost criterion is important, determining who bears the costs of repairing damaged parts of the vehicle, i.e. the operator, a supplier, a part manufacturer, a guilty party, an insurer. The research is based on the fleet of a new type of vehicles covered by a warranty agreement between the operator and the supplier, which includes also further post-warranty maintenance. The problem formulated in the title of this paper consists in developing a method for assessing the risk of costs incurred by the supplier during selected corrective maintenance procedures. The stochastic Poisson processes [1, 12, 19], taking into account a repair costs table, were used to solve this research problem.

The basis for the applied stochastic modelling of cost processes are the theorems and properties of a homogeneous Poisson process. The definitions and variable naming convention used were taken from the papers [1, 6, 10, 12, 19, 22]. An interesting example of using the Poisson process to model the number of accidents in the Baltic Sea, in order to increase the safety of navigation is presented in the paper [11].

3. The Poisson model of the vehicle corrective maintenance process

The number of failures to a vehicle during a specified period of its operation or in a specified range of its mileage changes randomly for subsequent periods of operation or subsequent intervals of the vehicle's mileage. The theory of stochastic processes enables modelling a random evolution of the number of failures and corrective maintenance costs in time or in mileage of a vehicle. The Poisson process and its extension [1, 12, 19] play a key role in building the maintenance counting models. These processes enable the construction of models of vehicle failure counting for both specific types of failures and for the assessment of the risk of incurring the costs of restoring the operational availability by the supplier.

If the conditional probability of a vehicle failure due to failure of its i-th part is constant for mileage from the interval $(l, l+\Delta l]$, regardless of the mileage of the vehicle with this part, provided that until the mileage of l this part was in working order, the random process of loss of operational availability of the vehicle caused by subsequent failures of its i-th part is a homogeneous Poisson process. Let G means the set of those parts of the vehicle that are being damaged in accordance with the described Poisson process. We assume that after the vehicle failure the damaged parts are replaced by new ones during corrective maintenance. In addition, we assume that the total cost c_i of materials and labour of the maintenance related to the replacement of the i-th part is specified in the supplier's cost table.

As a measure of the risk of incurring the cost of repairing the vehicle due to failure of the i-th part of the group G by the mileage of $(l_i, l_i+\Delta l_i)$ we accept the product $c_i\lambda_i\Delta l_i$, where λ_i it is the failure intensity of the i-th part during the specified period of vehicle use. In order to assess the level of risk of incurring costs, it is necessary to estimate the conditional probability of a vehicle failure due to the failure of the i-th part and the total costs of corrective maintenance related to this failure. We make a simplifying assumption that the cost of corrective maintenance related to the replacement of a certain part is constant in the considered periods of vehicle operation. With the assumptions made, the vehicle mileages between subsequent maintenance related to the replacement of the i-th part from the group G have the same exponential distribution with the parameter λ_i, i.e. a cumulative distribution function (CDF) of the vehicle mileage with i-th part is given by the formula $F_i(l) = 1 - \exp(-\lambda_i l)$, $l > 0$. Parameter λ_i is the failure intensity of the i-th part for a fixed vehicle mileage unit, for which we take 1 km. The arrivals of vehicle corrective maintenance caused by failure of the i-th part is a homogeneous Poisson process HPP (λ_i) with parameter λ_i.

Let $L_{i;j}$, $j = 1,2,\ldots$ denote random mileage of the vehicle at which the j-th corrective maintenance caused by replacement of the i-th part of the group G takes place and let the process $\{N_i(l),\ l \geq 0\}$ count vehicle maintenance to mileage l (in km) due to replacements of the i-th part. For a set mileage $l \geq 0$ random number $N_i(l)$ of corrective maintenance of the vehicle due to failure of the i-th part of the group G has a Poisson distribution, i.e.

$$\Pr\left(N_i(l) = j\right) = \frac{(\lambda_i l)^j}{j!} e^{-\lambda_i l}, \tag{1}$$

what we denote as $N_i(l) \sim \text{Poisson}(\lambda_i l),\ i \in G$.

Random variables $L_{i;j}$, $j = 1,2,\ldots$ determine the vehicle's mileage for subsequent corrective maintenance due to replacement of the i-th part. Assuming that $L_{i;0} = 0$, we can write a difference:

$$X_{i;j} \stackrel{\text{def}}{=} L_{i;j} - L_{i;j-1},\ j = 1,2,\ldots \tag{2}$$

Difference $X_{i;j}$ for $j = 1$ means the mileage of the vehicle for the first corrective maintenance, and for $j \geq 2$ means vehicle mileage between successive corrective maintenance, still due to replacement of the i-th part. With all the assumptions made and within the naming convention used, all random variables $X_{i;1}, X_{i;2}, \ldots$ have the same exponential distribution with the parameter λ_i what we symbolically denote as $X_{i;j} \sim \text{EXP}(\lambda_i)$ for $j = 1,2,\ldots$. A consequence of the adopted assumptions regarding the corrective maintenance processes related to the replacement of parts from the group G is the possibility to apply the following properties adapted from the general theory of stochastic processes.

Property 1. Random number $N_i(l_1,l_2)$ of vehicle corrective maintenance due to replacements of the i-th part ($i \in G$), with a planned further mileage of it (l_1,l_2), has a Poisson distribution with a parameter $(l_2 - l_1)\lambda_i$, i.e.

$$N_i(l_1,l_2) \sim \text{Poisson}\left((l_2 - l_1)\lambda_i\right). \tag{3}$$

Property 2. Random cost $C_i(l_1,l_2|c_i)$ of vehicle corrective maintenance due to replacements of the i-th part during the mileage (l_1,l_2) assuming that the cost related to one replacement of this part c_i is given by the formula:

$$C_i(l_1,l_2|c_i) = c_i N_i(l_1,l_2). \tag{4}$$

The following property follows directly from property 2.

Property 3. Expected vehicle corrective maintenance cost $\mathbb{E}C_i$ and cost variance $\mathbb{D}^2 C_i$ for maintenance caused by replacements of the i-th part, with further mileage (l_1,l_2) are defined by the formulas:

$$\mathbb{E}C_i(l_1,l_2|c_i) = c_i\lambda_i(l_2 - l_1), \tag{5}$$

$$\mathbb{D}^2 C_i(l_1,l_2|c_i) = c_i^2\lambda_i(l_2 - l_1). \tag{6}$$

So the expected cost $\mathbb{E}C_i(l_1,l_2|c_i)$ is a measure of the risk of incurring the vehicle corrective maintenance costs due to replacements of the i-th part from the group G with a planned further mileage (l_1,l_2). Additionally, the variance can be applied to estimating the confidence interval for the risk of incurring these costs.

In the case of a fleet composed of n homogeneous vehicles operated under the same conditions, expected vehicle corrective maintenance cost $\mathbb{E}C_i(l_1,l_2;n|c_i)$ due to replacements of the i-th part of the group G with planned further mileage of vehicles (l_1,l_2) is expressed by the formula (7):

$$\mathbb{E}C_i(l_1,l_2;n|c_i) = nc_i\lambda_i(l_2 - l_1). \tag{7}$$

An alternative measure of the risk of incurring the costs of corrective maintenance due to replacements of the i-th part of the group G for the fleet of n vehicles is the most probable cost of corrective maintenance C_i^{Mo} formulated in property 4 assuming that $n\lambda_i(l_2 - l_1)$ is not an integer.

Property 4. Most probable cost (modal cost) $C_i^{Mo}(l_1,l_2;n|c_i)$ of corrective maintenance for n vehicles of a homogeneous fleet due to exchanges of the i-th part, with further mileage of the vehicles (l_1,l_2) is given by the formula:

$$C_i^{Mo}(l_1,l_2;n|c_i) = c_i\left\lfloor n\lambda_i(l_2 - l_1)\right\rfloor, \tag{8}$$

where $\lfloor \cdot \rfloor$ means the floor function.

Another very useful property of the maintenance process is determining the distribution of the vehicle mileage of the tested fleet to a given number of maintenance m .

Property 5. The mileage of the vehicle $L_{i;m}$ to its m-th corrective maintenance due to replacement of the i -th part is a random variable with the Erlang distribution, expressed as:

$$L_{i;m} \sim \mathrm{ERL}\left(m;\lambda_i\right), \tag{9}$$

i.e. the probability density function of a random vehicle's mileage $L_{i;m}$ can be expressed as:

$$f_{i;m}\left(l\right) = \frac{\lambda_i^m l^{m-1} e^{-\lambda_i l}}{\left(m-1\right)!}, \ l > 0 \tag{10}$$

Hence for a random vehicle mileage $L_{i;m}$ it is possible to designate a cumulative distribution function $F_{i;m}\left(l\right)$ and its basic characteristics, i.e. the expected value $\mathbb{E}L_{i;m}$, mode $\mathrm{Mo}\left(L_{i;m}\right)$ and variance $\mathbb{D}^2 L_{i;m}$ of the vehicle mileage to the m -th corrective maintenance due to replacement of the i -th part:

$$F_{i;m}\left(l\right) = 1 - \frac{\sum_{k=0}^{m-1} e^{\lambda_i l} \left(\lambda_i l\right)^k}{k!} \ \text{dla } l > 0 \tag{11}$$

$$\mathbb{E}L_{i;m} = \frac{m}{\lambda_i} \tag{12}$$

$$\mathrm{Mo}\left(L_{i;m}\right) = \frac{m-1}{\lambda_i} \ \text{ dla } m \ge 2 \tag{13}$$

$$\mathbb{D}^2 L_{i;m} = \frac{m}{\lambda_i^2} \tag{14}$$

Using the properties (12) and (14) and the cost table of corrective maintenance, it is possible to determine the expected cost and cost variance.

4. Assessment of the risk of incurring costs due to replacement of the event recorder system

The operational database applies to the fleet of $n = 45$ new vehicles used for 5 years. During this time, out of several thousand parts of which the vehicle is made, more than 500 have been damaged. The fleet supplier has granted a warranty for this time and is considering the possibility of extending it for further years. At the time under study, vehicles reached mileages of around 300,000 km. From the database, the mileages of the parts between failures have been designated as vehicle mileage with a given part.

The analysis of reliability based on the operational database shows that the mileage of the vehicles between the loss of their operational availability caused by replacements of individual parts within corrective maintenance belong to probability distribution families of the types: gamma, Weibull, exponential, normal, lognormal.

The basic problem that had to be solved was to identify these parts which force corrective maintenance of vehicles and, at the same time, meet the assumption of the Poisson-distributed failures. For this purpose, the hypotheses on the exponentiality of the distribution of mileage between failures for selected parts were verified. Weibull ++ software was used for testing, focusing on this stage of the research primarily on parts from the 15-th vehicle construction group. This group includes electronic and electrical devices, the failure of which is immediately detected and their replacement is relatively fast within the scope of corrective maintenance, and the aging processes of the selected parts are marginal. In vehicles of the tested fleet, this part group includes: an event recorder system, a recorder module, a main monitoring module, a drive controller and a pressure aggregate from 11-th construction group. These parts have been subjected to a reliability analysis.

To assess the risk of incurring corrective maintenance costs of the vehicle fleet, an event recorder system was chosen first. This system is intended for the registration and monitoring of electric meter systems and for recording events regarding emerging hazards and failures of urban transport vehicles. The basic element of the event recorder system is the parameter recorder shown in Fig. 1. The cognitive goal is to assess the risk of incurring the corrective maintenance costs of all fleet vehicles due to the failure of this recorder.

Fig. 1. Parameter recorder.

The recorder saves on the memory card **analogue signals,** such as mileage counter, driving speed, traction current, traction network voltage, control circuit voltage and **logic signals,** such as condition of control devices, feedback signals for braking systems activation, wagon door status, switch changer state, slip signal and other signals important for safety reasons. During the tested period of operation, corrective maintenance at the supplier's cost due to exchanges of this device was registered 11 times. The total costs of the maintenance amounted to approximately 237,000 PLN, and the average cost of one maintenance amounted to approximately 21,500 PLN.

In the reliability analysis, in addition to the mileages of the 11 event recorder systems subjected to corrective maintenance, 45 censored observations concerning the mileage of recorders undamaged on the day ending the research were also included.

On the basis of the vehicle maintenance database provided by the vehicle supplier, failure data for a selected group of parts were developed. For the needs of the conducted reliability tests, data on the mileage of the damaged i -th part for the whole fleet was compiled in the form of pairs $\left(l_k;\delta_k\right), k = 1,2,\ldots,n_i$, where l_k is the mileage of the k -th instance of the i -th part (expressed in kilometres), n_i is the number of data related to the i -th part and

$$\delta_k = \begin{cases} 1, & \text{if } l_k \text{ is the observed mileage,} \\ 0, & \text{if i } l_k \text{ is the censored mileage.} \end{cases} \tag{15}$$

For the event recorder system, from the operational data for the entire fleet the following pairs were obtained:

(268,707; 0),	(166,033; 1),	(106,638; 0),	(360,404; 0),	(114,686; 1),	(128,846; 0),
(284,834; 0),	(369,436; 0),	(314,522; 0),	(186,073; 1),	(153,834; 0),	(311,679; 0),
(342,241; 0),	(328,494; 0),	(354,100; 0),	(292,038; 0),	(339,492; 0),	(292,266; 0),
(319,088; 0),	(352,654; 1),	(257,803; 1),	(79,301; 0),	(310,513; 0),	(305,087; 0),
(89,378; 1),	(245,637; 0),	(308,114; 0),	(328,878; 0),	(314,589; 0),	(91,939; 1),
(176,804; 0),	(319,733; 0),	(328,480; 0),	(69,514; 1),	(258,654; 0),	(287,122; 0),
(189,518; 1),	(107,834; 0),	(271,868; 0),	(244,005; 0),	(284,525; 0),	(207,679; 0),
(314,146; 0),	(271,384; 0),	(316,331; 0),	(286,225; 0),	(58,298; 1),	(231,227; 0),
(254,974; 0),	(290,198; 0),	(299,174; 0),	(80,172; 1),	(184,615; 0),	(310,800; 0),
(339,499; 0),	(315,371; 0).				

The research covered 56 instances of the event recorder system, of which 45 were in good working order at the time the tests were completed and their observations of the mileage were cut off. The test carried out at the 5% significance level did not give grounds for rejecting the hypothesis on the exponentiality of the vehicles' mileage between their failures caused by failures of the event recorder systems. Based on the operational data, the failure intensity was estimated $\hat{\lambda} \approx 0{,}0000009463[1/\text{km}]$. Assuming that during the next year of operation of the fleet, the failure intensity of the used recorder systems will not change significantly, it is possible to determine the risk of incurring the replacement costs. The annual mileage of vehicles in the fleet under investigation is about 60,000 km. Hence the risk of incurring costs measured by the expected cost of corrective maintenance caused by the replacement of this system with the mileage of vehicles from $l_1 = 300000$ [km] to $l_2 = 360000$ [km] is about 1,224 [PLN], what with the fleet of 45 vehicles makes the amount of about 55,081 [PLN]. However, the most likely total corrective maintenance cost for the fleet of 45 vehicles due to replacement of this system determined from the formula (8) is 43,000 [PLN]. The expected mileage of the vehicle for the third maintenance related to the replacement of this system is 3,170,242 km.

5. Costs of corrective maintenance of the selected group of damaged parts of the vehicle

Very practical properties of the Poisson process in assessing the cost of corrective maintenance of the vehicle due to the independent failures to parts from the selected group G is both the property of the superposition of Poisson's processes as well as the property of the Poisson process decomposition [1], [12].

Property 6 (superposition of Poisson's processes). If the numbers of corrective maintenance of the vehicle due to the replacement of parts from the set G for the mileage of $l > 0$ are independent and have Poisson distributions, i.e. $N_i(l) \sim \text{Poisson}(\lambda_i l)$, $i \in G$, than the total number of corrective vehicle maintenance $N_G(l_1,l_2)$ due to the replacement of parts from the set G with a planned further mileage of (l_1,l_2) has a Poisson distribution with a parameter $(l_2 - l_1)\sum_{i\in G}\lambda_i$, i.e.

$$\Pr\left(N_G(l_1,l_2) = n\right) = \frac{\left((l_2 - l_1)\sum_{i\in G}\lambda_i\right)^n}{n!}\exp\left((l_2 - l_1)\sum_{i\in G}\lambda_i\right). \quad (16)$$

Using property 4, it is now possible to designate the total random cost $C_G\left(l_1,l_2 \middle| c_i, i \in G\right)$ of vehicle corrective maintenance due to the replacement of parts from the group G for the mileage (l_1,l_2)

$$C_G\left(l_1,l_2 \middle| c_i, i \in G\right) = \sum_{i\in G} C_i\left(l_1,l_2 \middle| c_i\right). \quad (17)$$

Hence the expected total cost $\mathbb{E}C_G\left(l_1,l_2 \middle| c_i, i \in G\right)$ and the variance of the total cost $\mathbb{D}^2 C_G\left(l_1,l_2 \middle| c_i, i \in G\right)$ are as follows:

$$\mathbb{E}C_G\left(l_1,l_2 \middle| c_i, i \in G\right) = \sum_{i\in G} c_i\lambda_i\left(l_2 - l_1\right), \quad (18)$$

$$\mathbb{D}^2 C_i\left(l_1,l_2 \middle| c_i\right) = \sum_{i\in G} c_i^2\lambda_i\left(l_2 - l_1\right). \quad (19)$$

Property 7 (decomposition of the Poisson process). If the total number $N_G(l)$ of corrective vehicle maintenance for mileage $l > 0$ due to independent replacements of parts from the set G is the Poisson process with the failure intensity λ_G and the probability of vehicle corrective maintenance due to a failure of the i-th part at this time equals to p_i, wherein $N_G(l) = \sum_{i\in G} N_i(l)$ and $\sum_{i\in G} p_i = 1$, than:

$$\Pr\left(N_i(l_1,l_2) = n_i, i \in G\right) = \prod_{i\in G}\frac{\left(p_i\lambda\left(l_2 - l_1\right)\right)^{n_i}}{n_i!}\exp\left(-p_i\lambda\left(l_2 - l_1\right)\right) \quad (20)$$

Equation (20) allows to determine the probability distribution of the number of replacements of individual parts from the set G with the planned further mileage of the vehicle (l_1,l_2).

By using property 7, it is also possible to determine the expected cost of corrective maintenance caused by independent failures to a part from the set G. The risk of incurring the corrective maintenance costs for the fleet of n vehicles (based on fleet operational data to mileage l_0) related to the replacement of parts from the set G with a planned further mileage (l_1,l_2) [km] ($0 < l_0 \le l_1 < l_2 < \infty$) – assuming the criterion of expected costs – is expressed by the following formula:

$$\mathbb{E}C_G\left(l_1,l_2,n \middle| c_i, i \in G\right) = \sum_{i\in G} c_i p_i \lambda n\left(l_2 - l_1\right) \quad (21)$$

If the criterion of the most likely cost is taken to assess the risk of incurring costs, then:

$$C_G^{Mo}\left(l_1,l_2,n \middle| c_i, i \in G\right) = \sum_{i\in G} c_i n\,\lambda_i\left(l_2 - l_1\right) \quad (22)$$

or:

$$C_G^{Mo}\left(l_1,l_2,n \middle| c_i, i \in G\right) = \sum_{i\in G} c_i n\,p_i\lambda\left(l_2 - l_1\right) \quad (23)$$

The presented methods of assessing the risk of incurring the costs of corrective maintenance require meeting quite strong assumptions regarding the distribution of failures. The reliability tests carried out show that only a small group of damaged parts of the vehicle meet these assumptions. However, the presented stochastic cost forecasting methods are a good basis for developing stochastic methods of cost forecasting for vehicle parts that meet weaker assumptions.

6. Case study for the selected group of parts

From statistical surveys of corrective maintenance of the most expensive parts from the 15-th construction group, the failure rate of the drive controller turned out to be the closest to the Poisson model. Additionally, the failure rate of the recorder module and the main monitoring module did not show a significant difference with the Poisson model, characterised by a constant failure intensity. In contrast, the failure process of the assembly set was characterised by a significantly decreasing intensity, so it did not meet the accepted requirement. Among the costly corrective maintenance from outside of the 15-th construction group, the failure rate of a pressure aggregate was similar to the Poisson model. The identified parts, which are characterised by failure rate that does not differ significantly from the Poisson model, were assigned to the group G. The theoretical basis of risk assessment methods for the corrective maintenance cost, which were described in the previous section, were applied to the identified group of parts G. Group of parts G is composed of a pressure aggregate, an event recorder system, a recorder module, a main monitoring module and a drive controller. There is only one instance of each of these parts inside one vehicle. Statistics on the number of corrective maintenance and its costs related to the replacement of these parts for the fleet of 45 vehicles are presented in Table 1.

Table 1. Data regarding costs of corrective maintenance due to replacement of selected parts of the vehicle (own work)

No.	Part name	Construction group	Number of maintenance	Total maintenance costs [PLN]
1	Pressure aggregate	11	27	610,200
2	Event recorder system	15	11	237,138
3	Recorder module	15	12	184,920
4	Main monitoring module	15	54	287,712
5	Drive controller	15	23	208,357

Table 2. Results of estimation of the parameters of the part's mileage assuming a two-parameter Weibull distribution

No.	Part name	Lower end of the interval for β	$\hat{\beta}$	Upper end of the interval for β	Lower end of the interval for η	$\hat{\eta}$	Upper end of the interval for η
1	Pressure aggregate	0.6954	1.0932	1.7187	293,870	635,599	1,145,688
2	Event recorder system	0.54656	0.9117	1.52082	450,409	1,209,913	3,250,134
3	Recorder module	0.71235	1.0414	1.52245	469,975	731,649	1,139,018
4	Main monitoring module	0.77217	0.96850	1.21475	190,126	253,444	337,850
5	Drive controller	0.62877	1.02535	1.67208	297,741	664,105	1,481,273

The results of the point and interval estimation performed for the parameters of the two-parameter Weibull distribution for selected vehicle parts are presented in Table 2. The assumption was made that the distributions of mileage between failures belong to families of two-parameter Weibull distributions with the parameters determined by the probability density function:

$$f\left(l|\beta,\eta\right)=\frac{\beta}{\eta}\left(\frac{l}{\eta}\right)^{\beta-1}\exp\left(-\left(\frac{l}{\eta}\right)^{\beta}\right)\mathbb{I}_{(0,\infty}\left(l\right), \tag{24}$$

where $\eta>0$ is a scale parameter and $\beta>0$ is a shape parameter.

For the estimation of Weibull distribution parameters with right-censored data, the maximum likelihood estimation (MLE) method was used [9]. Table 2 presents the results of point and interval estimation for Weibull distribution parameters. A 95% confidence level was assumed.

On the basis of operational data it is not possible to reject the hypotheses about the exponentiality of mileage to failure of a part from the group G on the assumed significance level of 0.05. Hence, further inference is based on the assumption of the Poisson distribution of failures. For the selected parts, the results of estimation of the expected mileage to failure, 95% confidence intervals for the expected mileage and the intensity of their failures are presented in Table 3.

Based on the estimated failure intensities and the unit costs of corrective maintenance related to replacements of the selected group of parts G, the risk of incurring corrective maintenance costs due to the replacement of these parts for the assumed further annual mileage of vehicles from the tested fleet from $l_1=300{,}000$ [km] to $l_2=360{,}000$ [km] was determined.

The estimated total cost of corrective maintenance related to the replacement of parts from the group G for one vehicle is 7,100 [PLN], which, with a fleet of 45 vehicles, amounts to almost 320,000 [PLN]. Alternatively, the forecast based on the most probable corrective maintenance cost for these parts is 274,590 [PLN] for the vehicle fleet.

Table 3. The results of the estimation of the parameters of the part mileage, assuming the exponentiality

No.	Part name	Confidence Bounds Lower	Mean Time (km)	Confidence Bounds Upper	Failure Rate [1/km]
1	Pressure aggregate	376,975	554,163	814,633	0.0000018045
2	Event recorder system	613,590	1,056,719	1,819,870	0.0000009463
3	Recorder module	508,933	765,372	1,151,023	0.0000013066
4	Main monitoring module	186,564	241,867	313,564	0.0000041345
5	Drive controller	397,799	602,362	912,119	0.0000016601

Table 4. Estimated costs of corrective maintenance of parts from the group G during the next year of operation

No.	Part name	Estimated costs		The most probable costs for the fleet
		for a vehicle	for the fleet	
1	Pressure aggregate	2,447	110,111	90,400
2	Event recorder system	1,224	55,081	43,116
3	Recorder module	1,208	54,364	46,230
4	Main monitoring module	1,322	59,477	58,608
5	Drive controller	902	40,605	36,236

7. Conclusions

Stochastic modelling of corrective maintenance costs presented in the paper, which takes into account reliability characteristics of repairable objects, such as vehicles, is an excellent method of supporting decision-making processes in their maintenance and allows for more rational use of the public transport fleet.

As a result of the conducted research, the main idea of stochastic cost forecasting of selected corrective maintenance of urban transport means was presented, which is of key importance in supporting effective management of vehicle fleet operations.

The developed methods have been implemented to assess the costs of corrective maintenance of the fleet of urban transport vehicles. Parameters of the vehicle corrective maintenance process models were estimated on the basis of the operational database containing information from the period of 5 years of operation of a homogeneous fleet of trams. To predict the costs of corrective maintenance, parts of the vehicle that meet the assumption regarding the Poisson-distributed failures were identified.

The issue of corrective maintenance costs is up-to-date due to the currently developed practical possibilities of the process perception of the activities of transport companies and is indispensable in the study of the life cycle costs of means of transport. In their final form, the presented methods will be used for supporting an IT system managing the operation and maintenance of fleet of urban transport vehicles.

The authors see further development of the conducted research in such a weakening the assumptions of the failure process models, so that the application possibilities in the field of corrective maintenance cost prediction can be broadened. Knowledge about corrective maintenance is essential in optimising preventive maintenance and reducing unscheduled vehicle downtime costs. In conclusion, it is worth noting that in recent years, intensive research has been carried out on the optimisation of strategies and methods for maintenance of technical objects [8, 15, 16, 26, 27].

Acknowledgements

Research was funded by the National Centre for Research and Development – Program for Applied Research PBS III/B6/30/2015, as well as funds for statutory activities 04/43/DSPB/0104 and 05/51/DSPB/3551.

References

1. Andrzejczak K.: Stochastic modelling of the repairable system, Journal of KONBiN 3(35) 2015, 5-14.
2. Andrzejczak K.: Methods of the forecasting in the modeling of means of transport maintenance (in Polish). Rozprawy nr 496. Wydawnictwo Politechniki Poznańskiej, Poznań 2013.
3. Andrzejczak K., Selech J.: Quantile analysis of the operating costs of public transport fleet, Transport Problems, vol. 12 (3), 103-111, 2017.
4. Andrzejczak K, Selech J., Investigating the trends of average costs of corrective maintenance of public transport vehicles, Journal of KONBiN 41(2017), 207-226. DOI 10.1515jok-2017-0011.
5. Andrzejczak K., Młyńczak M., Selech J. Assessment model of operational effectiveness related to newly operated public means of transport. Proceedings of the 27th European Safety and Reliability Conference (ESREL 2017), Portorož, Słowenia, Editors: Marko Čepin, Radim Briš, pages 3455-3461.
6. Ascher H., Feingold H.: Repairable Systems Reliability: Modeling, Inference, Misconceptions and their Causes. Marcel Decker, New York, 1984.
7. Bobrowski D.: Modele i metody matematyczne teorii niezawodności w przykładach i zadaniach. 1985, Warszawa, WNT.
8. Chen Y., Cowling P., Polack F., Remde S., Mourdjis P., Dynamic optimisation of preventative and corrective maintenance schedules for a large scale urban drainage system. European Journal of Operational Research, 257 (2), 494-510. https://doi.org/10.1016/j.ejor.2016.07.027
9. Ferreira L.A., Silva J.L. Parameter estimation for Weibull distribution with right censored data using EM algorithm. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19(2): 310–315, http://dx.doi.org/10.17531/ein.2017.2.20 .
10. Grabski F.: Stochastyczny model bezpieczeństwa obiektu w procesie eksploatacji. Problemy Eksploatacji, 1/2011 (80), 89-102.
11. Grabski F.: Nonhomogenous Poisson proces application to modeling accidents number at Baltic Sea waters and ports, Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars, Vol. 8, Number 1, 2017, 39-46.
12. Jokiel-Rokita A., Magiera R.: (2010). Parameter estimation in non-homogeneous Poisson process models for software reliability. Technical report, Wrocław University of Technology, Institute of Mathematics and Computer Science.
13. Knopik L, Migawa K. Multi-state model of maintenance policy. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (1): 125–130, http://dx.doi.org/10.17531/ein.2018.1.16.
14. Kusz A., Marciniak A., Skwarcz J. Implementation of computation process in a bayesian network on the example of unit operating costs determination. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2015; 17 (2): 266–272, http://dx.doi.org/10.17531/ein.2015.2.14.
15. 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, http://dx.doi.org/10.17531/ein.2017.1.19.
16. Lee H, Cha J H. New stochastic models for preventive maintenance and maintenance optimization. European Journal of Operational Research 2016; 255 (1): 80-90, https://doi.org/10.1016/j.ejor.2016.04.020.
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, http://dx.doi.org/10.17531/ein.2017.2.1
18. Młyńczak M. 2012. Metodyka badań eksploatacyjnych obiektów technicznych, Wrocław, Oficyna Wydawnicza Politechniki Wrocławskiej, s. 204.

19. Nakagawa T.: Stochastic Processes with Applications to Reliability Theory. Springer, 2011.
20. Nataša Kontrec, Stefan Panić, Milena Petrović, Hranislav Milošević. A stochastic model for estimation of repair rate for system operating under performance based logistics. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (1): 68–72, http://dx.doi.org/10.17531/ein.2018.1.9.
21. Nowakowski T, Młyńczak M, Werbińska-Wojciechowska S, Dziaduch I, Tubis A. Life Cycle Costs of passenger transportation system. Case study of Wroclaw city agglomeration 5. JPSRA 2014; 5: 109-120.
22. Omdahl T.P., ed.: Reliability, Availability and Maintainability (RAM) Dictionary, ASQC Quality Press, Milwaukee, Wisconsin, 1988.
23. Pogačnik B, Duhovnik J, Tavčar J. Aircraft fault forecasting at maintenance service on the basis of historic data and aircraft parameters. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (4): 624–633, http://dx.doi.org/10.17531/ein.2017.4.17.
24. Raposo H, Farinha JT, Ferreira L, Galar D. An integrated econometric model for bus replacement and determination of reserve fleet size based on predictive maintenance. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (3): 358–368, http://dx.doi.org/10.17531/ein.2017.3.6.
25. Rymarz J, Niewczas A, Krzyżak A. Comparison of operational availability of public city buses by analysis of variance. Eksploatacja i Niezawodnosc–Maintenance and Reliability 2016; 18 (3):373–378, http://dx.doi.org/10.17531/ein.2016.3.8.
26. Sharma A, Yadava G S, Deshmukh S G. A literature review and future perspectives on maintenance optimization, Journal of Quality in Maintenance Engineering 2011; 17 (1): 5 – 25, https://doi.org/10.1108/13552511111116222.
27. Stenström C, Norrbin P, Parida A, Kumar U. Preventive and corrective maintenance - cost comparison and cost-benefit analysis. Structure and Infrastructure Engineering 2016; 12 (5): 603-617, https://doi.org/10.1080/15732479.2015.1032983

Karol ANDRZEJCZAK
Poznan University of Technology
Faculty of Electrical Engineering
Piotrowo 3A, 60-965 Poznan, Poland

Marek MŁYŃCZAK
Wrocław University of Science and Technology
Faculty of Mechanical Engineering
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

Jarosław SELECH
Poznan University of Technology
Faculty of Machines and Transport
Piotrowo 3, 60-965 Poznań, Poland

E-mails: karol.andrzejczak@put.poznan.pl,
marek.mlynczak@pwr.edu.pl, jaroslaw.selech@put.poznan.pl

Article citation info:
LI J, DUAN R. Dynamic diagnostic strategy based on reliability analysis and distance-based VIKOR with heterogeneous information. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 610–620, http://dx.doi.org/10.17531/ein.2018.4.12.

Jing LI
Rongxing DUAN

DYNAMIC DIAGNOSTIC STRATEGY BASED ON RELIABILITY ANALYSIS AND DISTANCE-BASED VIKOR WITH HETEROGENEOUS INFORMATION

DYNAMICZNA STRATEGIA DIAGNOSTYCZNA Z WYKORZYSTANIEM INFORMACJI HETEROGENICZNYCH BAZUJĄCA NA ANALIZIE NIEZAWODNOŚCI I OPARTYM NA ODLEGŁOŚCIACH ALGORYTMIE VIKOR

This paper presents a dynamic diagnostic strategy based on reliability analysis and distance-based VIKOR with heterogeneous information. Specifically, the proposed method uses a dynamic fault tree (DFT) to describe the dynamic fault characteristics and evaluates the failure rate of components using interval numbers to deal with the epistemic uncertainty. Furthermore, DFT is mapped into a dynamic evidential network (DEN) to calculate some reliability parameters and these parameters together with test cost constitute a decision matrix. In addition, a dynamic diagnostic strategy is developed based on an improved VIKOR algorithm and the previous diagnosis result. This diagnosis algorithm determines the weights of attributes based on the Entropy concept to avoid experts' subjectivity and obtains the optimal ranking directly on the original heterogeneous information without a transformation process, which can improve diagnosis efficiency and reduce information loss. Finally, the performance of the proposed method is evaluated by applying it to a train-ground wireless communication system. The results of simulation analysis show the feasibility and effectiveness of this methodology.

Keywords: *reliability analysis, dynamic evidential network, VIKOR, decision matrix, heterogeneous information.*

W artykule przedstawiono dynamiczną strategię diagnostyczną, w której wykorzystuje się oryginalne informacje heterogeniczne. Metoda ta bazuje na analizie niezawodności i opartym na odległościach algorytmie VIKOR. Dokładniej, przedstawiona strategia polega na wykorzystaniu dynamicznego drzewa błędów (DFT) do opisu dynamicznych charakterystyk błędów oraz ocenie intensywności uszkodzeń komponentów przy użyciu liczb przedziałowych, co pozwala rozwiązać problem niepewności epistemicznej. Ponadto, w proponowanej metodzie, DFT zostaje odwzorowane w dynamiczną sieć dowodową (DEN) w celu obliczenia niektórych parametrów niezawodności, a parametry te wraz z kosztem badań diagnostycznych tworzą matrycę decyzyjną. Opracowana dynamiczna strategia diagnostyczna opiera się na udoskonalonym algorytmie diagnostycznym VIKOR oraz wynikach wcześniejszej diagnostyki. Algorytm VIKOR określa wagi atrybutów w oparciu o koncepcję Entropii, co pozwala wyeliminować subiektywność oceny eksperckiej i ustalić optymalną kolejność działań diagnostycznych bazując bezpośrednio na oryginalnych informacjach heterogenicznych bez konieczności ich transformacji, co może poprawić efektywność diagnozy i zmniejszyć utratę informacji. Działanie proponowanej metody oceniano poprzez zastosowanie jej do diagnostyki systemu łączności radiowej pociąg–ziemia. Wyniki analizy symulacyjnej wskazują na możliwość praktycznego wykorzystania i skuteczność omawianej metodologii.

Słowa kluczowe: *analiza niezawodności, dynamiczna sieć dowodowa, VIKOR, matryca decyzyjna, informacja heterogeniczna.*

1. Introduction

The application of high technology to the engineering system has significantly improved the performance of modern systems and at the same time greatly increased the complexity of the systems structure. Manufacture cost of these systems is too high. Once these systems fail, it will cause a great loss. Therefore, it will be extremely important to establish a fault diagnosis model based on their unique fault characteristics and develop a dynamic diagnosis strategy which can locate the fault component quickly and reduce the maintenance cost when these systems break down. Usually, fault diagnosis requires a large amount of historical fault data. However, in engineering practice, application of redundant technologies have improved the reliability of these systems, which raises some challenges in fault diagnosis. For one thing, the behaviours of components in these systems, such as failure priority, functional dependent failures, and sequentially dependent failures should be taken into account. For another, high reliability makes it extremely difficult to obtain complete fault data because these systems may still be in the early life cycle, which results in the epistemic uncertainty. Aiming at these challenges, many researchers have put forward a large number of efficient fault diagnosis methods over the last few decades. Johnson presented a sequential diagnostic method based on heuristic information search [10], which constructed a sequential test procedure to locate the failure using information theory. However, the diagnostic result was not satisfied. A novel diagnosis strategy for multi-value attribute system was proposed based on rollout algorithm, and it obtained an optimal diagnostic sequence [9]. Based on these researches, Tian et al. proposed a fault diagnostic strategy of multivalued attribute system based on growing algorithm, which chose failure states and found an appropriate test set for these states [21]. This growing algorithm could avoid the backtracking approach of traditional algorithms and obtained good diagnostic results with a high efficiency. A real-time fault diagnosis approach was presented based on reliability analysis and Bayesian networks (BN) [6]. BN was used to calculate the system reliability, and the real-time system reliability was monitored and compared with the previous values. If the deviations exceeded the preset threshold, a heuristic algorithm was used to locate the failed component which had the greatest changes between the prior probability and posterior probability. In the literature [3], a real-time fault diagnosis method for complex systems using object-oriented BN was proposed. It included an off-line BN construction phase and an on-line fault diagnosis phase. Nevertheless, these methods constructed the BN model based on the parameter learning algorithm, which needed a large amount of fault data and could not handle epistemic uncertainty. Chiremsel et al. proposed a probabilistic fault diagnosis method of safety instrumentation system using the fault tree and BN [4]. A static fault tree was used to construct the fault model of safety instrument system and was mapped into BN to calculate the importance measure which was used to design the diagnosis algorithm. Nevertheless, this method is unable to model the dynamic fault behaviours and deal with epistemic uncertainty.

For dynamic fault characteristics, Dugan introduced a DFT to model the dynamic fault behaviours and used diagnostic importance factor (DIF) to determine the diagnostic sequence [1-2]. However, this method calculated DIF based on Markov chains which had a state space explosion problem and determined the diagnosis sequence only by components' DIF which is a single attribute decision making problem, thereby influencing the diagnosis efficiency. Besides, it assumed that the failure rates of the components are expressed in defined values describing their reliability characteristics and failed to cope with the epistemic uncertainty. Although some researchers put forward interval analysis [24], the possibility theory [19, 22], imprecise probability [12], fuzzy set theory [5, 11] and evidence theory [25], these theories were only used for the reliability analysis and risk assessment and were not further applied to the fault diagnosis. Therefore, Duan et al. presented a novel fault diagnosis method based on fuzzy set and DFT analysis [8]. The fuzzy information obtained by fuzzy set theory and domain expert was transformed into quantitative information to obtain the fuzzy failure rates of components. Discrete time Bayesian Networks was used to calculate some reliability results, and an efficient diagnosis algorithm was developed based on qualitative structural information and quantitative parameters. However, it is usually difficult to determine the corresponding membership function of each language value, and this diagnosis algorithm was also a single attribute decision making problem. To overcome these limitations, multiple attributes decision-making was used in [7, 20]. However, these methods usually used the attributes with defined values and could not make decisions under uncertainty. Besides, the proposed methods dealt with the decision problems regarding one particular type of values. It was more reasonable to express the different attributes in their appropriate data types. Only a few work took into consideration the heterogeneous information [13, 23]. However, there was litter work connected with the diagnostic strategy for complex systems. Furthermore, diagnostic algorithms failed to update the diagnostic decision table according to the previous diagnosis result.

Motivated by the problems mentioned above, this paper proposes a dynamic diagnostic strategy based on reliability analysis and distance-based VIKOR, a multi-criteria decision analysis method, with heterogeneous information considering epistemic uncertainty shown in Fig. 1. A DFT is used to establish the system fault model to describe the dynamic fault characteristics. Interval numbers are used to describe the failure rate of components to deal with epistemic uncertainty. Furthermore, a DFT is converted into a DEN to obtain the reliability parameters such as DIF and risk achievement worth (RAW). In addition, DIF, RAW, test cost and previous diagnosis result are taken into account comprehensively to obtain the optimal diagnostic ranking order using a distance-based VIKOR with heterogeneous information. Finally, a train-ground wireless communication system is given to demonstrate the efficiency of this proposed method.

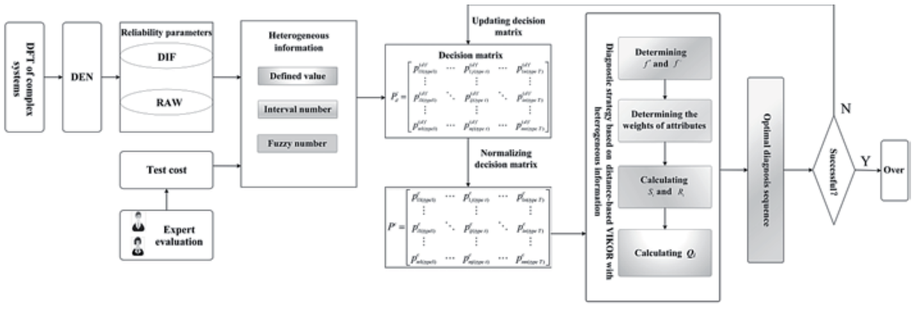


Fig. 1. A dynamic diagnostic framework based on reliability analysis and distance-based VIKOR with heterogeneous information.

The remainder of this article is organized as follows. Section 2 presents the DFT model construction and quantitative analysis of DFT based on DEN. A novel dynamic diagnostic strategy based on reliability analysis and distance-based VIKOR with heterogeneous information considering the epistemic uncertainty is given in Section 3. Section 4 is devoted to a simple illustration example of the proposed approach. Some conclusions are given in the final section.

2. DFT analysis

2.1. Model Construction of DFT

Fault tree is a deductive method to decide the potential causes that may cause the occurrence of a predefined undesired event, generally denoted as the top event. DFT extends a static fault tree to describe the dynamic failure behaviours such as priorities of failure events, spares, and sequence-dependent events. Dynamic gates in DFT include the priority AND gate, the functional dependency gate (FDEP), the sequence enforcing gate, the cold, hot, and warm spare gates. The model construction of the fault tree usually requires an in depth knowledge of the system and its components. It includes the construction of a network topology and the failure rates estimation of components. The former can resort to fault mode and effect analysis and the latter needs to obtain lots of fault data, which is almost impossible to estimate precisely the failure rates of the basic events in the practical engineering application. In this paper, interval numbers are used to describe the failure rates of the basic events based on the expert elicitation and some data sheet at the design stage.

2.2. Quantitative analysis of DFT based on DEN

Traditional DFT assumes that the failure rates of the components are expressed in defined values is inadequate to deal with epistemic uncertainty. To this end, the failure rates of the basic events in DFT are considered as interval numbers in this paper and a new DFT solution is proposed to calculate the reliability results by mapping a DFT into a DEN.

In evidence theory, $\Theta = \{W_i, F_i\}$ is the knowledge framework of the component i and the focal elements are defined by:

$$2^{\Theta} = \{\{\varnothing\}, \{W_i\}, \{F_i\}, \{W_i, F_i\}\} \tag{1}$$

where $\{W_i\}$ and $\{F_i\}$ denote the working state and failure state respectively. The state of $\{W_i, F_i\}$ corresponds to the epistemic uncertainty.

Belief measure *(Bel)* defines the lower bound of the probabilities that the focal element exists, and plausibility measure *(Pl)* defines the upper bound of the probabilities that the focal element exists. The basic belief assignment on the system state expresses an epistemic uncertainty, where *Bel* and *Pl* measures are not equal and bound the system reliability. Therefore, the basic probability assignment (BPA) of component i can be computed as:

$$\begin{cases} m(\{W_i\}) = Bel(\{W_i\}) \\ m(\{F_i\}) = 1 - Pl(\{W_i\}) \\ m(\{W_i, F_i\}) = Pl(\{W_i\}) - Bel(\{F_i\}) \end{cases} \tag{2}$$

If a component i follows the exponential distribution with the interval failure rate $[\underline{\lambda}, \overline{\lambda}]$, the interval failure probability of the component at a mission time T can be calculated as follows:

$$[\underline{P_i(x)}, \overline{P_i(x)}] = 1 - exp([\underline{\lambda}, \overline{\lambda}]T) \tag{3}$$

where $\underline{P_i(x)}$ and $\overline{P_i(x)}$ denote respectively the lower failure probability of the component and the corresponding upper failure probability.

Presumably, the upper and lower bounds of the component's failure probability is equivalent to the BPA of component i in the DEN:

$$\begin{cases} m(\{W_i\}) = 1 - \overline{P_i(x)} \\ m(\{F_i\}) = \underline{P_i(x)} \\ m(\{W_i, F_i\}) = \overline{P_i(x)} - \underline{P_i(x)} \end{cases} \tag{4}$$

where $Bel(\{F_i\}) = \underline{P_i(x)}$ and $Pl(\{F_i\}) = \overline{P_i(x)}$.

2.2.1. Mapping a static logic gate into an DEN

Static logic gates mainly include three gates, AND gate, OR gate and voting gate. This section takes an OR gate as an example and provides the schemes to map an OR gate into a DEN. When any of the input components X_i (i=1,…, n) of an OR gate fails, the output of the gate fails too. Fig. 2 shows an OR gate and the equivalent DEN. Table 1 gives the conditional probabilities of node *A (T+ΔT)* in the DEN. Equation (5) gives the conditional probabilities of output node *E (T+ΔT)*. A more detailed description of this work can be found in [16].

Table 1. The conditional probabilities of node A (T+ΔT)

$A(T)$	$A(T+\Delta T)$		
	{W}	{F}	{W,F}
{W}	m_A(W)	m_A(F)	m_A(W,F)
{F}	0	1	0
{W,F}	0	m_A(F)	1− $m_A(F)$

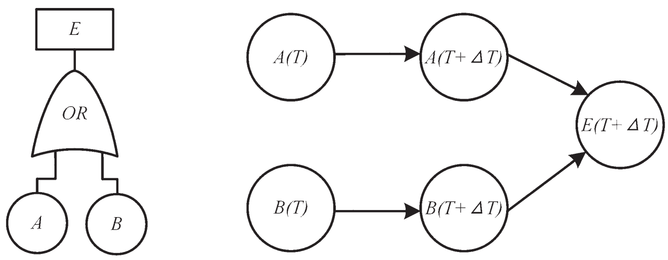


Fig. 2 An OR gate and the equivalent DEN

$$\begin{cases} P(E=1 \mid A(T+\Delta T)=0, B(T+\Delta T)=1)=1 \\ P(E=1 \mid A(T+\Delta T)=1, B(T+\Delta T)=0)=1 \\ P(E=1 \mid A(T+\Delta T)=1, B(T+\Delta T)=1)=1 \\ P(E=1 \mid A(T+\Delta T)=1, B(T+\Delta T)=\{0,1\})=1 \\ P(E=1 \mid A(T+\Delta T)=\{0,1\}, B(T+\Delta T)=1)=1 \\ P(E=\{0,1\} \mid A(T+\Delta T)=0, B(T+\Delta T)=\{0,1\})=1 \\ P(E=\{0,1\} \mid A(T+\Delta T)=\{0,1\}, B(T+\Delta T)=0)=1 \\ P(E=\{0,1\} \mid A(T+\Delta T)=\{0,1\}, B(T+\Delta T)=\{0,1\})=1 \\ P(E=1 \mid A(T+\Delta T)=0, B(T+\Delta T)=0)=0 \end{cases} \tag{5}$$

2.2.2. Mapping a dynamic logic gate into a DEN

Some dynamic logic gates are introduced to model the functional and sequential in the DFT. These logic gates include priority AND gate, the sequence enforcing gate, FDEP and spare gates. An FDEP gate will be used to describe how the dynamic logic gates are mapped into DEN. An FDEP gate includes a trigger event and some dependent basic events. The trigger event can be a basic event or an output of another gate in the DFT. The occurrence of a trigger event will force all basic events to occur, which means all basic events functionally depend upon the trigger event. Fig. 3 shows an FDEP gate and the equivalent DEN. Table 2 and Table 3 show the conditional probabilities of the node $A(T+\Delta T)$ and $E(T+\Delta T)$ respectively.

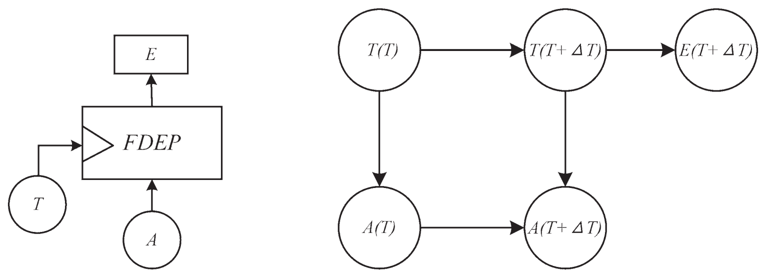


Fig. 3 An FDEP gate and the equivalent DEN.

Table 2. The conditional probabilities of the node A (T+ΔT)

$T(T+\Delta T)$	$A(T)$	$A(T+\Delta T)$		
		{W}	{F}	{W,F}
{W}	{W}	m_A(W)	m_A(F)	m_A(W,F)
{W}	{F}	0	1	0
{W}	{W,F}	0	0	1
{F}	{W}	0	1	0
{F}	{F}	0	1	0
{F}	{W,F}	0	1	0
{W,F}	{W}	0	0	1
{W,F}	{F}	0	1	0
{W,F}	{W,F}	0	0	1

Table 3. The conditional probabilities of the node E(T+ΔT)

T(T+ΔT)	E(T+ΔT)		
	{W}	{F}	{W,F}
{W}	1	0	0
{F}	0	1	0
{W,F}	0	0	1

2.2.3. Calculating reliability results

(1) DIF

DIF is usually defined as the probability that a basic event has occurred given that the top event has also occurred [2]. The DIF of a component i is given by:

$$DIF_i = P(i \mid S) = [Bel(\{F_{i|S}\}), Pl(\{F_{i|S}\})] \tag{6}$$

where i is a component in the system S; $P(i \mid S)$ is the probability that the basic event i has occurred given the top event has occurred.

(2) RAW

RAW, one of the most widely used importance measures, is defined as the ratio of the system unreliability if a component has failed over the system unreliability [17]. Traditionally, the definition of RAW does not take the uncertainties into account. An extension of RAW is introduced which allows us to deal with epistemic uncertainty. The interval RAW of a component i can be defined as follows under uncertainties.

$$I_{X_i}^{RAW} = \frac{P(S=1 \mid X_i=1)}{P(S=1)} = \frac{[Bel(\{F_{S=1|X_i=1}\}), Pl(\{F_{S=1|X_i=1}\})]}{[Bel(\{F_{S=1}\}), Pl(\{F_{S=1}\})]} = \frac{[\underline{Q_{S=1|X_i=1}}, \overline{Q_{S=1|X_i=1}}]}{[\underline{Q_{S=1}}, \overline{Q_{S=1}}]} \tag{7}$$

where $I_{X_i}^{RAW}$ is the RAW for the event X_i, $Bel(\{F_{S=1|X_i=1}\})$ and $Pl(\{F_{S=1|X_i=1}\})$ respectively denote the belief and plausibility measures that the system is in a failed state given that the component i has failed.

3. Dynamic diagnosis algorithm based on heterogeneous information

3.1. Multi-attribute decision-making problem description in the fault diagnosis

If a fault tree has m root nodes, each root node represents a diagnostic scheme. All diagnostic schemes can be expressed in root node set $X = \{X_1, X_2, \cdots, X_m\}$ and each root node has n attributes to evaluate the performance. Evaluation attributes are expressed in attribute set $v = \{v_1, v_2, \cdots, v_n\}$. Different attributes may have different weights and the weights vector is expressed in $\omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$, $\sum_{j=1}^{n} \omega_j = 1,\ 0 < \omega_j < 1$. As for the complexity of decision problem in fault diagnosis and uncertainty, the evaluations for each attribute may be described in different types of values. For example, for precise information, defined value is used; otherwise, due to the epistemic uncertainty, some parameters can be evaluated by some experts. In this situation, the interval number, fuzzy number and linguistic term are more reasonable. In this paper, attribute values are expressed with defined value v^n, interval value v^i and triangle fuzzy number v^f, where $v^n = \{v_1, v_2, \cdots v_{n_1}\}$, $v^i = \{v_{n_1+1}, v_{n_2+2}, \cdots v_{n_2}\}$, $v^f = \{v_{n_2+1}, v_{n_2+2}, \cdots v_n\}$ and $v^n \cup v^i \cup v^f = v$; $N_1 = \{1, 2, \cdots, n_1\}$, $N_2 = \{n_1+1, n_1+2, \cdots n_2\}$, $N_3 = \{n_2+1, n_2+2, \cdots n\}$.

3.2. Distance measure for heterogeneous information [18]

3.2.1. Interval numbers

Definition 1 Let $A = [a^-, a^+]$ and $B = [b^-, b^+]$ be two interval numbers, the distance between A and B is defined as in 1-norm concept:

$$d(A,B) = \| A - B \| = | \underline{a} - \underline{b} | + | \overline{a} - \overline{b} | \tag{8}$$

The larger the distance $d(A,B)$, the greater the degree of separation will be. In particular, when $d(A,B)$ is 0, it means that A and B are equal.

3.2.2. Triangular fuzzy numbers

A triangular fuzzy number is usually given in the form $A=(a,b,c)$, where b is the median value, a is the left distribution of the confidence interval and c is the right distribution of the confidence interval of the fuzzy number A. The membership function of A which associated with a real number in the interval [0, 1] can be defined as:

$$\mu(x) = \begin{cases} (x-a)/(b-a), a \le x \le b \\ (c-x)/(c-b), b \le x \le c \\ 0 \qquad ,others \end{cases} \tag{9}$$

Definition 2 Let $A = (a_1, b_1, c_1)$ and $B = (a_2, b_2, c_2)$ be two triangular fuzzy numbers, the distance between them is defined as in 1-norm concept:

$$d(A,B) = \| A - B \| = | a_1 - a_2 | + | b_1 - b_2 | + | c_1 - c_2 | \tag{10}$$

Similarly, the larger the distance $d(A,B)$, the greater the degree of separation will be. In particular, if $d(A,B)$ is 0, it means that A and B are equal.

3.3. VIKOR algorithm based on generalized distance aggregation function

3.3.1. Generalized distance aggregation function

In the decision making situations where the evaluation values are represented by more than two values types, it is necessary to deal with the heterogeneous information to make full use of this information as much as possible. The base for VIKOR approach is an aggregation function which measures the distance for multi-attributes to compromise ranking. Due to the different types of values for each attribute, a generalized distance aggregation function, $G\text{-}L_P$ [15, 23], is used and is defined as follows:

$$G\text{-}L_{p,i} = \left\{ \sum_{j=1}^{J} [\omega_j \cdot d(f_j^*, f_{ij}) / d(f_j^*, f_j^-)]^p \right\}^{1/p} \tag{11}$$

where $1 \le p \le \infty$; $i = 1,2,\cdots I$. $d(x, y)$ is generalized distance measure function; I is the number of candidate alternatives and J is the number of attributes; $\omega_j$is the weight of j^{th} attribute; For an alternative X_i, its rating on j^{th} attribute is represented as f_{ij}; the positive and negative ideal solution on j^{th} attribute is represented as f_j^* and f_j^- respectively;

$L_{1,i}$ (represented as R_i) is represented as majority rule to satisfy a maximum group utility, while $L_{\infty,i}$ (represented as S_i) is interpreted as a rule to satisfy minimum individual regret [23]. S_i and R_i are used to compromise ranking in group decision and they are calculated by the following equations:

$$S_i = \sum_{j=1}^{J} \omega_j (f_j^* - f_{ij}) \Big/ (f_j^* - f_j^-) \tag{12}$$

$$R_i = \max_j [\omega_j (f_j^* - f_{ij}) \Big/ (f_j^* - f_j^-)] \tag{13}$$

The generalized distance aggregation function is used to eliminate the units of different attribute functions. Because $d(x, y)$ is precise real number belonging to the interval [0,1], VIKOR algorithm with heterogeneous information is similar to the idea of traditional VIKOR.

3.3.2. Determine the best value f_j^* and the worst value f_j^- of all attributes

Dynamic diagnostic strategy is essentially an optimization decision process. For fault diagnosis of systems with heterogeneous information, the first task is to build a decision matrix $F = [f_{ij}]_{m\times n}$. And then the positive ideal solution f_j^* and the negative ideal solution f_j^- of all attributes are calculated as follows according to the different types of values.

If the value type of the attributes is a defined number, the positive ideal solution f_j^* and the negative ideal solution f_j^- can be solved by the following equations:

$$f_j^* = \begin{cases} \max\limits_{1\le i\le m} \{f_{ij}\}, \text{ if the } j_{th} \text{ attribute is a benefit attribute;} \\ \min\limits_{1\le i\le m} \{f_{ij}\}, \text{ if the } j_{th} \text{ attribute is a cost attribute.} \end{cases} \tag{14}$$

$$f_j^- = \begin{cases} \min\limits_{1\le i\le m} \{f_{ij}\}, \text{ if the } j_{th} \text{ attribute is a benefit attribute;} \\ \max\limits_{1\le i\le m} \{f_{ij}\}, \text{ if the } j_{th} \text{ attribute is a cost attribute.} \end{cases} \tag{15}$$

If the value type of the attributes is an interval number, the positive ideal solution f_j^* and the negative ideal solution f_j^- can be calculated by the following equations:

$$f_j^* = \begin{cases} [\max\limits_{1\le i\le m}\{f_{ij}^L\}, \max\limits_{1\le i\le m}\{f_{ij}^U\}], \text{ if the } j_{th} \text{ attribute is a benefit attribute;} \\ [\min\limits_{1\le i\le m}\{f_{ij}^L\}, \min\limits_{1\le i\le m}\{f_{ij}^U\}], \text{ if the } j_{th} \text{ attribute is a cost attribute.} \end{cases} \tag{16}$$

$$f_j^- = \begin{cases} [\min\limits_{1\le i\le m}\{f_{ij}^L\}, \min\limits_{1\le i\le m}\{f_{ij}^U\}], \text{ if the } j_{th} \text{ attribute is a benefit attribute;} \\ [\max\limits_{1\le i\le m}\{f_{ij}^L\}, \max\limits_{1\le i\le m}\{f_{ij}^U\}], \text{ if the } j_{th} \text{ attribute is a cost attribute.} \end{cases} \tag{17}$$

where $[f_{ij}^L, f_{ij}^U]$ is **an interval evaluation value of the i^{th} alternative on the j^{th} attribute.**

If the value type of the attributes is a triangular fuzzy number, the positive ideal solution f_j^* and the negative ideal solution f_j^- can be calculated by the following equations:

$$f_j^* = \begin{cases} (\max\limits_{1\le i\le m}\{f_{ij}^L\}, \max\limits_{1\le i\le m}\{f_{ij}^M\}, \max\limits_{1\le i\le m}\{f_{ij}^U\}), \text{ if the } j_{th} \text{ attribute is a benefit attribute;} \\ (\min\limits_{1\le i\le m}\{f_{ij}^L\}, \min\limits_{1\le i\le m}\{f_{ij}^M\}, \min\limits_{1\le i\le m}\{f_{ij}^U\}), \text{ if the } j_{th} \text{ attribute is a cost attribute.} \end{cases} \tag{18}$$

$$f_j^- = \begin{cases} (\min\limits_{1\le i\le m}\{f_{ij}^L\}, \min\limits_{1\le i\le m}\{f_{ij}^M\}, \min\limits_{1\le i\le m}\{f_{ij}^U\}), \text{ if the } j_{th} \text{ attribute is a benefit attribute;} \\ (\max\limits_{1\le i\le m}\{f_{ij}^L\}, \max\limits_{1\le i\le m}\{f_{ij}^M\}, \max\limits_{1\le i\le m}\{f_{ij}^U\}), \text{ if the } j_{th} \text{ attribute is a cost attribute.} \end{cases} \tag{19}$$

where $(f_{ij}^L, f_{ij}^M, f_{ij}^U)$ is a triangular fuzzy value of the i^{th} alternative on the j^{th} attribute given by domain experts; $\|f_j^* \text{-} f_j^-\| \ne 0$.

3.3.3. Normalize the decision matrix

Usually, evaluation values of different attributes have different dimensions, which are not directly comparable. So the decision matrix $F = [f_{ij}]_{m\times n}$ for heterogeneous information with different dimensions should be normalized. A normalized decision matrix $P^C = [p_{ij}]_{m\times n}$ is obtained based on the 1-norm concept using the following equation:

$$p_{ij} = \frac{\| f_j^* - f_{ij} \|}{\| f_j^* - f_j^- \|} = \begin{cases} \dfrac{| f_j^* - f_{ij} |}{| f_j^* - f_j^- |}, i \in M, j \in N_1 \\ \dfrac{| f_j^{L*} - f_{ij}^L | + | f_j^{U*} - f_{ij}^U |}{| f_j^{L*} - f_j^{L-} | + | f_j^{U*} - f_j^{U-} |}, i \in M, j \in N_2 \\ \dfrac{| f_j^{L*} - f_{ij}^L | + | f_j^{M*} - f_{ij}^M | + | f_j^{U*} - f_{ij}^U |}{| f_j^{L*} - f_j^{L-} | + | f_j^{M*} - f_j^{M-} | + | f_j^{U*} - f_j^{U-} |}, i \in M, j \in N_3 \end{cases} \tag{20}$$

3.3.4. Calculate the weights of attributes based on the Entropy concept

There are several attributes in the multi-attribute decision making, and their weights may be unknown. Subjective evaluation method and objective evaluation method can be used to determine the weights of attributes. However, the former usually uses the subjective judgment of the decision maker to determine the weights of attributes and it has subjectivity and arbitrariness to a certain degree. Objective evaluation method uses some algorithms to calculate the weights of attributes according to the attributes information and it is more scientific. Entropy weight method [14] is widely used to determine the weights in practical engineering. Shannon Entropy is a measure of information uncertainty based on probability theory. It is very suitable for measuring the relative contrast intensities of attributes to represent the average intrinsic information transmitted to the decision makers. The smaller the entropy value of evaluation attribute v_j is, the more the value of this attribute plays in the decision. That is to say, its weight is larger. The steps for determining the weights of attributes based on the entropy weight method are as follows:

Step 1: Normalize the normalized decision matrix and calculate the weighted proportion of the i^{th} alternative on the j^{th} attribute using the following equation.

$$P_{ij} = \frac{p_{ij}}{\sum_{i=1}^{n} p_{ij}} \tag{21}$$

Step 2: Calculate the entropy H_j value of the j^{th} attribute as follows:

$$H_j = -K \sum_{i=1}^{n} p_{ij} \ln p_{ij} \tag{22}$$

where $K = 1/\ln n$ $(K > 0, 0 \le p_{ij} \le 1)$ and assume $p_{ij} \ln p_{ij} = 0$ if p_{ij} is 0.

Step 3: Calculate the value of α_j defined as follows:

$$\alpha_j = 1 - H_j \tag{23}$$

where α_j is the divergence degree of the intrinsic information of the j^{th} attribute. The greater the value of α_j, the more important the attribute is in the decision making process.

Step 4: Calculate the weights of attributes using the following equation:

$$\omega_j = \frac{\alpha_j}{\sum_{j=1}^{m} \alpha_j} \tag{24}$$

where $\sum_{j=1}^{n} \omega_j = 1, 0 \le \omega_j \le 1$.

3.3.5. Calculate the values S_i, R_i and Q_i

A generalized distance aggregation function is used to obtain the optimal ranking in the decision making according to VIKOR algorithm. The optimal ranking should satisfy the maximum group utility and satisfy the minimum individual regret. S_i, R_i and Q_i are defined as follows.

$$S_i = \sum_{j=1}^{n} \omega_j \cdot p_{ij} \tag{25}$$

$$R_i = \max_{1 \le j \le n} \{\omega_j \cdot p_{ij}\} \tag{26}$$

$$Q_i = v \frac{S_i - S^+}{S^- - S^+} + (1-v) \frac{R_i - R^+}{R^- - R^+} \tag{27}$$

where $S^+ = \min_i S_i$, $S^- = \max_i S_i$, $R^+ = \min_i R_i$ $R^- = \max_i R_i$ and v is introduced as the weight for the strategy of maximum group utility, whereas 1-v is the weight of the individual regret. If v>0.5, it means that a decision making is based on the conditions agreed by the vast majority of policy makers. If v<0.5, the decision making is based on the circumstances refused by the vast majority of policy makers. Usually, v can take any value from 0 to 1 and the value of v is set to 0.5 in the paper. Finally, we can obtain the optimal diagnosis ranking by the value Q_i in ascending order.

3.3.6. Updating the decision matrix using the previous diagnosis result

The component with a smaller Q_i value should be diagnosed first. This assures a reduced number of system checks while bringing the system back to life. Nevertheless, this approach fails to update the reliability parameters in order to optimize the diagnosis process using the previous diagnosis result. That is to say, DIF and RAW are not updated by the previous diagnosis result, thereby having a significant effect on the diagnosis efficiency. When the component diagnosed at the present time works we should feed this evidence information to a DEN and obtain the updating DIF and RAW. In addition, the decision matrix should be updated too and the corresponding value of Q_i can be calculated to determine the next optimal ranking. And so on, the final optimal diagnostic ranking can be obtained.

4. A case study

Train-ground wireless communication system, a vital subsystem of urban rail transit, is responsible for data transmission between vehicle equipment and ground equipment. To ensure safe operation, application of high technologies has been used to improve its reliability greatly. Once train-ground wireless communication system breaks down, it may decrease the operation performance and even causes a great loss. Therefore, an efficient diagnosis strategy should be taken to bring it back to life as soon as possible when it fails. Fig.4 shows the DFT model of a train-ground wireless communication system. It is assumed that all components have the exponential distribution and failure rates of components expressed in interval values are shown in Table 4.

Table 4. Interval failure rates of components

Components	Interval failure rates	Components	Interval failure rates
X1	[4.22e-6, 5.28e-6]	X8,X9	[5.49e-6, 6.71e-6]
X2	[5.94e-6, 7.26e-6]	X10,X11	[3.15e-5, 3.85e-5]
X3	[4.86e-5, 5.94e-5]	X12,X13	[6.12e-5, 7.48e-5]
X4,X5	[3.78e-5, 4.62e-5]	X14	[5.04e-5, 6.11e-5]
X6,X7	[6.48e-5, 7.92e-5]	X15	[5.04e-5, 6.11e-5]

Table 5. DIF of all components

Components	DIF of components	Components	DIF of components
X1	[0.0508,0.0518]	X8,X9	[0.0857,0.0939]
X2	[0.0709,0.0722]	X10,X11	[0.0708,0.0756]
X3	[0.5681,0.5727]	X12,X13	[0.2012,0.2156]
X4,X5	[0.0751,0.0822]	X14	[0.1788,0.1914]
X6,X7	[0.1963,0.2148]	X15	[0.1788,0.1914]

The DFT is mapped into a corresponding DEN for quantitative analysis using the method mentioned above. Assuming the task time T=1000 h, the probability of system failure can be obtained using the inference algorithm, and it is [0.08293, 0.10714]. In addition, the DIF and RAW of all components can be calculated shown in Table 5 and

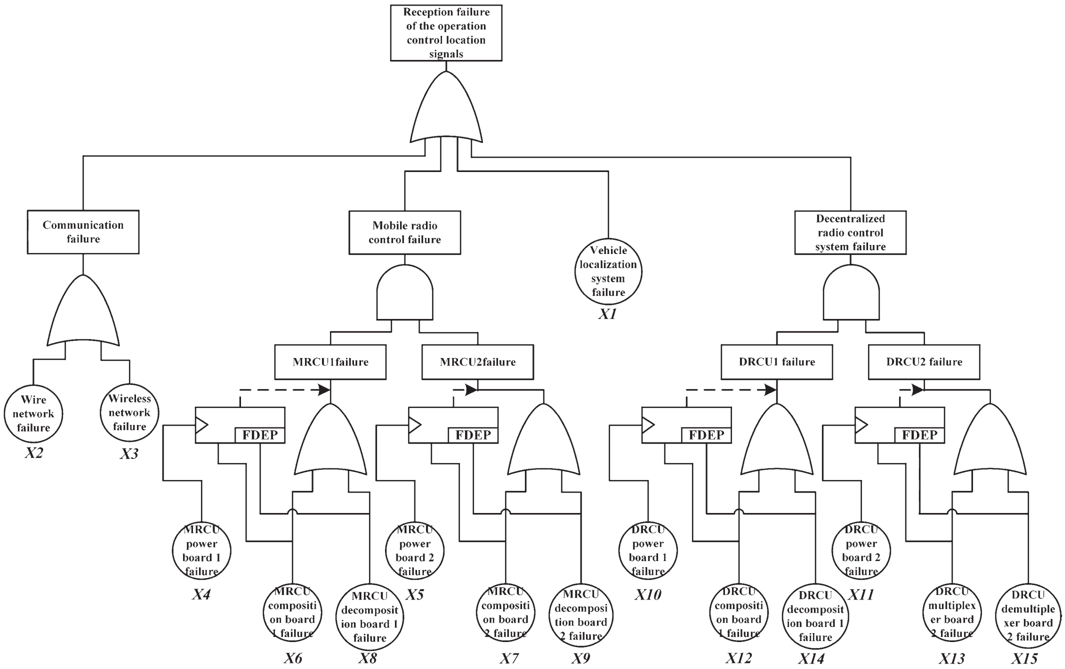


Fig. 4. DFT model of train-ground wireless communication system

Table 6. RAW of all components

Components	$P(S=1 \mid X_i=1)$	$I_{X_i}^{RAW}$
X1	1	[9.3337,12.0694]
X2	[0.99214,1]	[9.2603,12.0694]
X3	[0.99214,1]	[9.2603,12.0694]
X4	[0.16768, 0.205466]	[1.5651,2.4799]
X5	[0.16768, 0.205466]	[1.5651,2.4799]
X6	[0.16768, 0.205466]	[1.5651,2.4799]
X7	[0.16768, 0.205466]	[1.5651,2.4799]
X8	[0.16768, 0.205466]	[1.5651,2.4799]
X9	[0.16768, 0.205466]	[1.5651,2.4799]
X10	[0.18919, 0.230323]	[1.7658,2.7799]
X11	[0.18919, 0.230323]	[1.7658,2.7799]
X12	[0.18919, 0.230323]	[1.7658,2.7799]
X13	[0.18919, 0.230323]	[1.7658,2.7799]
X14	[0.18919, 0.230323]	[1.7658,2.7799]
X15	[0.18919, 0.230323]	[1.7658,2.7799]

Table 7. Linguistic assessment of Components' test cost

Components	test cost
X1	High
X2	Moderate
X3	Very High
X4,X5	Very Low
X6,X7	Low
X8,X9	Low
X10,X11	Very Low
X12,X13	Low
X14,X15	Low

Table 8. Evaluation standards of the test cost

Linguistic expression for test cost	Fuzzy numbers
Very High	(0.7, 0.9, 1)
High	(0.5, 0.7, 0.9)
Moderate	(0.3, 0.5, 0.7)
Low	(0.1, 0.3, 0.5)
Very Low	(0.1, 0.2, 0.3)

Table 9. A decision matrix with heterogeneous information

Components	DIF	RAW	Test cost
X1	[0.0508,0.0518]	[9.3337,12.0694]	(0.5,0.7,0.9)
X2	[0.0709,0.0722]	[9.2603,12.0694]	(0.3,0.5,0.7)
X3	[0.5681,0.5727]	[9.2603,12.0694]	(0.7,0.9,1)
X4	[0.0751,0.0822]	[1.5651,2.4799]	(0.1,0.2,0.3)
X5	[0.0751,0.0822]	[1.5651,2.4799]	(0.1,0.2,0.3)
X6	[0.1963,0.2148]	[1.5651,2.4799]	(0.1,0.3,0.5)
X7	[0.1963,0.2148]	[1.5651,2.4799]	(0.1,0.3,0.5)
X8	[0.0857,0.0939]	[1.5651,2.4799]	(0.1,0.3,0.5)
X9	[0.0857,0.0939]	[1.5651,2.4799]	(0.1,0.3,0.5)
X10	[0.0708,0.0756]	[1.7658,2.7799]	(0.1,0.2,0.3)
X11	[0.0708,0.0756]	[1.7658,2.7799]	(0.1,0.2,0.3)
X12	[0.2012,0.2156]	[1.7658,2.7799]	(0.1,0.3,0.5)
X13	[0.2012,0.2156]	[1.7658,2.7799]	(0.1,0.3,0.5)
X14	[0.1788,0.1914]	[1.7658,2.7799]	(0.1,0.3,0.5)
X15	[0.1788,0.1914]	[1.7658,2.7799]	(0.1,0.3,0.5)

Table 10. A normalized decision matrix with heterogeneous information

components	DIF	RAW	Test cost
X1	1.0000	0.0000	0.2500
X2	0.9610	0.0042	0.5500
X3	0.0000	0.0042	0.0000
X4	0.9473	1.0000	1.0000
X5	0.9473	1.0000	1.0000
X6	0.7029	1.0000	0.8500
X7	0.7029	1.0000	0.8500
X8	0.9258	1.0000	0.8500
X9	0.9258	1.0000	0.8500
X10	0.9578	0.9712	1.0000
X11	0.9578	0.9712	1.0000
X12	0.6974	0.9712	0.8500
X13	0.6974	0.9712	0.8500
X14	0.7422	0.9712	0.8500
X15	0.7422	0.9712	0.8500

Table 11. The positive and negative ideal solutions

Attributes	Positive ideal solutions	Negative ideal solutions
DIF	[0.5681，0.5727]	[0.0508，0.0518]
RAW	[9.3337，12.0694]	[1.5651，2.4799]
Test cost	(0.1，0.2，0.3)	(0.7，0.9，1)

Table 6 respectively. DIF enables us to discriminate between components by their importance from a diagnostic point of view. RAW is defined as the ratio of the system unreliability if a component has failed over the system unreliability and it plays an important role in the diagnostic sequence. Furthermore, test cost of the components has a significant impact on diagnostic strategy. However, test cost of all components is usually very difficult to express as defined values because of uncertainties. So the linguistic assessments are used for generating criteria and alternative ratings, which are transformed into triangular fuzzy numbers to describe test cost of all components. Table 7 and Table 8 show the linguistic assessment of the test cost and alternative ratings of all components.

DIF, RAW and test cost are used to build a decision matrix. The former two, expressed in interval numbers, belong to the benefit attributes. The latter belongs to the cost attribute, which is expressed in a triangular fuzzy number. Table 9 and table 10 show the deci-

Table 12. Values of S_i, R_i and Q_i for all components

Components	S_i	R_i	Q_i
X1	0.5782	0.2905	0.4156
X2	0.4532	0.2792	0.1275
X3	0.3850	0.3836	0.5000
X4	0.6010	0.3259	0.6275
X5	0.6010	0.3259	0.6275
X6	0.5876	0.3259	0.6023
X7	0.5876	0.3259	0.6023
X8	0.6524	0.3259	0.7234
X9	0.6524	0.3259	0.7234
X10	0.5947	0.3165	0.5706
X11	0.5947	0.3165	0.5706
X12	0.5766	0.3165	0.5367
X13	0.5766	0.3165	0.5367
X14	0.5896	0.3165	0.5611
X15	0.5896	0.3165	0.5611

Table 13. An updating decision matrix with the previous diagnosis result

Components	DIF	RAW	Test cost
X1	[0.0544,0.0554]	[9.9374,12.9137]	(0.5,0.7,0.9)
X3	[0.6078,0.6120]	[9.8593,12.9137]	(0.7,0.9,1.0)
X4	[0.0778,0.0850]	[1.6174,2.5785]	(0.1,0.2,0.3)
X5	[0.0778,0.0850]	[1.6174,2.5785]	(0.1,0.2,0.3)
X6	[0.2039,0.2222]	[1.6174,2.5785]	(0.1,0.3,0.5)
X7	[0.2039,0.2222]	[1.6174,2.5785]	(0.1,0.3,0.5)
X8	[0.0890,0.0971]	[1.6174,2.5785]	(0.1,0.3,0.5)
X9	[0.0890,0.0971]	[1.6174,2.5785]	(0.1,0.3,0.5)
X10	[0.0739,0.0790]	[1.8336,2.9019]	(0.1,0.2,0.3)
X11	[0.0739,0.0790]	[1.8336,2.9019]	(0.1,0.2,0.3)
X12	[0.2100,0.2242]	[1.8336,2.9019]	(0.1,0.3,0.5)
X13	[0.2100,0.2242]	[1.8336,2.9019]	(0.1,0.3,0.5)
X14	[0.1866,0.1990]	[1.8336,2.9019]	(0.1,0.3,0.5)
X15	[0.1866,0.1990]	[1.8336,2.9019]	(0.1,0.3,0.5)

Table 14. Revised values of S_i, R_i and Q_i for all components

Components	S_i	R_i	Q_i
X1	0.5369	0.2746	0.2790
X3	0.3512	0.3496	0.3709
X4	0.6372	0.3757	0.9298
X5	0.6372	0.3757	0.9298
X6	0.6245	0.3757	0.9107
X7	0.6245	0.3757	0.9107
X8	0.6839	0.3757	1
X9	0.6839	0.3757	1
X10	0.6289	0.3649	0.8636
X11	0.6289	0.3649	0.8636
X12	0.6117	0.3649	0.8377
X13	0.6117	0.3649	0.8377
X14	0.6237	0.3649	0.8557
X15	0.6237	0.3649	0.8557

sion matrix with heterogeneous information and normalized decision matrix respectively. The positive and negative ideal solutions can be obtained shown in table 11. Based on the entropy methodology, the weights of the three attributes, ω_1=0.2905，ω_2=0.3259，ω_3=0.3836 are obtained using the Eq. (21) - (24). Table 12 presents the values of S_i, R_i and Q_i for all components. The optimal diagnosis sequence is as follows according to the corresponding Q_i in ascending order.

X2≻X1≻X3≻X12(X13)≻X10(X11)≻X14(X15)≻X6(X7)≻X4(X5)≻X8(X9)

If a train-ground wireless communication system broke down, we should diagnose X2 firstly. If X2 fails, then diagnosis is over. Otherwise, we should feed this evidence information (X2 works) to the DEN and recalculate DIF and RAW. An updating decision matrix with the previous diagnosis result is shown in table 13. Similarly, the revised values of S_i, R_i and Q_i for all components can be obtained shown in table 14. The updating optimal diagnosis sequence is as follows.

X1≻X3≻X12(X13)≻X14(X15)≻X10(X11)≻X6(X7)≻X4(X5)≻X8(X9)

So we can draw a conclusion that the next component diagnosed is X1. If X1 fails, then diagnosis is over. Otherwise, we input this evidence information to the DEN and update the decision matrix again. These steps are repeated several times, and the final optimal diagnostic ranking can be obtained as follows.

X2≻X1≻X3≻X12(X13)≻X10(X11)≻X4(X5)≻X14(X15)≻X6(X7)≻X8(X9)

Obviously, the diagnostic strategy which takes the previous diagnosis result into account is more reasonable and efficient because it can update the decision matrix dynamically. To avoid subjectivity and arbitrariness, the proposed method determines the weights of attributes based on the Entropy concept. Besides, the optimal ranking is obtained directly based on the original heterogeneous information without a transformation process using a generalized distance-based function, which can improve diagnosis efficiency and reduce information loss.

5. Conclusion

In this paper, a novel dynamic diagnostic strategy for complex systems is proposed based on reliability analysis and distance-based VIKOR with heterogeneous information, which aims to deal with two important issues that arise in engineering applications, such as failure dependency and epistemic uncertainty. For the challenge of the failure dependency, a DFT is used to describe the dynamic fault behaviours. For the challenge of the epistemic uncertainty, the failure rates of components in complex systems are expressed in interval numbers. Furthermore, DFT is converted into a DEN to calculate some reliability results and these parameters together with test cost constitute a decision matrix. In addition, a dynamic diagnostic strategy is developed based on an improved VIKOR algorithm and the previous diagnosis result. This diagnosis algorithm determines the weights of attributes based on the Entropy concept to avoid experts' subjectivity and obtains the optimal ranking directly on the original heterogeneous information without a transformation process, which can improve diagnosis efficiency and reduce information loss. Finally, a train-ground wireless communication system is given to demonstrate the efficiency of the proposed method. This method takes full advantages of DFT for modelling, DEN for the uncertainty inference and VIKOR for dynamic decision making, which is especially suitable to diagnose complex systems.

Acknowledgement

This research was funded by National Natural Science Foundation of China [71461021].

References

1. Assaf T, Dugan J B. Design for diagnosis using a diagnostic evaluation measure. Instrumentation & Measurement Magazine 2006; 9(4): 37-43, https://doi.org/10.1109/MIM.2006.1664040.
2. Assaf T, Dugan J B. Diagnosis based on reliability analysis using monitors and sensors, Reliability Engineering & System Safety 2008; 93(4): 509-521, https://doi.org/10.1016/j.ress.2006.10.024.
3. Cai B, Liu H, Xie M. A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks. Mechanical Systems and Signal Processing 2016; 80: 31–44, https://doi.org/10.1016/j.ymssp.2016.04.019.
4. Chiremsel Z, Said R N, Chiremsel R. Probabilistic Fault Diagnosis of Safety Instrumented Systems based on Fault Tree Analysis and Bayesian Network. Journal of Failure Analysis & Prevention 2017; 16(5): 747-760, https://doi.org/10.1007/s11668-016-0140-z.
5. Cao Y, Li T, Wang J, Xie R, Wang X. Fuzzy dynamic fault tree analysis for electro-mechanical actuator based on algebraic model with common-cause failures. Automatic Control and Computer Sciences 2016; 50(2): 80-90, https://doi.org/10.3103/S0146411616020024.
6. Doguc O. and Ramirez-Marquez J. E. Using Bayesian approach for sensitivity analysis and fault diagnosis in complex systems. Journal of Integrated Design and Process Science 2009; 13(1):33–48.
7. Duan R, Zhou H, Fan J. Diagnosis strategy for complex systems based on reliability analysis and MADM under epistemic uncertainty. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2015; 17(3): 343–352, https://doi.org/10.17531/ein.2015.3.4.
8. Duan R X, Zhou H L. Diagnosis strategy for micro-computer controlled straight electro-pneumatic braking system using fuzzy set and dynamic fault tree. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2014; 16(2): 217–223.
9. Huang Y F, Jing B. Diagnosis strategy for multi-value attribute system based on Rollout algorithm. Control and Decision 2011; 26(8): 1269-1272.
10. Johnson R A. An information theory approach to diagnosis. IRE Transactions on Reliability and Quality Control 1960; 9(1): 35-35, https://doi.org/10.1109/IRE-PGRQC.1960.5007263.
11. Kabir S, Walker M, Papadopoulos Y, Rüde E, Securius P. Fuzzy temporal fault tree analysis of dynamic systems. International Journal of Approximate Reasoning 2016; 77: 20-37, https://doi.org/10.1016/j.ijar.2016.05.006.
12. Liu Z, Li Y F, He L P, Yang Y Z, Huang H Z. A new fault tree analysis approach based on imprecise reliability model. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 2014; 228(4): 371-381, https://doi.org/10.1177/1748006X14520824.
13. Lourenzutti R. Krohling R A. A generalized TOPSIS method for group decision making with heterogeneous information in a dynamic environment. Information Sciences 2016; 33:01-18, https://doi.org/10.1016/j.ins.2015.10.005.

14. Liu H C, You J X, You X Y. A novel approach for failure mode and effects analysis using combination weighting and fuzzy VIKOR method. Applied Soft Computing 2015; 28:579-588, https://doi.org/10.1016/j.asoc.2014.11.036.
15. Opricovic S. Tzeng G. Extended VIKOR method in comparison with outranking methods. European Journal of Operational Research 2007; 178(2):514-529, https://doi.org/10.1016/j.ejor.2006.01.020.
16. Philippe W, Christophe S. Dynamic evidential networks in system reliability analysis: A Dempster-Shafer Approach. Proceedings of the 16th Mediterranean Conference on Control & Automation, Ajaccio, France 2008; 603-608.
17. Sallak M, Schon W. Aguirre F. Extended Component Importance Measures Considering Aleatory and Epistemic Uncertainties. IEEE Transactions on Reliability 2013; 62(1):49-65, https://doi.org/10.1109/TR.2013.2240888.
18. Sanayei A, Mousavi S F. Yazdankhah A. Group decision making process for supplier selection with VIKOR under fuzzy environment. Expert Systems with Applications 2010; 37(1):24-30, https://doi.org/10.1016/j.eswa.2009.04.063.
19. Şahin R, Liu P. Possibility-induced simplified neutrosophic aggregation operators and their application to multi-criteria group decision-making. Journal of Experimental & Theoretical Artificial Intelligence 2017; 29(4): 769-785, https://doi.org/10.1080/0952813X.2016.1259266.
20. Selim H, Yunusoglu M G, Yılmaz S. Balaman, A dynamic maintenance planning framework based on fuzzy TOPSIS and FMEA: application in an international food company. Quality and Reliability Engineering International 2016; 32(3): 795-804, https://doi.org/10.1002/qre.1791.
21. Tian H, Duan F, Fan L. Fault diagnostic strategy of multivalued attribute system based on growing algorithm. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 2018; https://doi.org/10.1177/1748006X18770356.
22. Wang D, Zhang Y, Jia X, Jiang P, Guo B. Handling Uncertainties in Fault Tree Analysis by a Hybrid Probabilistic–Possibilistic Framework. Quality and Reliability Engineering International 2016; 32(3): 1137-1148, https://doi.org/10.1002/qre.1821.
23. Wang X, Cai J. A group decision-making model based on distance-based VIKOR with incomplete heterogeneous information and its application to emergency supplier selection. Kybernetes 2017; 46(3): 501-529, https://doi.org/10.1108/K-06-2016-0132.
24. Zhang R, Zhang L, Wang N, Wang W. Reliability evaluation of a multi-state system based on interval-valued triangular fuzzy Bayesian networks. International Journal of System Assurance Engineering and Management 2016; 7(1):16-24, https://doi.org/10.1007/s13198-015-0335-9.
25. Zhang X, Mahadevan S, Deng X. Reliability analysis with linguistic data: An evidential network approach. Reliability Engineering & System Safety 2017; 162: 111-121, https://doi.org/10.1016/j.ress.2017.01.009.

Jing LI
School of Economics and Management
Nanchang University
999 Xuefu Rd., Honggutan new district
Nanchang, Jiangxi, China

Rongxing DUAN
School of Information Engineering
Nanchang University
999 Xuefu Rd., Honggutan new district
Nanchang, Jiangxi, China

E-mail: lijingncu@126.com, duanrongxing@ncu.edu.cn

Article citation info:
KALUĐER S, FEKETE K, JOZSA L, KLAIĆ Z. Fault diagnosis and identification in the distribution network using the fuzzy expert system. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 621–629, http://dx.doi.org/10.17531/ein.2018.4.13.

Slaven KALUĐER
Krešimir FEKETE
Lajos JOZSA
Zvonimir KLAIĆ

FAULT DIAGNOSIS AND IDENTIFICATION IN THE DISTRIBUTION NETWORK USING THE FUZZY EXPERT SYSTEM

IDENTYFIKACJA I DIAGNOZA BŁĘDÓW W ELEKTROENERGETYCZNEJ SIECI ROZDZIELCZEJ Z WYKORZYSTANIEM ROZMYTEGO SYSTEMU EKSPERCKIEGO

In this paper, a fuzzy expert off-line system has been developed for fault diagnosis in the distribution network based on the structural and functional operation of the relay and circuit breakers. Functional operations (correct operation, false operation and failure to operate) of the relays and circuit breakers are described by fuzzy logic. Input data for the proposed fuzzy expert fault diagnosis system (FDS) are status and time stamps of the alarms, associated with relays and circuit breakers. The diagnostic system from a huge number of alarms sets, logically organizes and quantifies the diagnosis. FDS can diagnose correct operation, false operation and failure to operate of the relays and circuit breakers. Also, it can identify and quantify fault location based on the Hamacher's operator of a fuzzy union. The additional contribution of this paper is in modeling unknown information using linear fuzzy membership function. Statuses of certain components may be unknown due to telemetry failures or are simply unavailable to the operator and proposed FDS can make diagnosis in such a situation. Developed fuzzy expert FDS is tested on the two examples of faults in real life distribution network.

Keywords: *fault diagnosis, alarm processing, fuzzy logic, expert system, distribution network.*

W prezentowanym artykule opracowano rozmyty system ekspercki typu off-line do diagnozowania błędów w elektroenergetycznej sieci rozdzielczej. System bazuje na strukturze i działaniu przekaźnika i wyłączników automatycznych. Działanie (prawidłowe działanie, błędne działanie i brak działania) przekaźników i wyłączników opisano za pomocą logiki rozmytej. Dane wejściowe do proponowanego rozmytego eksperckiego systemu diagnostyki błędów (FDS) stanowią stany i sygnatury czasowe alarmów, związane z przekaźnikami i wyłącznikami. System diagnostyczny logicznie porządkuje i określa ilościowo diagnozę na podstawie ogromnej liczby zestawów alarmów. FDS pozwala zdiagnozować prawidłowe działanie, błędne działanie oraz awarię (brak działania) przekaźników i wyłączników. Ponadto umożliwia identyfikację i lokalizację błędów w oparciu o sumę Hamachera. W artykule dodatkowo omówiono metodę modelowania informacji nieznanych przy użyciu liniowej funkcji przynależności dla zbiorów rozmytych. Stany niektórych elementów mogą być nieznane z powodu awarii telemetrii lub mogą być po prostu niedostępne dla operatora. Proponowany FDS umożliwia postawienie diagnozy w takich sytuacjach. Opracowany rozmyty ekspercki FDS testowano na dwóch przykładach błędów powstałych w funkcjonującej sieci rozdzielczej.

Słowa kluczowe: *diagnoza błędów, przetwarzanie alarmów, logika rozmyta, system ekspercki, sieć rozdzielcza.*

1. Introduction

Nowadays, electric power systems around the world are becoming more and more complex, and their operation is often close to their limits. Uninterrupted supply of electricity is an important factor influencing a series of events and relationships within the society itself. Disruption of the power systems caused by faults results in significant financial damage due to unsupplied electricity (possible consumers' lawsuits for damage caused by interruption of electricity supply). One of the ways of dealing with the mentioned problems is automating and systematically handling of the information which help power system operators make the right decisions. The nature of this type of problem is diagnostic, and the general term is a failure diagnosis. According to Sekine [26], ″fault diagnosis of power systems involves identifying the location and cause of faults occurring in the power system due to lightning strokes, and so on″. With occurrence of a few hundred alarms in a short period of time, the situation for a power system operator is complicated and it is very difficult to find a section of a failure and a cause of a failure. This prevents the operator from reacting in a proper way and to establish initial topology of the network thus making operator's assumptions erroneous. Due to such situations, various methods for analysing the alarms are used. In order to assist operators in making decisions and diagnosis, automated fault-diagnosis and decision-making systems are being developed. Off-line systems for the diagnosis of faults are primarily developed. On-line systems for fault diagnosis, often called alarm processing, is developed from the off-line systems [29]. Interesting examples of fault diagnosis in complex systems can be found in the literature [6, 8, 10]. Modern online diagnostic systems work in the form of advanced DMS (Distribution System Management) applications. The current situation in the Croatian distribution system is such that most diagnostics are carried out by the operator (human) on the basis of received alarms that are generated in the case of the fault. In many parts of the distribution network, equipment is not connected to the local dispatching centre and operators are often faced with a lack of information. The first step towards the establishment of a modern on-line diagnostic system is a communication connection of all relevant equipment with a control center. This is the way towards the development of the so-called smart grid where it will rely less and less on the network operator response and failures will be addressed through a network self-recovery.

Artificial Intelligence (AI) has proven to be successful in solving diagnosis problems when compared to traditional numerical methods [32]. Expert systems (ES) are one of the AI methods which are commonly used to solve these problems. They use and import methods developed in the area of qualitative reasoning [13]. ES possesses a lack of generality in an application for failure diagnosis and restoration, which is not considered as a negative characteristic in such an extensive and combinatorial problem, since the system requires: understanding, flexibility and high performance [18]. ES emulates the solution, which means that the solution can be closer to the reality than the solution obtained by simulation [11]. However, some problems remain, such as shortcomings or complete lack of information, quantifying fault identification and black-and-white thinking. For such a type of problem, fuzzy rule-based expert systems are suitable.

In [12] fuzzy set is developed to deal with the uncertainty involved in the process of locating faults in distribution networks. The fuzzy set theory has been introduced as a mechanism to incorporate uncertainties and qualitative judgments in the status of relays and circuit breakers as well as correctness or incorrectness of their operations in the research papers [4, 20]. An integrated fuzzy expert system is presented in [15] to diagnose faults in a transmission network and substations. Besides the application in power system, fuzzy expert systems are widely used for diagnosing problems in many fields [1]. Methods by which researchers are trying to overcome the shortcomings of fuzzy expert systems are: artificial neural networks (ANNs) combined with the fuzzy logic systems [3, 24], genetic algorithms (GA) [2, 14], fuzzy Petri net [34], mixed integer programming model [21], multi-agent systems [25] etc. A detailed review of articles on intelligent systems used for fault diagnosis in transmission networks can be found in the literature [7]. On the other hand, trends in the fault diagnosis in distribution systems include systems with distributed generation (photovoltaics, wind generation etc.) [17, 27, 28, 30, 33, 35, 36].

Based on the literature review, it can be concluded that most commonly used methods for fault diagnosis in distribution networks include expert systems, neural networks (NN), fuzzy logic (FL) and genetic algorithms (GA). The strengths of the ES are in the representation of expert knowledge and the interpretation of causative-consequence relationship. The shortcomings of the ES are in the lack of generality, inability to learn and adapt. On the other hand, NN and GA are superior in the case of learning and adapting and dealing with uncertainty and missed data. The disadvantages of NN and GA lie in poor expert knowledge representation and interpretation. The FL is very good in dealing with uncertainty and missed data while it is weaker regarding learning and adapting. Research trends show that solutions related to fault diagnosis in the distribution network are found in the hybridization of different methods in order to combine the strengths of each method and overcome weaknesses. A combination of ES and FL has been shown to be effective and it is chosen in this paper.

It is noticed that fuzzy expert systems developed in [4, 12, 15, 20] for fault diagnosis use information only in the form of relays and circuit breakers (CBs) status and they do not use the alarm time stamp. Also, the expert base of knowledge are not shown by the functional activity of the relay and CB. In this paper, a model for functional activity of the relay and the circuit breaker, described by fuzzy logic, is presented and incorporated in off-line fault diagnosis system (FDS) for application to electric distribution system. Fuzzy expert FDS uses relay and CB statuses and their timestamps as input data. The diagnostic system sets, logically organizes and quantifies the diagnosis from a huge number of alarms. Developed fuzzy expert FDS successfully diagnoses the correct operation, false operation and failure to operate of the relay and CB. It also has the ability to locate and quantify the fault on the basis of the operator of the fuzzy union, which is expressed as Hamacher's union operator [37]. Locating and quantifying the fault refers to finding a section of distribution network which is faulted and quantifying this diagnose with a certain probability. In addition, it also successfully deals with telemetry breakdowns by using a new way of modeling non-existent information using a linear time-dependent fuzzy membership function. Since faults in telemetry, noise or lack of connection can make certain statuses of relays or CBs unknown, modeling of non-existent information enables successful diagnosis even in these cases. Developed fuzzy expert FDS is tested on the two examples of faults in real life distribution network.

The structure of this paper is as follows: first, a detailed mathematical description of the developed fuzzy expert FDS is done in Section 2. Modeling of the real-life distribution network is briefly described in Section 3. In Section 4, examples of the diagnoses for the real-life events are presented. Short conclusion and comments are made at the end of the paper.

2. Mathematical description of the developed fuzzy expert FDS

2.1. General description of the model

The flow chart diagram of a developed fuzzy expert FDS is shown in Fig. 1. The diagram starts with the input data relating to the alarms with their statuses and time stamps. Alarms used for fault diagnosis are related to relays and CBs and they consist of statuses and time stamps. Data are in linguistic and numerical form. The linguistic form describes component status (for example relay is activated or relay starts) - R_{start}, relay trip - R_{trip}, circuit breaker open – CB_{open} and circuit breaker closed – CB_{closed}). Numerical form is time-stamp.

The input data are then transferred to the expert knowledge database which is composed of the three parts: the network model, the fuzzy expert database for relays and the fuzzy expert database for circuit breakers. The network model is composed of individual sections. The expert databases for relays and CBs are modeled using fuzzy rules outlined by functional knowledge and causative-consequent mode of component functioning. Since it is not possible to make a diagnosis when only one component status is available (this can be often in practice), a nonexistent status is modeled. The part of the knowledge database is modeled with fuzzy rules that use the model of nonexistent information to model the functionality of components with one known and one unknown information. Thus developed fuzzy expert FDS can diagnose relay and CB operation even when only one status is known because it uses the model of nonexistent information.

After the expert knowledge databases (Fig. 1), the diagnosis of the relay and CB operation is made using fuzzy membership functions in order to diagnose the correct operation, false operation and failure to operate of the appropriate component. A detailed description of the used fuzzy rules can be found in the Chapters 2.2 and 2.3. Once, when the diagnosis of relay and CB operations are made, using Hamacher's fuzzy union operator, the last diagnosis is made i.e. the fault is identified and quantified.

The above-mentioned system can diagnose the correct operation, false operation and failure to operate of the relay and CB and it makes a quantitative diagnosis by the numerical amount of membership function in order to rank a different diagnosis thus helping the operator make the right decision.

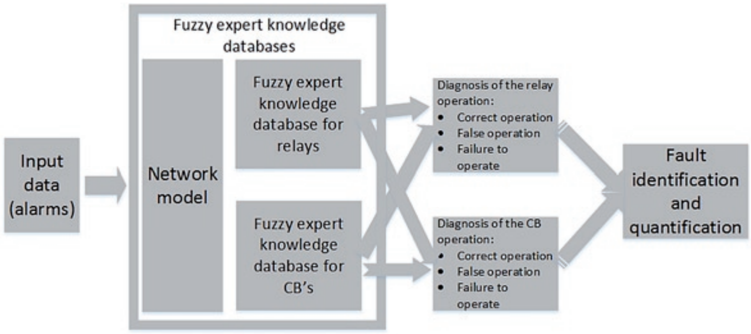


Fig. 1. Block diagram of the developed fuzzy expert FDS

2.2. Fuzzy model for the relay operation diagnosis

The functional operation of the relay is described by means of correct operation, false operation and failure to operate. Expressions (1) and (2) represent a functional description of these operations based on the cause and effect of the fault and related relay and are referred to as logical equations of operations [16, 22]:

$$R_{operate} = R_{start} \cap R_{trip} \tag{1}$$

$$R_{failure\_to\_operate} = R_{start} \cap N_o R_{trip} \tag{2}$$

where: $R_{operate}$ – correct operation of the relay, R_{start} – relay status when it is activated i.e. active relay status, R_{trip} – relay status when it trips, $R_{failure\_to\_operate}$ – failure to operate of the relay and N_0R_{trip} – non-existent trip status of the relay (the relay has not sent the signal for switch off to CB).

Relation (1) represents the correct operation of the relay which means that after the relay is activated R_{start} (at the moment t_{start} when fault occurs) it waits for a certain time and then at the moment that is defined by the so-called setup tripping time t_{sup} it sends the tripping signal R_{trip} to associated CB to open. In the practice, actual trip time of the relay t_{trip} can be slightly different from the set up time t_{sup} (most often t_{trip} is in the time interval ±10% of t_{sup}). Relation (2) is the failure to operate of the relay which means that relay is activated but it has not sent the signal to the circuit breaker. In this case, status R_{trip} is not available, thus the new non-existent trip status N_0R_{trip} is modeled. The second functional operation of the relay i.e. false operation means that relay tripped but outside of the allowed time interval. In practice most commonly allowed time interval is $0.9t_{sup} – 1.1t_{sup}$. This functional operation can't be modeled by previously introduced logical equations but it is modeled using fuzzy logic.

A new way of modeling the functional operation of the relay by fuzzy logic in the time domain is described by the following expressions:

$$\mu_{R\_operate}(t)=\mu_{Rstart}(t)\cap\mu_{Rtrip}(t)\ ,\ t\in\left(0.9t_{\sup},1.1t_{sup}\right) \tag{3}$$

$$\mu_{R\_false\_operation}(t)=\mu_{Rstart}(t)\cap\mu_{Rtrip}(t)\ ,t\in\left[t_{start},t_{sup}\right)\cup\left(t_{\sup},2t_{\sup}\right. \tag{4}$$

$$\mu_{R\_failure\_to\_operate}(t)=\mu_{Rstart}(t)\cap\mu_{NoRtrip}(t)\ ,t\in\left[t_{start},\ \ t_1\right) \tag{5}$$

where: $\mu_{R\_operate}(t)$ – fuzzy membership function that represents the correct operation of the relay, $\mu_{Rstart}(t)$ – fuzzy membership function of the active relay status, $\mu_{Rtrip}(t)$ – fuzzy membership function of the trip relay status, t_{sup} – setup tripping time of the relay, $\mu_{R\_false\ operation}(t)$ – fuzzy membership function that represents the false operation of the relay, t_{start} – the time when relay is activated (relay activation time), $\mu_{R\_failure\ to\ operate}(t)$ – fuzzy membership function that represents the failure to operate of the relay, $\mu_{NoRtrip}(t)$ – fuzzy membership function of the non-existent trip status of the relay and t_1 – arbitrary chosen time which needs to be larger than double setup tripping time t_{sup}.

In order to obtain intersection of fuzzy sets, standard fuzzy intersection defined by the relation (6) is chosen in this paper:

$$\mu_{A\cap B} = \min(\mu_A, \mu_B) \tag{6}$$

where: μ_A – membership function of the fuzzy set A and μ_B – membership function of the fuzzy set B.

Fuzzy membership functions of the active relay status and trip relay status are modeled as trapezoidal shape membership functions. For the fault diagnosis in power system using fuzzy-expert systems, researchers mostly use triangular fuzzy membership function in order to model non-existent state (or alarm signal) [20]. In some other application, the constant membership function is also used [23]. For FDS developed in this paper, a new way of modeling non-existent information consists of using linear membership function (as time passes the value of the membership function to the non-existent state linearly increases) for relay non-existent trip status. The background of this model consists of a combination of deductive (logical) insertion, regression insertion and longitudinal insertion expressed in fuzzy logic and is based on the theory of missing data explained in [5, 31]. The fuzzy membership function of the non-existent trip status of the relay is defined:

$$\mu_{NoRtrip}(t) = \frac{t}{t_1} + a \tag{7}$$

where: a – is intersection of linear membership function and y-axes, $1/t_1$ is the slope of the liner membership function.

Unknown time stamp (unknown time t_n) of this non-existent relay signal can be found from:

$$t_n = \frac{\int_0^{t_1} t\cdot\mu_{NoRtrip}(t)\,dt}{\int_0^{t_1}\mu_{NoRtrip}(t)\,dt} \tag{8}$$

Correct operation of the relay is defined by the membership function $\mu_{R\_operate}(t)$ (relation (3)) which is equal to the standard fuzzy intersection of the relay active status membership function $\mu_{Rstart}(t)$ and relay trip status membership function $\mu_{Rtrip}(t)$ in the time interval of $\pm10\%$ of the relay setup tripping time. Example of the fuzzy memebership function that describes correct operation of overcurrent protection of the relay is shown in Fig. 2. False operation of the relay is defined by the membership function $\mu_{R\_false\_operation}(t)$ (relation (4)) which is equal to the standard fuzzy intersection of the relay active status membership function $\mu_{Rstart}(t)$ and relay trip status membership function $\mu_{Rtrip}(t)$ in the time interval that starts with the relay activation time t_{start} and ends with double setup tripping time $2t_{sup}$. Because the value of membership function $\mu_{Rtrip}(t)$ in the moment $t=t_{sup}$ is one, setup tripping time t_{supp} is excluded from this time interval. Failure to operate of the relay is defined by the membership function $\mu_{R\_failure\_to\_operate}(t)$ (relation (5)) which is equal to the stan-

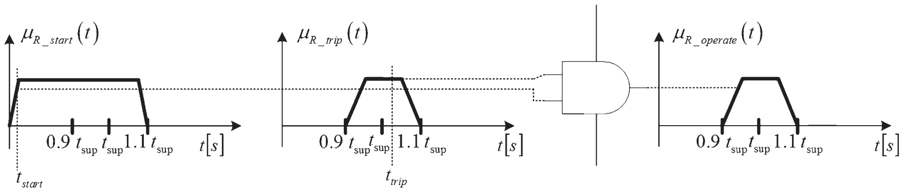


Fig. 2. Example of correct operation of the relay overcurrent protection relay

dard fuzzy intersection of the relay active status membership function $\mu_{Rstart}(t)$ and relay non-existent trip status membership function $\mu_{NoRtrip}(t)$ in the time interval that starts with the relay activation time t_{start} and ends with time t_1 which is arbitrary chosen time (it needs to be larger than double time t_{sup}).

2.3. Fuzzy model for the circuit breaker operation diagnosis

The relays and circuit breakers in the power system are causal-consequently connected due to their functional operation. During the fault, the relay sends signal R_{trip} to the circuit breaker to switch off (to open) and circuit breaker opens the faulted section (line, transformer etc.). The functional operation of the CB is described by means of correct operation, false operation and failure to operate. These actions are described by the logical equations (9-11) derived from the [16, 22]. All relations are combinations of relay status and CB status:

$$CB_{operate} = R_{trip} \cap CB_{open} \tag{9}$$

$$CB_{false\_operation} = N_O R_{trip} \cap CB_{open} \tag{10}$$

$$CB_{failure\_to\_operate} = R_{trip} \cap CB_{close} \tag{11}$$

where: $CB_{operate}$ – correct operation of the circuit breaker, CB_{open} – CB status when it opens or switch off (open status), $CB_{false\_operation}$ – CB false operation, $CB_{failure\_to\_operate}$ – failure to operate of the CB and CB_{close} – CB status when it close (closed status).

Relation (9) represents the correct operation of CB which means that CB opens when it receives the corresponding signal from the connected relay. Expression (10) represents the false operation of CB which means that CB opens without the signal from the connected relay. Expression (11) represents the failure to operate of CB which means that CB receives the signal from the connected relay but it doesn't open.

A new way of modeling the functional operation of the relay by fuzzy logic in the time domain is described by the following expressions:

$$\mu_{CB\_operate}(t) = \mu_{Rtrip}(t) \cap \mu_{CBopen}(t)\,,\ t \in \left(0.9t_{sup} + t_{CBopen}, 1.1t_{sup}\right) \tag{12}$$

$$\mu_{CB\_false\_operation}(t) = \mu_{NoRtrip}(t) \cap \mu_{CBopen}(t)\,,\ t \in \left[0.9t_{sup} + t_{CBopen}, t_1\right) \tag{13}$$

$$\mu_{CB\_failure\_to\_operate}(t) = \mu_{Rtrip}(t) \cap \mu_{CBclosed}(t)\,,\ t \in \left(0.9t_{sup}, 1.1t_{sup}\right) \tag{14}$$

where: $\mu_{CB\_operate}(t)$ – fuzzy membership function that represents the correct operation of the CB, $\mu_{CBopen}(t)$ – fuzzy membership function of the CB open status, t_{CBopen} – time needed CB to open, $\mu_{CB\_alse\_operation}(t)$ – fuzzy membership function that represents the false operation of CB, $\mu_{CB\_failure\ to\ operate}(t)$ – fuzzy membership function that represents the failure to operate of the CB, $\mu_{CBclose}(t)$ – fuzzy membership function of the CB closed status.

The fuzzy membership function of the open CB status is modeled as gamma membership function. Correct operation of the CB is defined by the membership function $\mu_{CB\_operate}(t)$ (relation (12)) which is equal to the standard fuzzy intersection of the relay trip status membership function $\mu_{Rtrip}(t)$ and CB open status membership function $\mu_{CBopen}(t)$ in the time interval that starts with $0.9t_{sup}+t_{CBopen}$ and ends with $1.1t_{sup}$. False operation of the CB is defined by the membership function $\mu_{CB\_false\_operation}(t)$ (relation (13)) which is equal to the standard fuzzy intersection of the relay non-existent trip status membership function $\mu_{NoRtrip}(t)$ and membership function of the CB open status $\mu_{CBopen}(t)$ in the time interval that starts with $0.9t_{sup}+ t_{CBopen}$ and ends with arbitrary chosen time t_1 (it needs to be larger than double tripping time t_{sup}). Failure to operate of the CB is defined by the membership function $\mu_{CB\_failure\_to\_operate}(t)$ (relation (14)) which is equal to the standard fuzzy intersection of the relay trip status membership function $\mu_{Rtrip}(t)$ and membership function of the CB closed status $\mu_{CBclosed}(t)$ in the time interval that starts with $0.9t_{sup}$ and ends with $1.1t_{sup}$. Example of the fuzzy membership function that describes the false operation of the CB is shown in Fig. 3. Fig. 3 shows that fuzzy membership function of the non-existent trip status of the relay is linear.

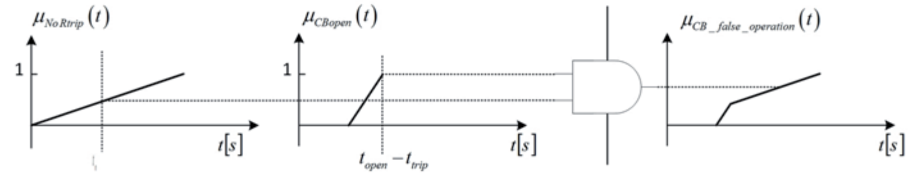


Fig. 3. Example of false operation of the CB

2.4. Fuzzy model for the fault identification and quantification

Fault identification is defined by the action of the relay and circuit breaker since their combination will protect a particular section or component in the electric distribution network. Knowing the information about the functional operation of the relay and/or circuit breaker, the faulted section of the network can be identified. The identification of the fault with proposed FDS is possible even when only one information is known. Nevertheless, if the information about both the relay and the CB are known, the probability that the failure occurred on a particular section that is protected with these specific relay and CB is greater than if only information for one component (CB or relay) is available. Fault identification and quantification using fuzzy logic are defined by the union of fuzzy sets of all CBs and all relays that are activated due to fault based on the expression [4, 12]:

$$\mu_{FDI}(t) = \mu_{R\_operate}(t) \cup \mu_{CB\_operate}(t) \tag{15}$$

where: $\mu_{FDI}(t)$ – fuzzy membership function that represents fault identification and quantification.

Fault identification presented with expression (15) is defined by the union of two fuzzy sets. Many alternative fuzzy union operators (t-conorms) are used for fault identification in electric networks. Standard fuzzy union operator (or maximum operator) is used in [4, 20]. Yager, Hamacher, Dubois and Dombi union operators are used in [19]. For FDS developed in this paper, Hamacher's union operator is chosen:

$$U(\mu_{R\_operate},\mu_{CB\_operate})=\frac{\mu_{R\_operate}+\mu_{CB\_operate}+(r-2)\cdot\left[\mu_{R\_operate}\cdot\mu_{CB\_operate}\right]}{r+(r-1)\cdot\left[\mu_{R\_operate}\cdot\mu_{CB\_operate}\right]} \tag{16}$$

where: $U(\mu_{R\_operate},\mu_{CB\_operate})$ – Hamacher's union of two fuzzy sets (first set is consisted of all relays that are activated due to fault and second set is consisted of all CBs that are activated due to a fault), r – is a positive number (r > 0) which is in this paper set to a value of 1.01.

3. Application of fuzzy expert FDS to the real distribution network

3.1. Model of the distribution network

The distribution network that is used as a model for developed fuzzy expert FDS is shown in Fig. 4 (part of 35 kV and 10 kV real distribution network in Croatia). The network consists of two 35 kV bus (B2 and B3) connected by a 35 kV transmission line (L2), two transformers 35/10 kV (T3 and T4), 10 kV bus (B4) and four 10 kV transmission lines (L3, L4, L5 and L6). The transmission line L2 (35 kV) is protected by the overcurrent (I>), short circuit (I>>) and earth fault protection (U_0I_0) incorporated in relay R1 that controls circuit breaker CB1. Circuit breaker CB2 is manual and it is not controlled by the any of the relays. The 35/10 kV transformers (T3 and T4) are protected by the overcurrent (I>) and differential (3DI) protection incorporated in relays R2 and R3 that control pairs of circuit breakers CB3-CB5 and CB4-CB6 respectively. 10 kV transmission lines (L3 – L6) are protected by the overcurrent (I>), short circuit (I>>) and earth fault protection (U_0I_0) incorporated in relays R4, R5, R6, and R7. These relays are connected to the appropriated circuit breakers (CB7, CB8, CB9, and CB10). Transformers in the 110/35 kV substation (T1 and T2) are the responsibility of the transmission system operator so their relays and circuit breakers are not considered in this paper. The time settings for the protection are shown in Fig. 4. The number of rules that are introduced in FDS for presented distribution network is 78. In this paper, only part of distribution network presented in Fig. 4 is modeled because this part makes one operational and functional section which is connected to one transmission transformer station (110/35 kV). This functional section is radially supplied from the transmission transformer station and is independent of the rest of the distribution network. The whole distribution network consists of many similar sections that are standard. In the practice, fault diagnosis is done for each section individually because the fault in one section doesn't affect other sections. Thus proposed FDS is tested for only one section. For the modeling fuzzy expert fault diagnosis system Prolog software tool is used.

3.2. Example of sagittal diagram for the test network

Sagittal diagrams were introduced for the first time in a power system fault diagnosis in 1997 [9]. They describe the causative consequence relationship of the alarm with the fault location. Most commonly, this causal connection is called the alarm path. In the background of the alarm path, there is a functional connection (cause-and-effect connection) of the fault and the relay and also of the relay and the CB. Sagittal diagrams represent a unified view of the alarm (membership function to a particular state) and fault identification.

Figure 5 shows the example of the sagittal diagram for the 10 kV transmission line L6 (see Fig. 4). The direction from the cause to the consequence goes from the left side to the right, marked by the arrow. The diagram consists of six functional operations; three of them are related to relays, and the remaining three are related to the CBs (correct operation, false operation and failure to operate).

The correct operation of the relay, false operation and failure to operate is shown for all three protections (overcurrent protection, short circuit protection and earth fault protection) that are incorporated in relay R7. Two alarms (states) are required for the diagnosis of functional operation, and these alarms are differently colored in Fig. 5. Light gray rectangles indicate the presence of these alarms on the alarm log, and diagnosis of the functional operation, in that

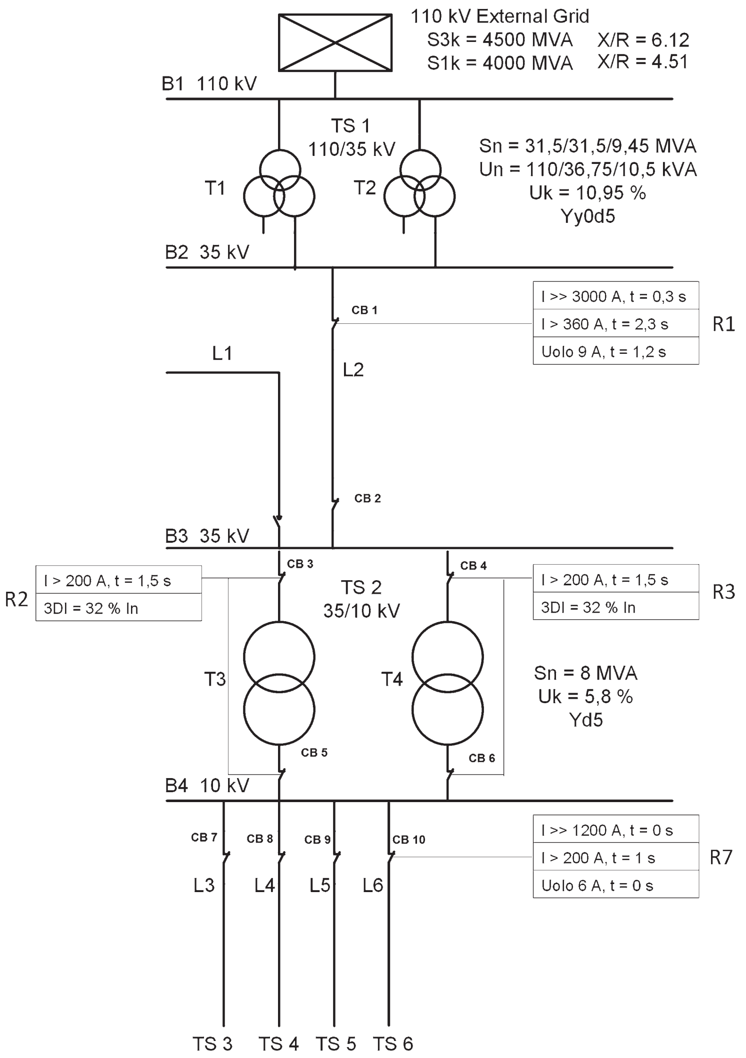


Fig. 4. Single line diagram of the analyzed distribution network

case, is simple. The situation is complicated when one of the necessary alarm is missing, which can happen in the actual distribution networks. In order to make diagnosis in such a situation, non-existent information is modeled as it is explained in section 2.2. Dark gray rectangles represent non-existent alarms (non-existent information;

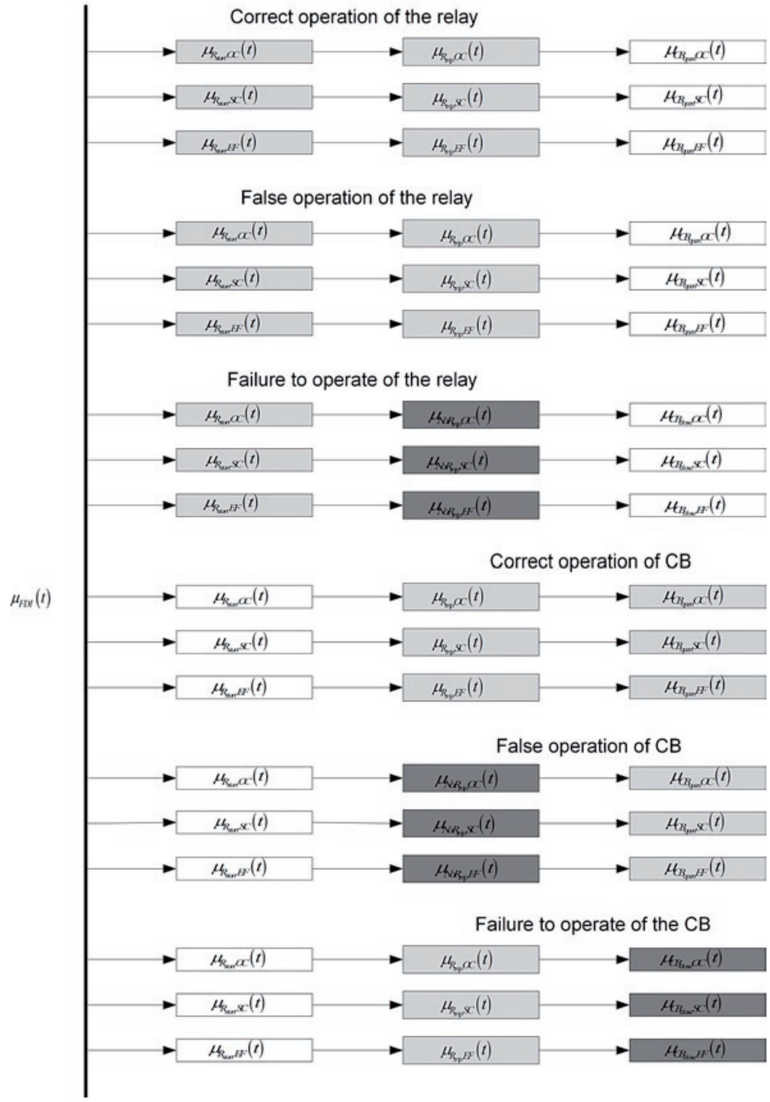


Fig. 5. Sagittal diagram for the line L6

membership function of non-existent state). White rectangles indicate alarms (state, membership function of the state) that are not used in diagnosis, and appear at the beginning or at the end of the alarm path.

From top to bottom in Fig. 5 the correct operation of the relay R7 is shown first. The diagram is further divided into three groups depending on the type of the associated protection: overcurrent (OC), short circuit (SC) and earth fault (EF). After the correct operation, false operation as well as failure to operate of the relay R7 are shown in Fig. 5. Relay R7 is connected to CB10 thus functional operations of CB10 are also shown.

4. Examples of the diagnosis for the real life events

Two examples of fault diagnosis for the real distribution network (Fig. 4) are shown in order to verify the developed FDS. The examples are based on the real-life events and they illustrate the possibilities and effectiveness of the proposed FDS. Fault activates the relay protection that isolates the faulted section by sending tripping to associated circuit breakers. Also, alarms are send to the dispatch center. There the operator can see the alarms through the SCADA system (Supervisory Control and Data Acquisition) which alongside having a possibility of remote management, measurement and control, also has a chronological events recorder (CER). The alarms that reach the dispatch center are visible in the CER and are retained there [38]. The operator needs to make decision how to reconfigurate the network and minimize the consequence of the faults (the intention is that the number of end consumers without the electricity due to fault is minimum while the network is repaired). Exact location of the faulted section is of vital importance for the quick intervention and developed FDS can help the operator in making decision. Actual fault diagnosis in the presented examples is done manually by the operator.

4.1. Example A

The fault occurs on the 10 kV transmission line L6 (actual fault location – AFL at Fig. 6a) and it started a series of alarms that are coming to the distribution network operator that are shown at Fig 7. The fault occurs due to the damage of insulator on a one 10 kV transmission line tower – it is single phase short-circuit.

After the fault occurs, overcurrent protection of the relay R7 reacts first (R_{trip}) and sends a signal to CB10 which opens (CB_{open}) after 47 ms. But overcurrent protection in relay R1 also reacts after 200 ms and sends a signal to the CB1 which opens after 60 ms. Two causes for this kind of events are possible: first, the fault is on the 10 kV line L6 (possible fault location – PFL1 at the Fig. 6a) and second, the fault is on the 35 kV line L2 (PFL2 at Fig. 6a). The operator starts the routine by disconnecting all the 10 kV lines (L3-L6) together with the associated end consumers and then starts to reclose all the equipment one by one in order to find the faulted section. This activated the whole sets of alarms that are shown in Fig. 7. The entire procedure lasted from 20:49:15 till 21:26:15 – approximately 37 minutes. As can be seen from the Fig. 7 at 10:26:15 the same four alarms appeared as in the beginning. This is the signal for the operator that faulted section was found. Fig. 7 shows original alarm list extracted

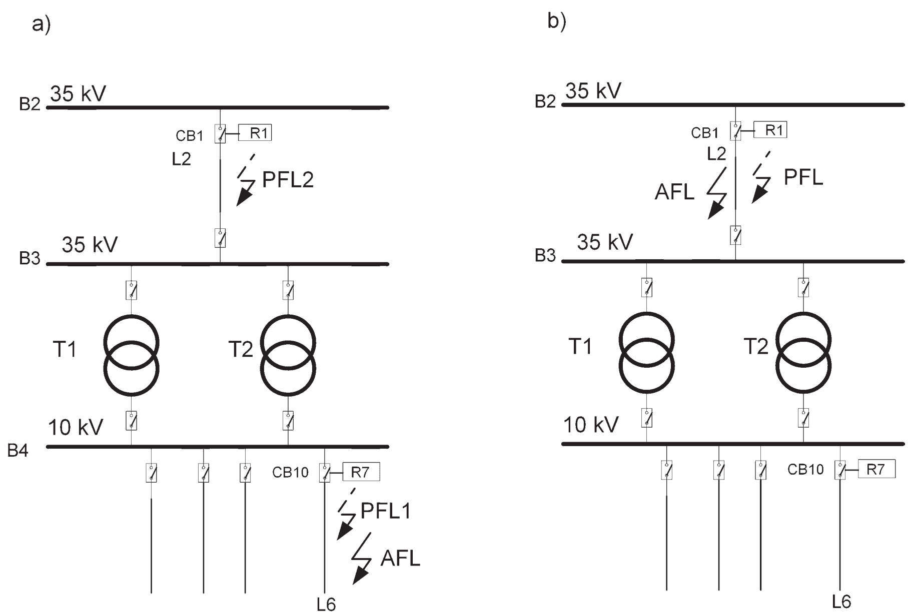


Fig. 6. Possible (PSF) and actual (AFL) fault locations for Examle A and Example B

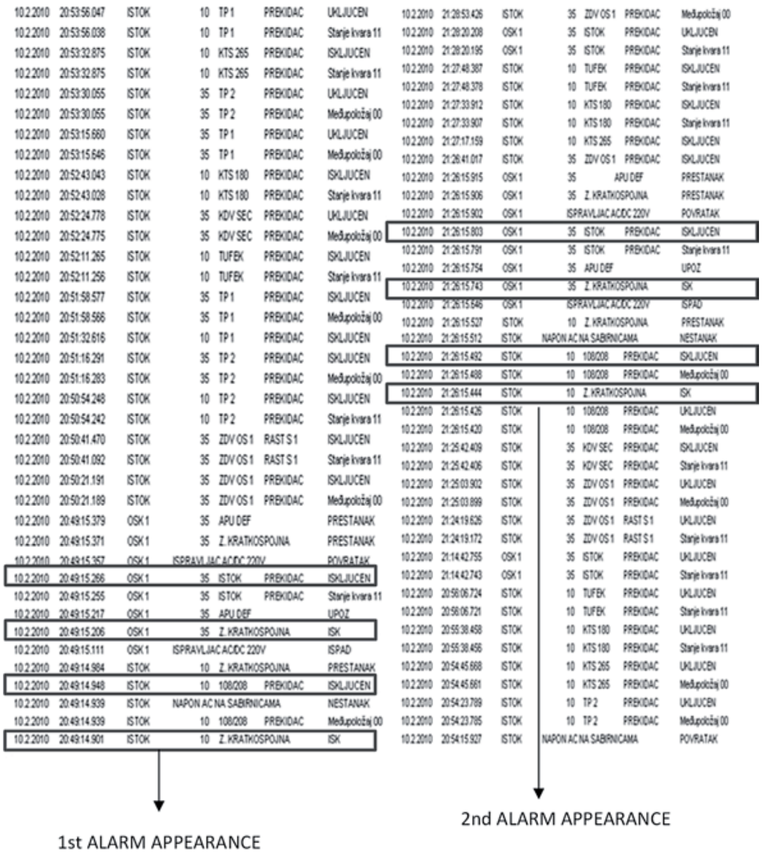


Fig. 7. Original alarm list exctracted from the CER

from the CER and four important alarms are highlighted with the red rectangle. Detail analysis of the fault can be found in [38].

Table 1. Alarm list for the Example A

No	Status	Component	Possible fault location	Time
1.	R_{trip}	R7	L6	20:49:14.901
2.	CB_{open}	CB10	L6	20:49:14.948
3.	R_{trip}	R1	L2	20:49:15.206
4.	CB_{open}	CB1	L2	20:49:15.266

Out of a whole range of alarms (see Fig. 7), only four are relevant for FDS and they are shown in Table 1 together with their statuses, time stamps, component, and location.

Table 2. Diagnose results for the Example A

Diagnose	Membership function
Correct operation of CB10	0.8879
Correct operation of CB1	0.6796
Fault identification on L6	0.8879
Fault identification on L2	0.6796

Developed fuzzy expert FDS can help the operator to make the right decision in uncertain situations, such as in Example A. After the statuses and time stamps of the alarms (shown in Table 1) are entered in FDS, diagnosis results are obtained and presented in Table 2.

Finally, in order to decide at which location (L6 or L2) the fault occurred, the maximum selection criterion is used: based on the results from Table 2, most probable fault location is 10 kV line L6.

Comparing the time needed to diagnose a faulted section when using the proposed FDS and without it (when the diagnosis is performed by the operator manually) it can be concluded that the proposed FDS shortens the diagnosis time because the FDS needed only a few seconds for diagnosing while the operator needed 37 minutes.

4.2. Example B

Unlike Example A where the fault occurred at one location and four relevant alarms arrived, in Example B there was also fault at one location but only one alarm arrived at the operator. The actual location of the fault is 35 kV line L2 (AFL in Fig. 6b). The alarm that arrived is shown in Table 3. It is connected with short circuit protection of the relay R1 (R_{start}).

Table 4. Diagnosis results for the Example B

Diagnose	Membership function
Failure to operate of relay R1	0.663
Fault identification on L2	0.663

From the alarm in Table 3, it can only be concluded that the possible fault refers to the 35 kV line L2 but it is impossible to make a diagnosis about the relay operation with only one status known. Thus unknown trip status is modeled ($N_O R_{trip}$) as it is explained in Section 2.2. Diagnosis results are presented in Table 4.

Based on the results, the operator can conclude, that relay R1 most probably failed to operate (it didn't send a signal to CB1) and that actual fault was on the 35 kV line L2. Diagnosis is made only with one available alarm and its probability is 66.3%.

Table 3. Alarm list for the Example B

No	Status	Component	Possible fault location	Time
1.	R_{start}	R1	L2	14:23:17.129

4. Conclusion

This paper presents the fuzzy expert fault diagnosis system (FDS) for application to distribution networks. Alarms with their statuses and time stamps are used as an input data for diagnosis of correct operation, false operation and failure to operate of the relays and circuit breakers. In the next step, using Hamacher's fuzzy union operator, the diagnosis of the fault identification and quantification is done. The advantage of the proposed FDS is that it is able to make a diagnosis in a situation when only one information is known by modeling the unknown information using a linear fuzzy membership function. The usability of the developed FDS is presented in two examples of the fault diagnosis of the actual events in the real-life distribution network. The first example illustrates the situation when arrived alarms indicate two fault locations and it is necessary that operator makes the decision and chooses the right fault location. The second example illustrates the situation when only one alarm is available and the diagnosis is made using fuzzy modeling of unknown information. Further research and development of the proposed fuzzy expert FDS will seek to complete the diagnosis with other types of information. Further testing of developed FDS will be done on a technically more advanced distribution network which contains fault indicators and digital fault recorders. For that case, new information will be added to the FDS such as statuses of fault indicators as well as currents and voltages of digital fault recorders. There are possibilities to include non-electrical data such as

GPS location data. Also, the FDS will be upgraded in order to make a diagnosis in the distribution networks with a high penetration level of distributed energy sources (like photovoltaic power plants etc.).

References

1. Amiri F, Khadivar A. A fuzzy expert system for diagnosis and treatment of musculoskeletal disorders in wrist. Tehnicki vjesnik - Technical Gazette 2017; 24(1): 147-155.
2. Bedekar P P, Bhide S R, Kale V S. Fault section estimation in power system using Hebb's rule and continuous genetic algorithm. International Journal of Electrical Power & Energy Systems 2011; 33: 457–465, https://doi.org/10.1016/j.ijepes.2010.10.008.
3. Bi T, Wen F, Ni Y, Wu F F. Distributed fault section estimation system using radial basis function neural network and its companion fuzzy system. International Journal of Electrical Power & Energy Systems 2003; 25(5): 377–386, https://doi.org/10.1016/S0142-0615(02)00083-2.
4. Chang C S, Chen J M, Srinivasan D, Wen F S, Liew A C. Fuzzy logic approach in power system fault section identification. IEE Proceedings - Generation, Transmission and Distribution 1997; 144(5): 406-414, https://doi.org/10.1049/ip-gtd:19971278.
5. Daniels M J, Hogan J W. Missing data in longitudinal studies: Strategies for Bayesian Modeling and Sensitivity Analysis, Chapman & Hall/ CRC, 2008.
6. Duan R, Hu L, Lin Y. Fault diagnosis for complex systems based on dynamic evidential network and multi-attribute decision making with interval numbers. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19(4): 580–589, https://doi.org/10.17531/ein.2017.4.12.
7. Ferreira V H, Zanghi R, Fortes M Z, Sotelo G G, Silva R B M, Souza J C S, Guimarães C H C, Gomes S. A survey on intelligent system application to fault diagnosis in electric power system transmission lines. Electric Power Systems Research 2016; 136: 135–153, https://doi.org/10.1016/j.epsr.2016.02.002.
8. Gómez Mu-oz C, García Marquez F, Arcos Jimenez A, Cheng L, Kogia M, Mohimi A, Papaelias M. A heuristic method for detecting and locating faults employing electromagnetic acoustic transducers. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19(4): 493–500, https://doi.org/10.17531/ein.2017.4.1.
9. Hyun-Joon C, Jong-Keun P. An expert system for fault section diagnosis of power systems using fuzzy relations. IEEE Transactions on power systems 1997; 12(1): 342-348, https://doi.org/10.1109/59.574957.
10. Isik M F, Haboglu M R, Yilmaz C, Yilmaz E N. Design and Implementation of Real Time Monitoring and Control System for Distributed Robotic Systems Supported with IOS/Android Application. Tehnicki vjesnik - Technical Gazette 2018; 25(2): 423-428.
11. Jadid S B, Jeyasurya B, Khaparde S A. Power system fault diagnosis expert system using PROLOG. Fourth IEEE Region 10 International Conference TENCON, IEEE; 778–781, https://doi.org/10.1109/TENCON.1989.177053.
12. Jarventausta P, Verho P, Partanen J. Using fuzzy sets to model the uncertainty in the fault location process of distribution networks. IEEE Transactions on Power Delivery 1994; 9(2): 954–960, https://doi.org/10.1109/61.296278.
13. Kimura T, Nishimatsu S, Ueki Y, Fukuyama Y. Development of an expert system for estimating fault section in control center based on protective system simulation. IEEE Transactions on Power Delivery 1992; 7(1): 167–172, https://doi.org/10.1109/61.108904.
14. Leão F B, Pereira R A F, Mantovani J R S. Fast fault section estimation in distribution control centers using adaptive genetic algorithm. International Journal of Electrical Power & Energy Systems 2014; 63: 787–805, https://doi.org/10.1016/j.ijepes.2014.06.052.
15. Lee H-J, Park D-Y, Bok-Shin Ahn, Park Y-M, Park J-K, Venkata S S. A fuzzy expert system for the integrated fault diagnosis. IEEE Transactions on Power Delivery 2000; 15(2): 833–838, https://doi.org/10.1109/61.853027.
16. Liacco T E, Kraynak T J. Processing by logic programming of circuit-breaker and protective-relaying information. IEEE Transactions on power apparatus and systems 1969; 88(2): 171-175, https://doi.org/10.1109/TPAS.1969.292418.
17. Liao Z, Wang D, Tang L, Ren J, Liu Z. A Heuristic Diagnostic Method for a PV System: Triple-Layered Particle Swarm Optimization–Back-Propagation Neural Network. Energies 2017; 10(2): 226, https://doi.org/10.3390/en10020226.
18. McDonald J R, Burt G M, Zielinski J S, McArthur S D J. Intelligent knowledge based systems in electrical power engineering. Boston, MA: Springer US; 1997, https://doi.org/10.1007/978-1-4615-6387-7.
19. Meza E M, de Souza J C S, Schilling M T, Do Coutto Filho M B. Exploring fuzzy relations for alarm processing and fault location in electrical power systems. 2001 IEEE Porto Power Tech Proceedings, https://doi.org/10.1109/PTC.2001.964909.
20. Monsef H, Ranjbar A M, Jadid S. Fuzzy rule-based expert system for power system fault diagnosis. IEE Proceedings - Generation, Transmission and Distribution 1997; 144(2); 186-192, https://doi.org/10.1049/ip-gtd:19970799.
21. Oliveira A L, de Araújo O C B, Cardoso G, de Morais A P, Mariotto L. A mixed integer programming model for optimal fault section estimation in power systems. International Journal of Electrical Power & Energy Systems 2016; 77: 372–384, https://doi.org/10.1016/j.ijepes.2015.11.090.
22. Park Y M, Kim G W, Sohn J M. A logic based expert system (LBES) for fault diagnosis of power system. IEEE Transactions on power systems 1997; 12(1): 363-369, https://doi.org/10.1109/59.574960.
23. Pospiech_Kurkowska S. Processing of Missing Data in a Fuzzy System. In: Pietka E, Kawa J (editors). Information Technologies in Biomedicine. Advances in Soft Computing 2008; 47.
24. Rafinia A, Moshtagh J. A new approach to fault location in three-phase underground distribution system using combination of wavelet analysis with ANN and FLS. International Journal of Electrical Power & Energy Systems 2014; 55: 261–274, https://doi.org/10.1016/j.ijepes.2013.09.011.
25. Rahman M S, Isherwood N, Oo A M T. Multi-agent based coordinated protection systems for distribution feeder fault diagnosis and reconfiguration. International Journal of Electrical Power & Energy Systems 2018; 97:106–119, https://doi.org/10.1016/j.ijepes.2017.10.031.
26. Sekine Y, Akimoto Y, Kunugi M, Fukui C, Fukui S. Fault diagnosis of power systems. Proceedings of the IEEE 1992; 80(5): 673–683, https://doi.org/10.1109/5.137222.

27. Shrivastava S, Jain S, Nema R K, Chaurasia V. Two level islanding detection method for distributed generators in distribution networks. International Journal of Electrical Power & Energy Systems 2017; 87: 222–231, https://doi.org/10.1016/j.ijepes.2016.10.009.
28. Smolinski M, Perkowski T, Mystkowski A, Egidijus D, Eidukynas D, Jastrzebski R P. AMB flywheel integration with photovoltaic system for household purpose. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19(1): 86–94, https://doi.org/10.17531/ein.2017.1.12.
29. Talukdar S N, Cardozo E, Leao Luiz V. Toast: The Power System Operator's Assistant. Computer 1986; 19(7): 53–60, https://doi.org/10.1109/MC.1986.1663279.
30. Topić D, Šljivac D, Stojkov M. Reliability model of different wind power plant configuration using sequential monte carlo simulation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18(2): 237–244, https://doi.org/10.17531/ein.2016.2.11.
31. Wall de T, Pannekoke J, Scholtus S. Handbook of statistical data editing and imputation, Wiley, 2011.
32. Wollenberg B F, Sakaguchi T. Artificial intelligence in power system operations. Proceedings of the IEEE 1987; 75(12): 1678–1685, https://doi.org/10.1109/PROC.1987.13935.
33. Yu D, Chen Z M, Xiahou K S, Li M S, Ji T Y, Wu Q H. A radically data-driven method for fault detection and diagnosis in wind turbines. International Journal of Electrical Power & Energy Systems 2018; 99: 577–584, https://doi.org/10.1016/j.ijepes.2018.01.009.
34. Zhang Y, Zhang Y, Wen F, Chung C Y, Tseng C-L, Zhang X, Zeng F, Yuan Y. A fuzzy Petri net based approach for fault diagnosis in power systems considering temporal constraints. International Journal of Electrical Power & Energy Systems 2016; 78: 215–224, https://doi.org/10.1016/j.ijepes.2015.11.095.
35. Zhao Y, Li D, Dong A, Kang D, Lv Q, Shang L. Fault Prediction and Diagnosis of Wind Turbine Generators Using SCADA Data. Energies 2017; 10(8): 1210, https://doi.org/10.3390/en10081210.
36. Zhou Q, Xiong T, Wang M, Xiang C, Xu Q. Diagnosis and Early Warning of Wind Turbine Faults Based on Cluster Analysis Theory and Modified ANFIS. Energies 2017; 10(7): 898, https://doi.org/10.3390/en10070898.
37. Zimmermann H J. Fuzzy Set Theory—and Its Applications, Fourth Edition. New York: Springer Science+Business Media, 2001.
38. Kaluđer S, Nikolovski S, Majdandžić Lj. Alarm processing in a power system by human expert. Tehnicki vjesnik - Technical Gazette 2013; 20(2): 343-349.Re

Slaven KALUĐER
HEP ODS Elektroslavonija Osijek
Cara Hadrijana 3, 31000 Osijek, Croatia

Krešimir FEKETE
Lajos JOZSA
Zvonimir KLAIĆ
Josip Juraj Strossmayer University of Osijek
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Kneza Trpimira 2B 31000 Osijek, Croatia

E-mails: slaven.kaluder@gmail.com, kresimir.fekete@ferit.hr, jozsa.lajos@ferit.hr, zvonimir.klaic@ferit.hr

Article citation info:
FIEDKIEWICZ Ł, PIELECHA I. Assessment of possible use of the ionization signal for the combustion process diagnostics in a spark ignition combustion engine powered by natural gas. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 630–637, http://dx.doi.org/10.17531/ein.2018.4.14.

Łukasz FIEDKIEWICZ
Ireneusz PIELECHA

ASSESSMENT OF POSSIBLE USE OF THE IONIZATION SIGNAL FOR THE COMBUSTION PROCESS DIAGNOSTICS IN A SPARK-IGNITION COMBUSTION ENGINE POWERED BY NATURAL GAS

OCENA MOŻLIWOŚCI WYKORZYSTANIA SYGNAŁU JONIZACJI DO DIAGNOSTYKI PROCESU SPALANIA W SILNIKU SPALINOWYM O ZAPŁONIE ISKROWYM ZASILANYM GAZEM ZIEMNYM*

The ionization signal, which is a result the presence of ions and electrons in the cylinder space of the internal combustion engine, is affected by many factors, including: temperature, pressure, fuel mixture composition, fuel type, presence of exhaust gases and others. The shape of the signal changes to a large extent from cycle to cycle, which indicates the stochastics of the combustion process. Nevertheless, its analysis provides a lot of useful information, such as the location of the maximum pressure or the maximum heat release rate. Using these signals allows supplementing the limited engine control systems of the combustion process in internal combustion engines. The paper presents a comparative analysis of the gas ionization current signal in the cylinder and the variable pressure at fixed operating points of a single-cylinder, four-stroke engine powered by natural gas. The analysis allowed to determine the relationship between the positions of the maximum thermal ionization signal value and of the maximum combustion pressure value. Additionally the relationship between the position of the maximum thermal fraction derivative and the maximum heat release rate was established.

Keywords: *ionization sensor, engine control, engine diagnostics, indicated pressure, heat release.*

Sygnał jonizacji wynikający z obecności jonów oraz elektronów w przestrzeni cylindra silnika spalinowego jest składową wielu czynników, między innymi: temperatury, ciśnienia, składu mieszanki, rodzaju paliwa, obecności reszty spalin oraz innych. Kształt sygnału zmienia się w znacznym stopniu z cyklu na cykl, co świadczy o stochastyce procesu spalania. Mimo tego, jego analiza dostarcza wielu przydatnych informacji, takich jak położenie maksymalnego ciśnienia czy maksymalnej szybkości wywiązywania się ciepła. Ich wykorzystanie pozwala uzupełnić ograniczone systemy kontroli procesu spalania w silnikach spalinowych. W artykule przedstawiono analizę porównawczą sygnału prądu jonizacji gazów w cylindrze oraz ciśnienia szybkozmiennego przy ustalonych punktach pracy jednocylindrowego, czterosuwowego silnika zasilanego gazem ziemnym. W wyniku analizy uzyskano zależność położenia maksymalnej wartości sygnału jonizacji termicznej od położenia maksymalnej wartości ciśnienia spalania, uzależniono również położenie maksimum pochodnej członu termicznego od położenia maksimum szybkości wywiązywania się ciepła.

Słowa kluczowe: *czujnik jonizacji, sterowanie silnikiem, diagnostyka silnika, ciśnienie indykowane, wywiązywanie ciepła.*

1. Introduction

To assess the internal combustion engine combustion process validity, it is necessary to analyze the thermodynamic indicators, such as pressure, the start and end points of the combustion, the amount of heat released and the heat release rate. The most common way to obtain the above values is to measure the variable cylinder pressure of the engine and its further processing [12, 13, 16, 20]. In order to achieve this it is necessary to use pressure sensors that allow high sampling frequency measurements under high pressure and temperature conditions, and the same goes for the equipment communicating with the sensor. The costs of indicated systems limit their application to only scientific test engines and higher class vehicles. Optical systems are an alternative, available method [18, 19, 21] along with systems based on ionization current measurement in the cylinder [13, 15]. The analysis of a light wave requires using complicated apparatus, similarly as for the variable pressure indication [24]. The ionization current measurement has the widest application possibilities due to the system structure and low cost [4, 5, 8].

The primary purpose of ionization current measurement is to detect ignition failure and occurrence of knocking combustion [11]. These systems are characterized by a faster response time. There are also attempts to use the signal for multiple engine applications, such as: the measurement of the recirculated exhaust gas content in the air/fuel mixture [2, 14], assessment of excess air coefficient in the combusted fuel dose [22, 23], temperature measurement in the cylinder [9], ignition advance angle control with feedback in the form of information on the maximum combustion pressure location [10]. In recent years, research work has been performed to allow using the ionization signal for quality control and combustion control in CI engines running on a homogeneous mixture, also known as HCCI engines [1, 17].

As a result of ignition, the flame front propagates in the mixture from the spark plug electrodes towards the cylinder walls. The strong chemical reactions caused by these events sustain the process, thus

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

leading to the production of ions and free electrons in the so-called chemical ionization phase. As a result of the increase of temperature and pressure in the final phase of the combustion process, the rate of ion formation significantly exceeds their recombination rate, which leads to an increase in their number. This phase is referred to as the thermal ionization phase.

Daniels [3] showed that the rate of ion formation is closely correlated with other significant thermodynamic combustion indicators, such as the ignition angle, the point of maximum flame front acceleration, the point of maximum heat release rate, the crankshaft angle of maximum pressure and the timing of the combustion process end. Eriksson [6] adds that the speed of the so-called the chemical ionization phase is closely related to the composition of the air-fuel mixture.

2. Ionization voltage signal

The measurement of ion number is performed using a spark plug, in the case of SI engines, or a sensor in CI engines. The idea is to create a large voltage potential difference (up to several hundred volts) between the spark plug electrodes, in the time period when the plug is not used to ignite the mixture (Fig. 1).

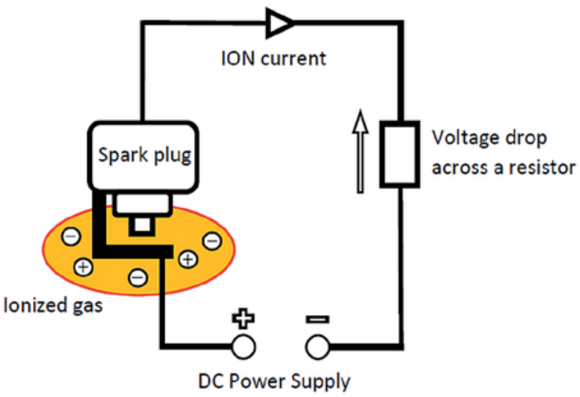


Fig. 1. Diagram of the ionization signal measuring circuit [7]

Due to the ability of ions and electrons to carry electric charge, a current flow appears (ionization current) proportional to the number of ions located in the vicinity of the electrodes. The measurement resistor, connected in series, has a potential difference proportional to the current of ionization according to the Ohm's law. This voltage is used directly by the measuring device and is referred to as the ionization voltage.

The characteristic ionization current signal has three distinct phases (Fig. 2):

a) ignition phase (210) – strong electromagnetic radiation from the ignition system – it causes interference in the ionization signal, making it difficult to analyze,
b) chemical ionization phase (220) – strongly associated with the emerging flame front,
c) thermal ionization phase (230) – is strongly dependent on the maximum pressure and temperature in the cylinder.

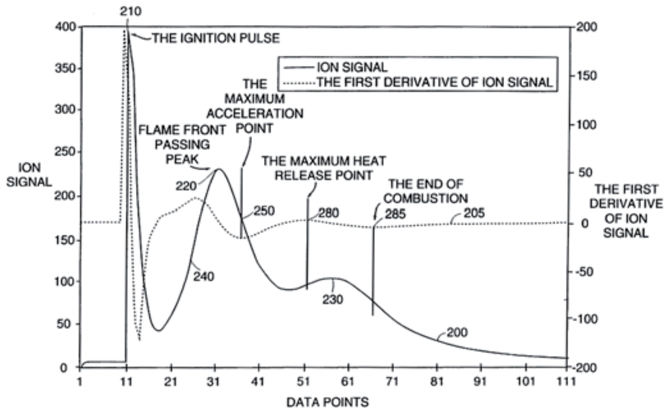


Fig. 2. The ionization voltage signal and its first derivative with the three phases marked [3]

In addition, the ion formation rate expressed as the first derivative of the ionization signal (205) contains the information regarding:

a) point of the maximum flame front acceleration (250) – inflection point occurring after the point of maximum chemical ionization; is the point of maximum flame acceleration, which signifies the end of ion formation in the vicinity of the spark plug electrodes (as a result of the flame front action as well as of further flame propagation into the combustion chamber),
b) the point of maximum heat release rate due to the mixture combustion process (280) – the local temperature around the spark plug increases with the intensity of the combustion process; thus, the ion formation rate in the thermal ionization phase in the vicinity of the spark plug is closely related to the heat release rate around the spark plug,
c) the end point of the combustion process (285) – the process of ion formation stops completely in the vicinity of the spark plug electrodes and the recombination intensity increases instead.

Due to the combustion process stochastics and the high sensitivity of the measurement method, the ionization current characteristics has a very low repeatability in relation to the cylinder pressure characteristic (Fig. 3a), which makes the signal analysis difficult (Fig. 3b). This is caused by the inability to ensure repetitive thermodynamic conditions and the required quality of the mixture in the space between the spark plug electrodes in each cycle of the engine's operation. This applies in particular to [6]:

a) the temperature – it affects the amount of energy provided for the ionization of molecules and the rate of ion recombination,
b) the air to fuel mass ratio – during the combustion of lean mixtures, the signal level decreases, which is caused by the lower combustion process temperature relative to the stoichiometric

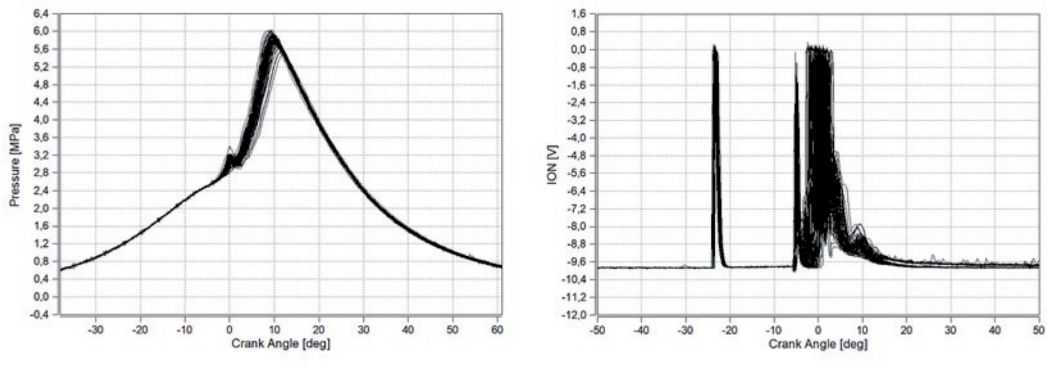


Fig. 3. The non-repeatability of in-cylinder processes: a) combustion pressure, b) ionization voltage signal for 100 consecutive engine cycles (mean indicated pressure IMEP = 0.5 MPa, ignition start crankshaft angle = 6 deg before TDC, excess air ratio λ = 1)

mixture; in addition, the lower density of fuel particles reduces the speed of flame front propagation, which again lowers the ion formation rate,

c) the fuel chemical composition – different types of fuels, depending on the arrangement of the hydrocarbon chains and the types of additives included, significantly affect the process of ions formation and recombination after ignition.

3. Aim of research

Usefulness of the ionization signal applies mainly to engines fueled with a stoichiometric gasoline mixture. There is little information in the literature on the subject of using this signal for combustion process diagnostics during the combustion of gas mixtures.

The aim of this research was to evaluate the possible use of the ionization signal for the diagnosis of thermodynamic processes occurring during the combustion of a fuel-air mixture where methane is used as the fuel. Such diagnostics will be deemed possible if the information it provides regarding the combustion conditions are sufficient to replace the cylinder pressure and the heat release rate characteristics with a diagnostic ionization signal. For this purpose, a comparative analysis of thermodynamic indicators was made, determined on the basis of the indicated pressure characteristics and the ionization voltage signal. The study attempted to establish relations between characteristic points of both methods and their location. Obtaining a correlation between the combustion process thermodynamic indicators and the ionization voltage signal will increase the diagnostic applicability of the signal, while allowing for greater combustion process control. Obtaining a correlation between these signals will also allow to eliminate additional sensors (e.g. combustion pressure), whose processing of the signal in real time in the engine control unit is challenging.

Analysis of the results of such tests will allow to extend the scope of diagnostics of internal combustion engines fueled not only with gasoline-air mixtures, but also with gaseous fuel mixtures.

4. Research methodology

4.1. Test object

The tests were performed on a single-cylinder four-stroke test engine marked as AVL 5804. The originally diesel-powered unit was modified to allow natural gas combustion. In order to achieve this, the injection system was modified (high-pressure direct injection was replaced by low-pressure indirect injection), the ignition system was installed and the compression ratio was reduced. The technical parameters of the test engine used are shown in Table 1.

4.2. Test bench

The tests were performed using specialized measuring equipment (Table 2) and dedicated control devices (Fig. 4). The ignition control (from Autoelektronika) enabled adjustment of the ignition advance angle and the discharge energy (function of charging time of the coil primary winding). An electronically controlled throttle was used to control the air intake. The electromagnetic natural gas injector from Bosch, controlled using the Autoelektonika company equipment, delivered the fuel dose at the specified crankshaft angle and for a specified injection duration. The pressure of the gas supplied from the high-pressure tank was regulated using a reducer and reached a level of 0.9 MPa. In order to limit wave phenomena, an additional volume of 2 dm^3 was installed in the natural gas supply system.

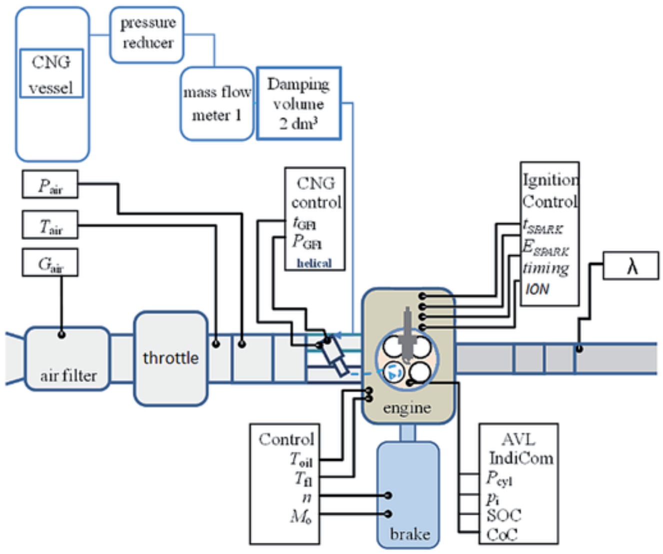


Fig. 4. Ionization signal measurement test bench schematic [7]

In order to maintain constant thermodynamic conditions, the stand was equipped with a liquid and oil conditioning system (constant temperature conditions of T_{ol} = 80°C and T_c = 80°C were maintained). A sensor integrated with the ignition coil, used in Mazda Skyactiv-G engines, was used to obtain the ionization voltage. The Kistler 6081A combustion pressure sensor was placed in the head of the test engine at a distance of 10 mm from the spark plug.

4.3. Scope of research

To determine the relationship between the combustion process thermodynamic indicators and the ionization signal, a test engine operating at a constant speed n = 1500 rpm and an excess air ratio λ = 1 was used. A constant initial value of the engine load in the form of indicated mean effective pressure IMEP = 0.43 MPa was assumed. This value was obtained with an ignition advance angle of 19 degrees before TDC. Fuel injection to the inlet channel was carried out at an

Table 1. The AVL 5804 test engine technical data

Parameter	Unit	Value/type
Engine	–	1-cylinder, 4-valves, SI
Displacement	dm^3	0.5107
Diameter × stroke	mm	85 x 90
Compression ratio	–	15.2
Fuel system	–	Indirect gas injection (electromagnetic injector)
Air intake system	–	naturally aspirated engine

Table 2 . Apparatus used in the research

Parameter	Name	Measurement range
Engine dynamometer	AVL AMK DW13-170	–50–300 Nm
Air intake rate	Sensycon Sensyflow	0–720 kg/h
Fuel consumption	Bronkhorst 111B	0.1–100 g/h
Lubrication system	AVL 577	0–150 °C
Cooling system	AVL 577	0–150 °C
Data acquisition system	AVL IndiSmart	8-channel system
	AVL Concerto	Post-processing
Broadband oxygen sensor	Bosch LSU 4.9	λ > 0.5

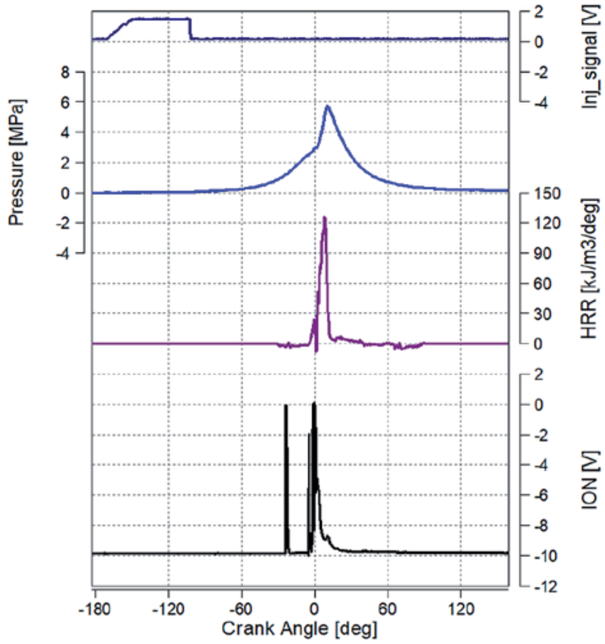


Fig. 5. Indicator chart showing the analyzed parameters

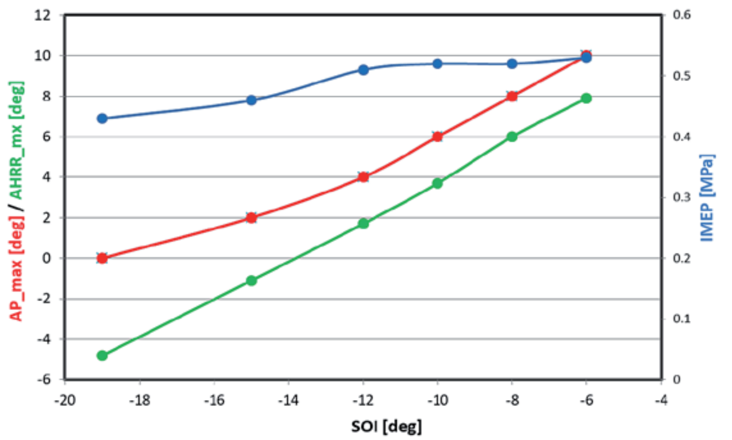


Fig. 6. Influence of start of ignition angle on the averaged values of: indicated mean effective pressure (IMEP), maximum cylinder pressure angle (AP_max) and maximum heat release rate angle (AHRR_mx) in a natural gas engine (λ = 1)

angle of 170 degrees before TDC. The fuel dose was kept constant at q_o = 16.9 mg/injection. The ignition angle (SOI) was a variable. These conditions have caused a change in the combustion process. The indicated mean effective pressure (IMEP), the angle of maximum combustion pressure (AP_mx) and the angle of the maximum heat release rate (AHRR_mx) were analyzed further. The scope of research is shown in Table 3.

The combustion process data acquisition was performed with an angular resolution of 0.1 deg when registering 100 motor cycles, which were then averaged. The non-repeatability of the combustion process indicators was determined as the standard deviation σ from 100 measuring cycles; the resulting values (averaged from 100 cycles) are included in Table 3.

Table 3. Test conditions and average resulting values

No.	Controlled variables			Resulting variables (average values)				
	constant		variable					
	n [rpm]	qo [mg/injection]	SOI [deg after TDC]	IMEP [MPa]	AP_mx [deg after TDC]	AHRR_mx [deg after TDC]	σ(AP_mx) [deg]	σ(AHRR_mx) [deg]
1.	1500	16.9	−19	0.43	0	−4.8	0.601	0.512
2.			−15	0.46	2	−1.1	0.422	0.578
3.			−12	0.51	4	1.7	0.475	0.497
4.			−10	0.52	6	3.7	0.445	0.432
5.			−8	0.52	8	6.0	0.452	0.447
6.			−6	0.53	10	7.9	0.596	0.538

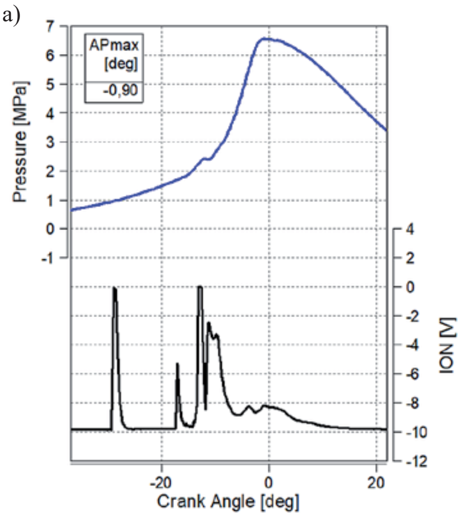
The engine operating conditions shown in Table 3 were used to determine the relationship between the maximum combustion pressure point, the maximum heat release rate point and the ionization signal. Control signals along with the ionization signal and thermodynamic analysis results for an example non-averaged run are shown in Fig. 5.

Test conditions presented in Fig. 5 and Table 3 indicate that the change in the ignition timing angle settings directly affects the combustion process (with other engine operation parameters kept at constant values). As a result of the ignition delay the occurrence of the maximum combustion pressure, and then the angle of the maximum heat release rate are also delayed. Changes in these values are not proportional, which is shown in Fig. 6 for 100 averaged runs.

Due to the increase of the standard deviation value at the extreme start of ignition angle settings, the combustion pressure and ionization signal were analyzed at these operating points (Fig. 7). Data analysis shows that ignition of the mixture results in a characteristic pressure peak. This change corresponds to the other ionization signal characteristics in the combustion chamber.

The use of natural gas as a fuel required maintaining a stable engine operating temperature, due to its high impact on the conditions of the combustion process. Such conditions result from the low ther-

a)



b)

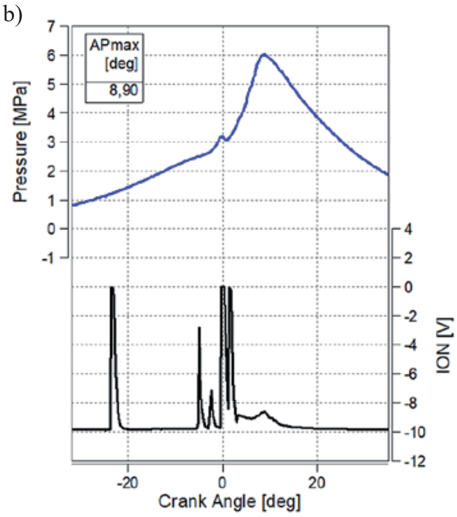


Fig. 7. The indicated pressure and ionization voltage signal characteristics from the combustion chamber for extreme measuring points still within the scope of the tests: a) SOI = 19 deg before TDC, b) SOI = 6 deg before TDC

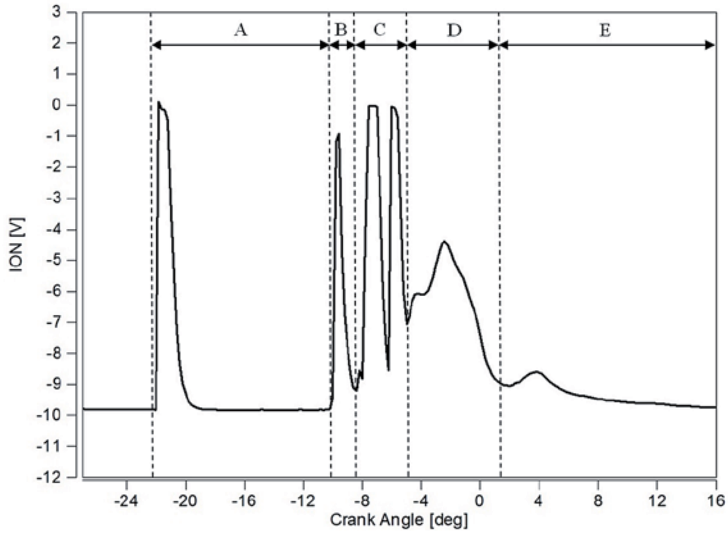


Fig. 8. The real ionization voltage signal along with markers indicating the different phases

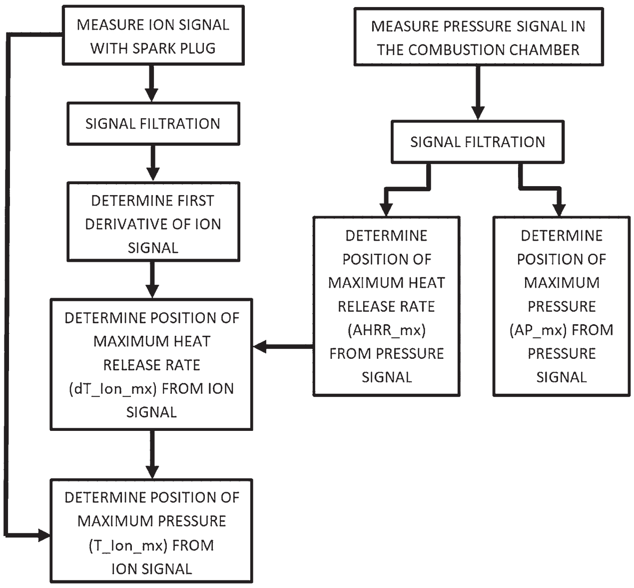


Fig. 9. Calculation algorithm that allows to determine the characteristic points of the ionization signal and its derivative

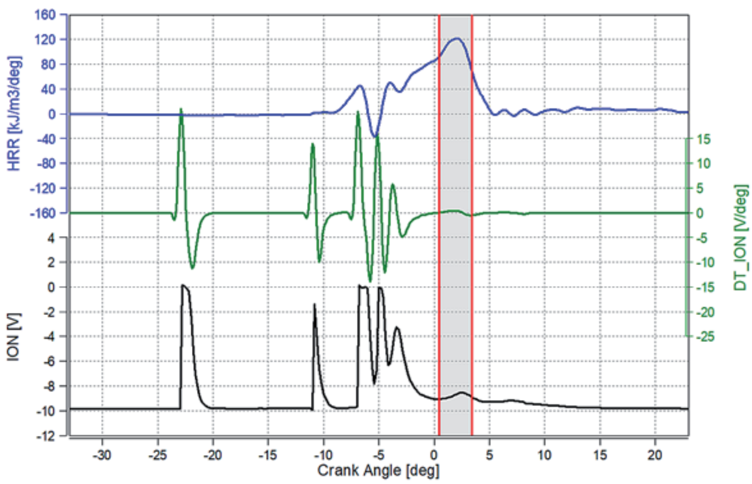


Fig. 10. The ionization signal analysis range (3° on the crankshaft) – marked in red

mal capacity of the natural gas as a fuel. The adopted temperature of 80°C was maintained by the liquid conditioning and oil conditioning systems.

5. Ionization signal test results analysis

5.1. Ionization signal characteristics

The obtained ionization voltage characteristics allow isolation of individual phases, both in the ignition phase and in the mixture combustion phase (Fig. 8). The first notable increase in the ionization voltage corresponds with the moment when the primary winding of the ignition coil starts charging. The duration of the charging phase (A) is limited by the second increase in the ionization signal that initiates the ignition phase (B). This phase results from the ignition coil finishing its charging process and the occurrence of the spontaneous induction phenomenon, which causes electric discharge of the spark plug on the electrodes. As a result of the discharge, the energy accumulated in the ignition coil is lost, resulting in the observed voltage oscillations on the secondary winding. They are reflected in the characteristic increases (C) of the ionization signal which make it difficult to measure the ionization current during the combustion process.

Analysis of the obtained ionization process characteristics in the combustion chamber indicates the existence of some discrepancies between the theoretical ionization signal and the real signal recorded. The real signal contains interference from the ignition system, which makes the analysis of the chemical ionization signal (phase D – Fig. 8) difficult. For this reason, further analysis of the ionization voltage signal concerns mainly its chemical part (phase E – Fig. 8).

This approach causes analyzes related to the ionization voltage signal to concern the combustion process, not the evaluation of the pre-flame processes. The result is that such a signal will be used to evaluate the combustion process, not to assess the timing of ignition or to estimate other quantities (such as the excess air ratio) before ignition in the vicinity of the spark plug.

5.2. Algorithm for determining the process indicators

To find the correlations (presented in chapter 3) a program was created (using the AVL Concerto software) to determine the characteristic points of the ionization signal (chapter 2). Using this calculation algorithm (Fig. 9), in the first stage, the ionization signal and pressure signal were filtered with a low-pass filter. Next, the first derivative (dT_Ion) was determined containing information about the angular position of the maximum heat release rate. The maximum heat release rate value obtained using the pressure characteristic (AHRR_mx) made it possible to limit the ionization signal analysis to within 3-degrees on the crankshaft. In this respect, the algorithm determined the maximum ionization signal value and its derivative (Fig. 10).

The selected range of ±1.5 degrees on the crankshaft in relation to the maximum heat release rate value location determined by using the pressure values, proved highly efficient in determining the characteristic points of the ionization signal (T_Ion_mx) and its derivative (dT_Ion_mx).

5.3. The relation between the ionization signal and the thermodynamic indicators of the combustion process

To determine the relationship between the ionization voltage signal and the thermodynamic indicators of the combustion process (obtained from the cylinder pressure signal) and the heat release rate, the focus was placed on their characteristic values:

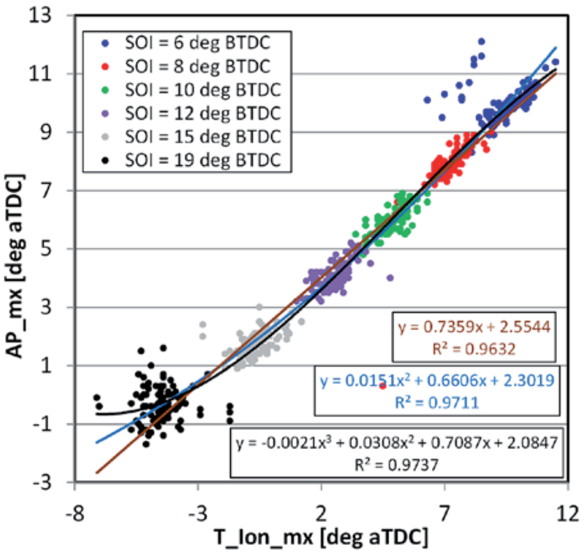


Fig. 11. The relation of the maximum pressure angle (AP_mx) and the maximum value of the thermal ionization voltage angle (T_Ion_mx) including all test points

a) angular position at the maximum cylinder pressure – AP_mx [deg]; this quantity was obtained through indication,
b) the angular position at the maximum heat release rate – AHRR_mx [deg]; this size was obtained using the equation:

$$\frac{\text{AHRR\_mx}}{d\alpha} = \frac{\kappa}{\kappa-1}\left(\frac{P_\alpha + P_{\alpha+1}}{2}\right)\left(V_{\alpha+1} - V_\alpha\right) + \frac{1}{\kappa-1}\left(\frac{V_\alpha + V_{\alpha+1}}{2}\right)\left(P_{\alpha+1} - P_\alpha\right)$$

where:

P – cylinder pressure,
V – volume above the piston,
κ – politropic compression and expansion factor (κ = 1.32),the indexes α and α+1 indicate the current and next crankshaft angle value.

c) angular position at the maximum ionization voltage value in the thermal phase – T_Ion_mx [deg]; this value was obtained using the ionization sensor (Fig. 1) and the designed algorithm (chapter 5.2),
d) the angular position at the maximum derivative value of the thermal phase ionization signal – dT_Ion_mx [deg]; this size was obtained using the designed algorithm.

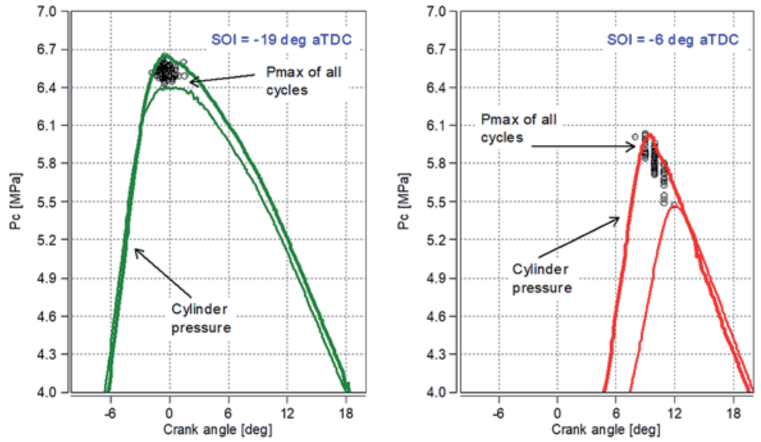


Fig. 12. Changes in the AP_mx angle value at the extreme ignition angle (SOI = 19 before TDC and 6 deg before TDC)

The analysis of the relationships between the values of AP_mx and T_Ion_mx as well as AHRR_mx and dT_Ion_mx reveals that it is possible to make a comparison between them and thus search for correlation. The analysis of the relationship between the maximum pressure crankshaft angle and the maximum value of the thermal ionization signal angle indicates a large correlation of these values with all the research points (Fig. 11).

These correlations are presented in relation to the linear function, quadratic and third order functions (logarithmic and exponential functions were not used due to the presence of negative values of both variables). Their determination coefficients have a similar value of 0.97. The differences between them are within 3%. This means that it is possible to adopt a linear function to determine the maximum cylinder pressure angle value based on the ionization signal. This function is also more useful in the implementation of such a solution, because it allows to increase the speed at which the AP_mx value can be de-

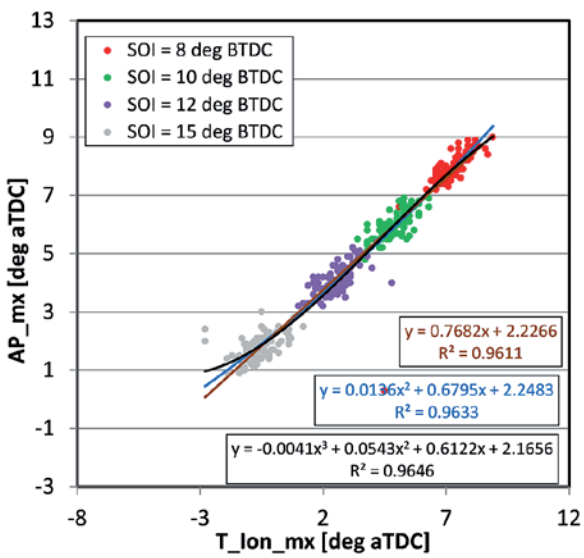


Fig. 13. The relation between the maximum pressure angle (AP_mx) on the crankshaft angle at the maximum thermal ionization voltage (T_Ion_mx) after reducing the number of test points

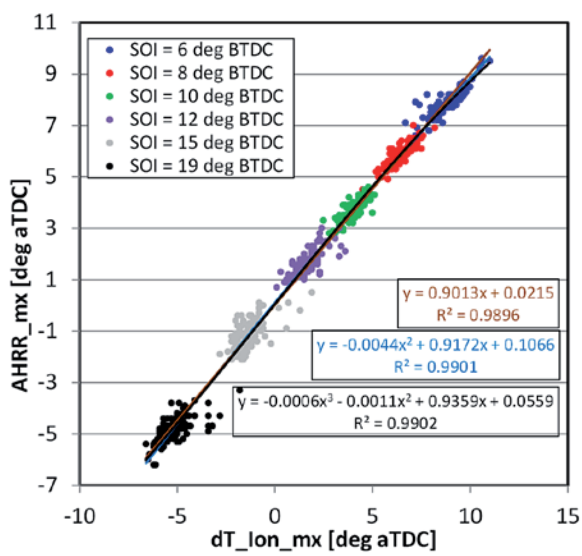


Fig. 14. The relation between the angle at maximum heat release rate (AHRR_mx) and the angle at the maximum derivative of the thermal phase (dT_Ion_mx) for all test points

termined in the engine controller in real time (shorter calculation time using a specific algorithm).

However, for extreme ignition timing values (SOI = 19 deg before TDC and 6 deg before TDC), large discrepancies in cylinder pressure were observed. This results in a significant variation in the angle value at the maximum cylinder pressure (Fig. 12). The measure of this dispersion (indicator) are the values of standard deviation σ(AP_mx) of 0.601 deg and 0.596 deg respectively. They are the largest values of standard deviation when compared to other research points (Table 3). As a result, the correlation between the AP_mx and T_Ion_mx signals diminishes for the extreme values of SOI.

Taking into account the above analyzes forced a limitation of the data range used to determine the relationship between AP_mx and T_Ion_mx signals, the criterion for selecting test points was that the value of standard deviation had to be below 0.5 deg. This limitation therefore requires not taking into account the extreme values of the start of ignition angle. The results of such analyzes are presented in Fig. 13.

The analysis of the results from Fig. 13 indicates the achievement of determination coefficients at the level of 0.96 (within the margin of error of 1%). It follows that it is possible to take into account a limited number of measurement data and adopt the linear relation of the function AP_mx = f(T_Ion_mx) for them. The determination coefficient of 0.9611 and the standard deviation σ(AP_mx) below 0.5 deg guarantee that the condition of linearity of these variables is met.

The analysis of the function AP_mx = f(T_Ion_mx) indicates the possibility of correlation between these variables considering only the standard deviation of the maximum cylinder pressure angle below 0.5 deg. This means that in order to obtain a specific correlation of these variables it is necessary to determine the above standard deviation and to adopt the criterion of its upper value limit.

The analysis of the relation between the angle at the maximum heat release rate (AHRR_mx) and the angular position of the maximum value of the thermal phase derivative (dT_Ion_mx) indicates a highly linear relationship (Fig. 14) for all the test points. In this case, the linear, quadratic and tertiary functions were also determined. For all of these considerations, the obtained determination coefficients are similar and the discrepancies are below 0.1%.

The obtained values of standard deviation σ(AHRR_mx) below 0.6 deg enable using all of the research points to determine the relationship between the values of AHRR_mx and dT_Ion_mx.

The determination coefficients for the function AHRR_mx = f(dT_Ion_mx) show a high correlation of thermodynamic signals with ionization voltage signals, for all test points. The higher value of the maximum heat release rate angle is linearly dependent on the angle at the maximum thermal ionization phase derivative value. Due to the lack of significant differences in the determination coefficients, using a linear relationship was proposed, since it can allow for faster determination of the thermodynamic indicators (angle at the maximum heat release rate) in real time.

Analysis of the standard deviation of both these relations indicates that the correlations obtained using a standard deviation below 0.6 deg are valid. Accepting the deviation AP_mx with a value of 0.596 deg (SOI = 6 deg before TDC) indicates the existence of a limited correlation. Adopting the deviation AHRR_mx with a value of 0.578 deg still allows to obtain a good correlation. Such small differences in standard deviations mean that further tests may be necessary to precisely specify the numerical criterion and limit values for determining the correlation of these variables.

The analyzes presented above indicate the possibility of replacing the selected thermodynamic engine performance indicators with ionization voltage signals, which are possible to achieve with much simpler methods than their thermodynamic counterparts.

6. Conclusions

The analysis of the test results indicates the possibility of using the ionization voltage signal to diagnose the combustion process of the spark-ignition engine powered by natural gas with the excess air ratio of $\lambda = 1$.

The performed tests and analyses have shown that:

1. There is a strong relationship between the angle at the maximum thermal ionization signal value and the angle at maximum combustion pressure – the determination coefficient is $R^2 = 0.9611$. However, this relationship makes it possible to reproduce the angle of maximum combustion pressure based on the ionization voltage signal using a linear (proportional) relationship between signals only in a limited range of the ignition advance angle (8-15 degrees before TDC).
2. There is a strong relationship between the angle at the maximum thermal phase derivative value and the angle at the maximum heat release rate – the determination coefficient is $R^2 = 0.9896$, when using all the research points. This dependence makes it possible to determine the crankshaft angle at the maximum heat release rate based on the ionization signal value by adopting a linear (proportional) relationship between the two signals.
3. It is necessary to precisely determine the criteria for the operation of the internal combustion engine in order to obtain a high correlation of these signals. The standard deviation value limit used in the previous analyzes may be one of such criteria.

The obtained research results indicate that there is some merit to using the ionization signal in modern diagnostic systems for gasoline-powered internal combustion engines and their control systems. The ionization signal obtained in each combustion engine cycle – strongly correlated with the cylinder pressure and the heat release rate characteristics – allows a precise control of the indicated parameters, contributing to a quick detection of incorrect cycles and improvement of the combustion engine performance indicators.

Further research on this subject will focus on the possibility of extending the applications of the ionization signal to include all other test points and eliminating the information noise associated with the electric discharge on the spark plug, which will also enable the diagnosis of the pre-flame phase of the combustion process (chemical ionization phase). The solution to this problem will allow to expand the applications of the ionization signal by measuring the quality of the air-fuel mixture, as well as to measure the temperature in the cylinder. These research results can significantly contribute to improving the combustion process control in order to improve the performance indicators of spark-ignition engines fueled with natural gas and, in result, to reduce the emission of toxic compounds.

Acknowledgements

The research presented in the article was conducted as part of the statutory work nr 05/52/DSPB/0261.

References

1. Butt R H, Chen Y, Mack J H, Saxena S, Dibble R W, Chen J Y. Improving ion current of sparkplug ion sensors in HCCI combustion using sodium, potassium, and cesium acetates: Experimental and numerical modeling. Proceedings of the Combustion Institute 2014; 35(3), doi: 10.1016/j.proci.2014.06.084.

2. Byttner S, Holmberg U. Closed-loop control of EGR using ion currents. Proceedings of the 27th IASTED International Conference Modelling, Identification and Control 2008; 978-0-88986-711-6.
3. Daniels C F. Mass fraction burned and pressure estimation through spark plug ion sensing. Patent US 6089077 A (2000).
4. Dev S, Sandhu N S, Ives M, Yu S, Zheng M. Ion current measurement of diluted combustion using a multi-electrode spark plug. SAE Technical Paper 2018; 2018-01-1134, doi:10.4271/2018-01-1134.
5. Dong G, Chen Y, Wu Z, Li L, Dibble R. Study on the phase relation between ion current signal and combustion phase in an HCCI combustion engine. Proceedings of the Combustion Institute 2015; 35 (3): 3097–3105, https://doi.org/10.1016/ j.proci.2014.08.033.
6. Eriksson L. Methods for ionization current interpretation to be used in ignition control. Diploma Thesis 1995.
7. Fiedkiewicz Ł, Pielecha I, Wisłocki K. Use of the gas ionization signal for combustion process diagnostics in the cylinder of a spark ignition engine. Combustion Engines 2017; 171 (4): 196-200, doi: 10.19206/CE-2017-433.
8. Filipek P, Wendeker M, Kamiński T, Mitraszewska I, Nowacki G. Opinion of ionization signal to estimating composition of mixture in combustion process about engine SI. Journal of KONES Powertrain and Transport 2008; 15 (3): 121–127.
9. Gao Z, Wu X, Man C, Meng X, Huang Z. The relationship between ion current and temperature at the electrode gap. Applied Thermal Engineering 2012; 33–34: 15–23, https://doi.org/10.1016/j.applthermaleng.2011.07.014.
10. Hellring M, Holmberg U. An ion current based peak-finding algorithm for pressure peak position estimation. SAE Technical Paper 2000; 2000-01-2829, https://doi.org/10.4271/2000-01-2829.
11. Hung D, Zhu G, Danne N, McKoskey J. Knock detection for a large displacement air-cooled V-twin motorcycle engine using in-cylinder ionization signals. SAE Technical Paper 2008; 2008-32-0028, https://doi.org/10.4271/2008-32-0028.
12. Koszałka G. Model of operational changes in the combustion chamber tightness of a diesel engine. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (1): 133–139.
13. Kruczyński S, Ślęzak M, Gis W, Orliński P. Evaluation of the impact of combustion hydrogen addition on operating properties of self-ignition engine. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (3): 343–347, http://dx.doi.org/10.17531/ein.2016.3.4.
14. Kusuhara T, Shinkai T, Yoshida K, Langley D. Development on internal EGR feedback control based on ion current. SAE Technical Paper 2017; 2017-01-0793, doi:10.4271/2017-01-0793.
15. Lagana A A M, Lima L L, Justo J F, Arruda B A, Santos M M. Identification of combustion and detonation in spark ignition engines using ion current signal. Fuel 2018; 227: 469–477, doi: 10.1016/j.fuel.2018.04.080.
16. Mikulski M, Wierzbicki S, Piętak A. Zero-dimensional 2-phase combustion model in a dual-fuel compression ignition engine fed with gaseous fuel and a divided diesel fuel charge. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2015; 17 (1): 42–48.
17. Phan T, Mack J H, Butt R H, Dibble R. A comparison of three ion sensing circuits in a homogeneous charge compression ignition engine. Combustion Science and Technology 2017; 189 (8), https://doi.org/10.1080/00102202.2017.1294587.
18. Pielecha I, Pielecha J, Skowron M, Mazanek A. The influence of diesel oil improvers on indices of atomisation and combustion in high-efficiency engines. Polish Maritime Research 2017; 24 (3): 99–105, http://doi.org/10.1515/pomr-2017-0094.
19. Pielecha I, Skowron M, Mazanek A. Evaluation of the injectors operational wear process based on optical fuel spray analysis. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (1): 83–89, http://dx.doi.org/10.17531/ein.2018.1.11.
20. Pielecha I, Wisłocki K, Cieślik W, Bueschke W, Skowron M, Fiedkiewicz Ł. Application of IMEP and MBF50 indexes for controlling combustion in dual-fuel reciprocating engine. Applied Thermal Engineering 2018; 132: 188–195, http://dx.doi.org/10.1016/j.applthermaleng.2017.12.089.
21. Piernikarski D, Hunicz J, Komsta H. Detection of knocking combustion in a spark ignition engine using optical signal from the combustion chamber. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2013; 15 (3): 214–220.
22. Rao R, Honnery D. A simplified mechanism for the prediction of the ion current during methane oxidation in engine-like conditions. Combustion and Flame 2015; 162 (7): 2928–2936, https://doi.org/10.1016/j.combustflame.2015.03.011.
23. Rivara N, Dickinson P B, Shenton A T. A transient virtual-AFR sensor using the in-cylinder ion current signal. Mechanical Systems and Signal Processing 2009; 23 (5): 1672–1682, https://doi.org/10.1016/j.ymssp.2009.01.004.
24. Wendeker M. Sterowanie wtryskiem w silniku samochodowym. Lublin: Lubelskie Towarzystwo Naukowe, 1999.

Łukasz FIEDKIEWICZ
Ireneusz PIELECHA
Institute of Combustion Engines and Transport
Poznan University of Technology
Piotrowo 3, 60-965 Poznan, Poland

E-mails: Lukasz.M.Fiedkiewicz@doctorate.put.poznan.pl,
Ireneusz.Pielecha@put.poznan.pl

Article citation info:
BOGDEVICIUS M, ZYGIENE R. Research of dynamic processes of the system "Vehicle – Track" using the new method of vehicle wheel with metal scale. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 638–649, http://dx.doi.org/10.17531/ein.2018.4.15.

Marijonas BOGDEVICIUS
Rasa ZYGIENE

RESEARCH OF DYNAMIC PROCESSES OF THE SYSTEM "VEHICLE – TRACK" USING THE NEW METHOD OF VEHICLE WHEEL WITH METAL SCALE

BADANIE DYNAMICZNYCH PROCESÓW ZACHODZĄCYCH W UKŁADZIE „POJAZD-TOR" Z WYKORZYSTANIEM NOWEJ METODY DLA KÓŁ Z METALOWĄ ŁUSKĄ

Mathematical models of vehicle wheel with metal scales are introduced in this article. When analysing the interaction between vehicle wheel with a metal scale and rail in the system "Vehicle – Track", the changes of the kinematic and dynamic parameters of the wheel and rail contact points in time are examined, depending on the height of the 2 mm metal scale, when the length of the metal scale is 100 mm and the speed of movement is $V = 40 - 100$ km/h. The results obtained after the research of the system "Vehicle – Track", when the wheel has a metal scale, help to better understand and evaluate the impact of metal scale on wheel on dynamic loads of rail and vehicle and the regularities of their movement. The appearance of a metal scale on the wheel's surface causes technical and maintenances problems for the rolling stock. Railway standards limit the speed of movement that depends on a certain size of metal scale.

Keywords: *rail-wheel interaction; spatial model of metal scale; contact area; impact force; friction force; vibration.*

W niniejszym artykule przedstawiono modele matematyczne koła pojazdu szynowego z powstałą w wyniku zużycia metalową łuską. Analizując oddziaływania pomiędzy kołem pojazdu z łuską a szyną w układzie "pojazd–tor", badano zmiany kinematycznych i dynamicznych parametrów punktów kontaktu koła z szyną zachodzące w czasie, w zależności od wysokości metalowej łuski (2 mm), przy długości łuski 100 mm i zakresie prędkości ruchu pojazdu V = 40–100 km/h. Wyniki uzyskane w badaniu układu "pojazd–tor" dla kół na powierzchni których powstała metalowa łuska, umożliwiają lepsze zrozumienie oraz ocenę wpływu łuski na dynamiczne obciążenia szyny i pojazdu oraz prawidłowości ruchu pojazdu. Pojawienie się metalowej łuski na powierzchni koła powoduje problemy techniczne i obsługowe w utrzymaniu ruchu taboru kolejowego. Normy kolejowe ograniczają prędkość ruchu pojazdów szynowych, uzależniając ją od rozmiaru łuski.

Słowa kluczowe: oddziaływania koło–szyna; przestrzenny model łuski metalowej; obszar kontaktu; siła uderzenia; siła tarcia; wibracja.

1. Introduction

The precise operation of modern machines and vehicles can only be ensured by complex measures, which consist of modern construction solutions for design, the use of high-quality and suitable operational and structural materials, and quality maintenance, which is done on time.

Multicriteria decision-making is widely used in all areas. Multicriteria assessment models are presented in the field of transport [24, 27]. Using the [20] multicriterion additive model, it was found that the main parameter of the technical condition of the railway track is the speed of the wagon.

Initial surface of vehicle wheel profile is symmetric. The radius of vehicle wheel is equal at all points. However, during exploitation, roughness appears on the surfaces of wheel and rail profiles, due interaction between vehicle wheel and rail, and their different geometrical surfaces. Furthermore, the wheel profile constantly changes and becomes asymmetrical.

Damages on surfaces of wheel and rail mostly appear due to their interaction. However, they may also occur due to various manufactory inaccuracies and poor quality of machined parts. Wheel and rail damages are rarely found on initial stages of exploitation, and over time, the damages increase and may cause irreversible consequences. To reduce the risk of train accidents, railway traffic risk management models are being developed [6], which enable railway managers to improve their traffic safety strategy by determining the priorities of the required measures.

Most of the vehicle wheel damage occurs due to braking, surface roughness, temperature differences, and other factors. Metal plasticity methods are used to evaluate the mechanical condition (determine strength and plasticity characteristics) of solid moulding wheels of the wagon [28]. The theoretical angular speeds of the loaded wheels are different due to disproportionate loads and although they are forced to move at the same speed as the wagon, the wheels are sometimes forced to tow and even slip unevenly.

The most common damage in wheel is flat, which occurs due to wheel slip or stuck brake pads [31]. Increase in wheel damages has been examined in the article [10].

Metal scales on wheel occur due to thermos-mechanical damages. Intensive plastic metal deformation appears due to sudden braking, short-term wheel skidding and jumping, when the metal of the wheel suddenly heats up and then suddenly cools down. Sudden braking means the braking of the train in exceptional cases, in these circumstances the largest braking force is used when the air is released from the brake pipe [26].

Many metal scales can appear on the rolling surface of the wheel, which can have one or more layers and differ in height. U-shaped scales are the most common to occur. The damages of scaled wheel are measured from the surface intact to the highest point of the scale. During the exploitation, the metal scale can become even more layered and can even come off the wheel surface. The hardness of metal scale is about 900 HV, which is typical for tempered steel with high residual stress.

In the case of scaled wheel, large impact forces appear in the contact between the wheel and rail, which load the rail and wheelset bearings. Wheel flats frequently appear on the rolling surface of the wheel due to metal scales.

The metal scale on wheel may form when the wheel is broken extremely. Due to large frictional forces and increased temperatures, the metal on the wheel melts and a certain layer of metal is cut off. A flat develops in the contact zone, and the molten metal flows outside the contact zone and to the surface of the wheelset. If the wheel continues to rotate, the moulded metal cools off and continues to deform when it comes in the contact with the rail. This is how the multi-layered metal scale forms on the surface of wheel (Fig. 1, Fig. 2).

Railway vehicle exploitation in Lithuanian Railways (JSC Lithuanian Railways) [22] is prohibited, when wheel has a scale with height higher than 0.5 mm in passenger cars and 1 mm in freight cars.

The authors [2-5, 27] usually choose flat as vehicle wheel damage to simplify the mathematical models, as it is the most common damage in vehicle wheel.

The contact zone of wheel and rail is described as a point and the geometry of the wheel as an analytical function in the mathematical models of interaction between wheel and rail [1, 15, 17, 18, 19, 20, 21, 22, 23].

Researchers [9, 16, 29, 30] describe variations of mass accelerations and displacements of wagon and rail in time or path length, in their studies of dynamical processes that occur in the interaction between wheel and rail. It was determined that the maximum acceleration value is obtained in the case of a wheel with a flat.

Parameters of the deformation of the isolated sinusoidal shape and the influence of the wagon speed on the vertical vibration of the body are evaluated in the dynamic models of interaction between the vehicle and track [12].

The influence of contact forces on the deformation of rolling carload wheels and rails, and the influence of this deformation on the redistribution of the contact stresses is also investigated [25].

Theoretical and practical study of the surface roughness of interaction between rail and train defined the values the surface roughness, when interacting with different rail profiles that affect the increase in exploitation time [11].

Research has been carried out [8] to improve the safe operation of trains and increase the efficiency of load assessments.

Rail condition is assessed [7,13,14] by using the movements of axle-box and bodies, estimating and analysing accelerations.

The authors of this article have failed to find other researches about metal scale in wheelset wheel.

The article consists of three chapters. In the first chapter the system „Vehicle – Track" and literature relevant to its elements are analysed. In the second chapter authors present a mathematical model of wheelset wheel with a metal scale and briefly describe the model of system „Vehicle – Track" and present a calculation algorithm for this system. The third chapter consists of the results of the calculations and discussion.

This paper presents mathematical models of a vehicle wheel with a 2 mm metal scale. The mathematical model allows examination of the interaction between wheel with a metal scale and the rail and displays its effect on dynamic loads. The new mathematical model allows evaluation of the rotation of the wheel around its longitudinal axis Y, in order to evaluate the rotation of the wheel and determine its slip on the rails. In this article, the authors seek to determine the slip of the wheel with a defect, when the defect is a metal scale and without a wheel defect. In the case of sliding the wheel in relation to the rail, frictional forces occur, and heat is released, which increases the wheel and rail temperature in the contact area.

The paper shows results of the changes in kinematic parameters of a wheel (angular velocity and acceleration, velocities, accelerations and other parameters) and dynamic parameters (forces and other parameters), depending on the geometry of metal wheel scale, movement speed and other parameters of the system "Vehicle – Track".

2. Materials and Methods

2.1. The scope of calculations and numerical characteristics of an adhesive joint

During exploitation, the metal scale on wheel can take on various forms. Therefore, this article presents a method for shape generation of metal scale of a wheel and allows generation of a wheel profile with a metal scale if the exact geometry is known.

The mathematical model created in respect to the geometry of metal scale shown in Figure 1 is described in this section. The metal scale of wheelset wheel that consists of three layers is shown in Figure 1 (real top view), where N_{layer} are the layers of metal scale.

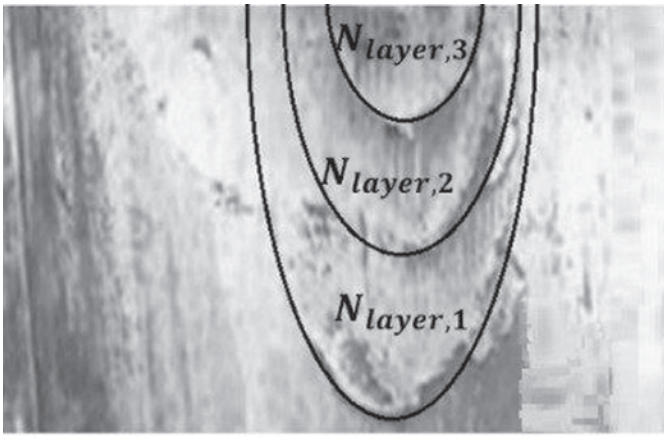


Fig. 1. The metal scale of wheelset wheel, that consists of three layers (real top view)

Metal scales are rarely included in the calculations of mathematical models of railway system "Vehicle – Track", due to complex and different geometrical shapes (Fig. 2). There are only a few known scientific articles about wheel damage, called metal scale, and its effects on the system "Vehicle – Track".

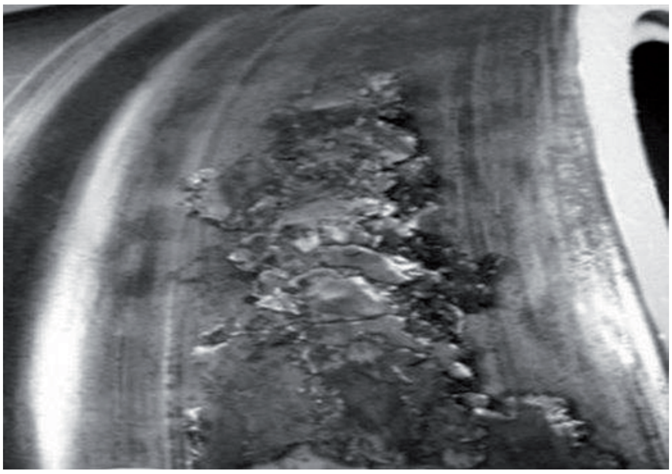


Fig. 2. Multi-layered metal scale of wheelset wheel

The developed mathematical model allows generating more metal scales on wheel and more layers of the metal scale (Fig. 2).

The geometry of the wheel and the metal scale on it is shown in Figure 3.

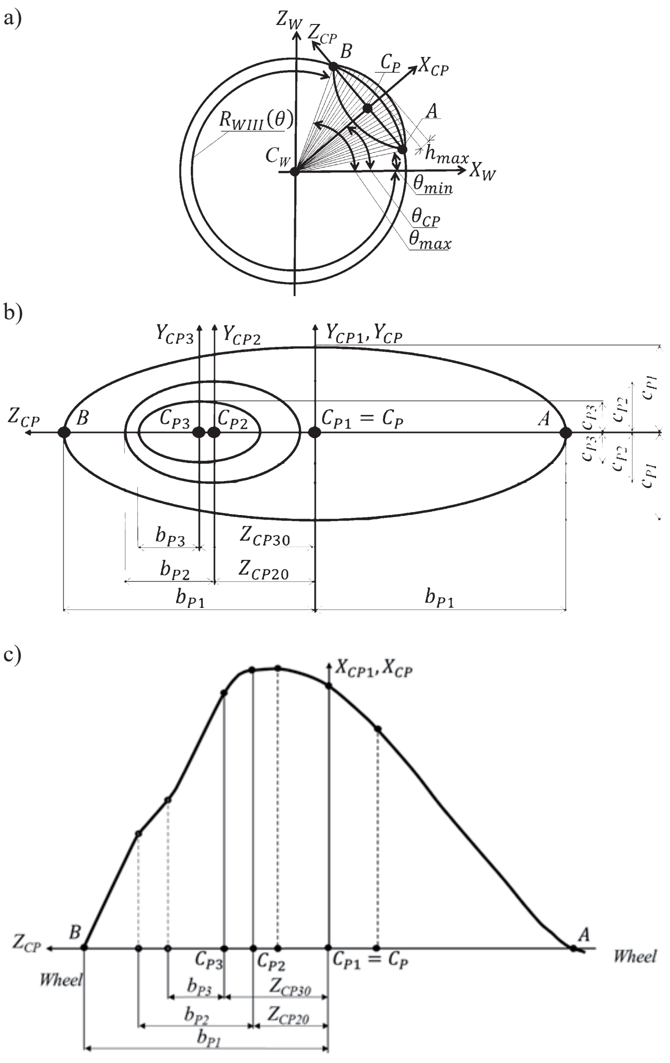


Fig. 3. Computational scheme of metal scale of wheel: a) wheel profile; b) geometrical parameters of metal scale of wheel in directions Y_{CP} and Z_{CP}; c) geometrical parameters of metal scale of wheel in directions X_{CP} and Z_{CP}

The width of the *i*-th layer of wheel metal scale is $2c_{Pi}$, and the length is $2b_{Pi}$. Maximum length of metal scale is $2b_{P1}$. Points A and B (Fig. 3) indicate the start and end of metal scale of vehicle wheels. The geometry of metal scale is described in local coordinate system X_{CP}, Y_{CP}, Z_{CP}. Points C_{P1}, C_{P2}, C_{P3} are the centre points of the metal scale layers. Z_{CP20}, Z_{CP30} (Fig. 3a and Fig. 3b) are the coordinates of centre points Z_{CP} of second and third layers, furthermore $Z_{CP10} = 0$.

The maximum height of metal scale h_{max} (Fig. 3a) is equal to the difference between maximum wheel radius R_{WI} and nominal wheel radius R_W. The size of angles θ_{max} and θ_{min} (Fig. 3a) depends on the position of scale on the surface of the wheel and indicates the maximum and minimum angle size, when generating wheel profile with a metal scale. The position of the scale centre C_P can be described as centre angle $\theta_{CP} = \dfrac{(\theta_{max} + \theta_{min})}{2}$.

When developing the geometrical model of metal scale of vehicle wheel it is accepted that:

- The metal scale of the vehicle wheel consists of N_{layer} layers (Fig. 1).
- The profile of metal scale is generated in local coordinate system X_{CP}, Y_{CP}, Z_{CP}.
- Each centre of metal scale layer can be moved in Z_{CP} axis by a value of Z_{CPi0}, but $Z_{CP10} = 0$.
- Geometrical parameters of each metal scale layer (starting from the second layer) are independent values that are selected in such way, that total profile of metal scales would be generated as accurate as possible.

It is assumed, that metal scale of vehicle wheel is between points A and B, when the centre angle θ (Fig. 3a) varies from θ_{min} to θ_{max}, i.e. $\theta \in [\theta_{min}, \theta_{max}]$. The section between points A and B is divided to $N_P - 1$ intervals, where N_P – total number of points on the surface of generated scale.

During interpretation of the geometry of a metal scale, it is assumed that the number of layers of the calculated metal sheet is $N_{layer} = 3$ (Fig.1, Fig. 3 b, c).

When generating a wheel profile with one scale, the perimeter of the vehicle wheel surface is divided into three: I zone, when $0 \le \theta < \theta_{min}$; II zone, when $\theta_{min} \le \theta < \theta_{max}$; III zone, when $\theta_{max} \le \theta < 2\pi$.

A change of radius in vehicle wheel with metal scale is described as a function $R_{WII}(\theta)$ in the second zone. Changes of radius in the first and the third zones of vehicle wheel with metal scale can be approximated in the same way as in the second zone, when the weariness of vehicle wheel is known in each of the zones.

By knowing the variation of vehicle wheel radius R_W in each zone, it is possible write the function $R_W(\theta)$ of changes in vehicle wheel radius, over the whole perimeter of the wheel surface:

$$R_W(\theta) = R_{WI}(\theta)\left[H(\theta) - H(\theta - \theta_{min})\right] + R_{WII}(\theta) + R_{WIII}(\theta)\left[H(\theta - \theta_{II}) - H(\theta - 2\pi)\right] \tag{2}$$

where $R_{WI}(\theta)$, $R_{WIII}(\theta)$ are known wheel radius functions in the first and third zones, $R_{WII}(\theta)$ is the radius function in the second zone (Fig. 3 a), H is Heaviside step function.

Wheel radius function H $R_{WII}(\theta)$ in the second zone is:

$$R_{WII}(\theta) = \sum_{i=1}^{N_P-1} R_{W,i,i+1}(\theta)\left[H(\theta - \theta_i) - H(\theta - \theta_{i+1})\right], \tag{3}$$

where $R_{W,i,i+1}(\theta)$ is radius function between points i and $i+1$,

$$R_{W,i,i+1}(\theta) = N_1(\xi)R_W(\theta_i) + N_2(\xi)R_W\left(\frac{(\theta_i + \theta_{i+1})}{2}\right) + N_3(\xi)R_W(\theta_{i+1}) \tag{4}$$

where ξ dimensionless coordinate, $N_j(\xi)$ is shape function, $j = 1,2,3$.

Dimensionless coordinate ξ is:

$$\xi = \frac{(\theta - \theta_i)}{\theta_{i+1} - \theta_i}, \text{ when } \xi \in [0,1]. \tag{5}$$

The shape functions $N_j(\xi)$ are equal:

$$N_1(\xi) = (2\xi - 1)(\xi - 1),\ N_2(\xi) = 4\xi(1 - \xi),\ N_3(\xi) = \xi(2\xi - 1). \tag{6}$$

In the coordinate system X_{CP}, Y_{CP} ir Z_{CP} (Fig. 3c) the total scale profile consists of multiple profiles of scale layers. The *k*-th profile of scale layer is described as a function:

$$\left(\frac{Z_{CP}-Z_{CPk0}}{b_{Pk}}\right)^{n_{ZPk}}+\left(\frac{X_{CP}}{a_{Pk}}\right)^{n_{XPk}}+\left(\frac{Y_{CP}-Y_{CPk0}}{c_{Pk}}\right)^{n_{YPk}}=1, \quad (7)$$

where n_{XPk}, n_{YPk}, n_{ZPk} are known exponents, a_{Pk}, b_{Pk}, c_{Pk} are halves of the layer length (Fig. 3 b), Z_{CPk0} and Y_{CPk0} is the *k*-th layer centre coordinate of metal scale.

The height of *k*-th metal scale layer of vehicle wheel is:

$$X_{CPk}=a_{Pk}\left[1-\left(\frac{Z_{CP}-Z_{CPk0}}{b_{Pk}}\right)^{n_{ZPk}}-\left(\frac{Y_{CP}-Y_{CPk0}}{c_{Pk}}\right)^{n_{YPk}}\right]^{n_{XPk}} \quad (8)$$

After adding the entire metal scale layer functions X_{CPk} general scale profile function is generated:

$$f_{P0}(Z_{CP})=\sum_{k=1}^{N_{layer}}a_{Pk}\left[1-\left(\frac{Z_{CP}-Z_{CPk0}}{b_{Pk}}\right)-n_{YPk}\right]^{1/n_{YPk}}\cdot\left[H\left(Y_{CP}+c_{Pk}\right)-H\left(Y_{CP}-c_{Pk}\right)\right]$$
$$\cdot H\left[\left(Z_{CP}+b_{Pk}\right)-H\left(Z_{CP}-b_{Pk}\right)\right]-\left(\sqrt{R_W^2-Z_{CP}^2}-R_{WCP}\right), \quad (9)$$

where R_W is nominal wheel radius, R_{WCP} is wheel radius from wheel centre C_W to point C_P (Fig. 3),

$$R_{WCP}=\sqrt{R_W^2-b_{P1}^2}, \text{ when } -b_{P1}\le Z_{CP}\le b_{P1} \quad (10)$$

Usually, the metal scale on the vehicle wheel has a smooth surface (Fig. 1), therefore, the generated scale function (9) is approximated in the following way:

$$f_P\left(Y_{CP},Z_{CP}\right)=f_{P0}\left(Y_{CP},Z_{CP}\right)\Phi\left(Y_{CP},Z_{CP}\right). \quad (11)$$

where $\Phi\left(Y_{CP},Z_{CP}\right)$ is one of the possible metal scale alignment functions:

$$\Phi\left(Y_{CP},Z_{CP}\right)=\left(cos\left(\frac{\pi Z_{CP}}{2b_{Pk}}\right)\right)^{\frac{1}{nzc}}\left(cos\left(\frac{\pi Y_{CP}}{2c_{Pk}}\right)\right)^{\frac{1}{nyc}}, \quad (12)$$

where *nzc* and *nyc* are exponents.

General function of metal scale profile is described:

$$f_P(Z_{CP})=\Phi(Z_{CP})\sum_{k=1}^{N_{layer}}a_{Pk}\left[1-\left(\frac{Z_{CP}-Z_{CPk0}}{b_{Pk}}\right)^{n_{ZPk}}\right]\left[H\left(Z_{CP}+b_{Pk}\right)-H\left(Z_{CP}-b_{Pk}\right)\right]$$
$$-\sqrt{R_W^2-Z_{CP}^2}-\sqrt{R_W^2-b_{P1}^2}. \quad (13)$$

General metal scale function $f_P\left(Z_{CP}\right)$ Equation (13) is used during examination of the metal scale profile X_{CP} in Z_{CP} direction, in further calculations.

2.2. Mathematical model of system "Vehicle – Track", when the vehicle wheel is scaled

The previously created mathematical model [5, 27] of "Vehicle – Track" (Fig. 4) has been used for most of calculations, which is composed of several mathematical models and designed to determine the forces acting during the interaction between rail and damaged wheel.

This mathematical model assesses the speed of the vehicle, the geometric parameters of the interacting bodies, their physical and mechanical properties, and allows determination of the changes in the forces acting during contact.

The mathematical model of the system "Vehicle – Track" for wheel with a metal scale is used on two-dimensional space. It is assumed that $Y=0$.

Many parameters that appear in the contact zone of wheel and rail, between two contacting surface points and at every time moment can be determined by using this mathematical model [5, 27]: wheel slip, friction forces, frictional torque, distributed load and other parameters. During analysis of the interaction between elements in system "Vehicle – Track", these assumptions and conditions are considered:

- Rail deformation in *X, Z* directions;
- Interaction between roadbed and rail, as an elastic foundation;
- Possible gap between the sleeper and roadbed;
- Length of wheel and rail contact and geometrical unevenness appearing on it;
- The effect of rail axial forces on rigidity (due to differences in temperature);
- Initial bending of the rails;
- Possible gap between rail and sleeper;
- Bending of rail that is between two sleepers;
- Interaction of soil layers, that is under two adjacent sleepers;
- Wheel profile with damages;
- Contact zone is examined as linear contact according to X coordinate.

The system "Vehicle – Track" is examined in vertical direction. The computational scheme and its elements are shown in Figure 4.

Vehicle in computational model of system "Vehicle – Track" consists of (Appendix B and Fig. 4): 1/8 wagon mass m_{bg4}, 1/4 bogie mass m_{bg3}, 1/2 wheelset mass. Wheelset mass is divided into two parts: m_{bg1} - wheel mass, in direct contact with the rail and m_{bg2} - main wheelset mass.

During interaction between the wheel and the rail, the use of the wheel mass m_{bg1} a, which is directly in a contact with the rail, allows a more accurate assessment of the forces occurring on the wheel-rail contact and the kinematic parameters of the individual wheelset.

Track in computational model of system "Vehicle – Track" consists of (Fig. 4): rail (m_r), sleeper (m_{sl}) and railway roadbed m_{si}. Roadbed consists of three layers (Appendix C): ballast (m_{s3}), subballast (m_{s2}) and soil (m_{s1}).

Linear (marked as ovals in Fig. 4 and non-linear (marked as triangles in Fig. 4 stiffness and damping elements are used in the systems "Vehicle – Track".

2.3. Nonlinear dynamical computational algorithm for movement equations of system "Vehicle – Track"

System of movement equations of "Vehicle – Track" with the metal scale of wheelset wheel is equal to:

$$[M]\{\ddot{q}\}+[C]\{\dot{q}\}+[K]\{q\}=\{F_{NL}(q,\dot{q})\}+\{F(t)\} \quad (14)$$

where $[M],[C],[K],\{F_{NL}(q,\dot{q},t)\},\{F(t)\}$ are mass, damping and stiffness matrices, nonlinear generalized force vector and external force vector, respectively. $\{q\}$, $\{\dot{q}\},\{\ddot{q}\}$ are the system generalized displacements, velocities and accelerations vectors, respectively.

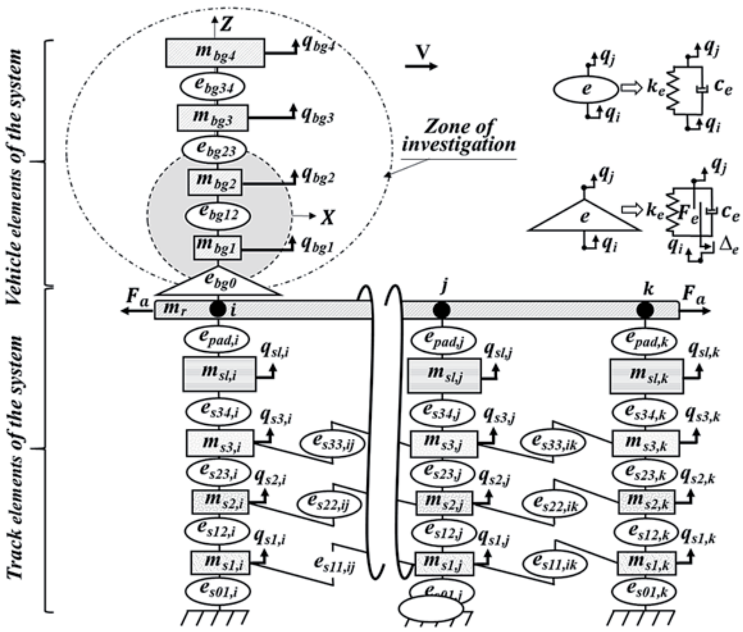


Fig. 4. Computational model of element interaction of system "Vehicle – Track"

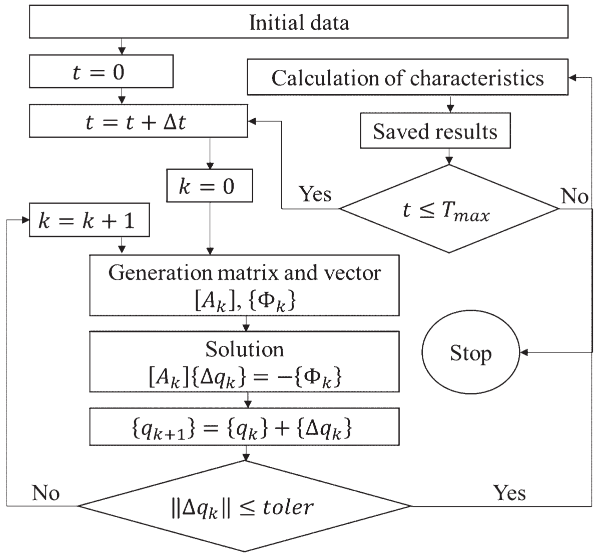


Fig. 5. Dynamical, nonlinear movement equation solving algorithm of system "Vehicle – Track"

Nonlinear generalized force $\{F_{NL}(q,\dot{q},t)\}$ is extracted in the Taylor series at the point $\{q_k\}$:

$$\{F_{NL}(q,\dot{q})\} = \{F_{NL,k}\} + [K_{T,k}]\{\Delta q_k\} + [C_{T,k}]\{\Delta \dot{q}_k\}, \quad (15)$$

where: $[K_{T,k}] = \left[\dfrac{\partial\{F_{NL}(q_k,\dot{q}_k)\}}{\partial\{q\}}\right]$; $[C_{T,k}] = \left[\dfrac{\partial\{F_{NL}(q_k,\dot{q}_k)\}}{\partial\{\dot{q}\}}\right]$;

$\{\Delta q_k\}$ and $\{\Delta \dot{q}_k\}$ *are* increments displacements and velocities vectors, respectively.

Then, the total system of equations (14), at the moment of time $t+\Delta t$, is equal to:

$$[M]\{\ddot{q}_{t+\Delta t}\} + [C]\{\dot{q}_{t+\Delta t}\} + [K]\{q_{t+\Delta t}\} - [C_T]\{\Delta \dot{q}_{t+\Delta t,k}\} - [K_T]\{\Delta q_{t+\Delta t,k}\}$$
$$= \{F_{NL}(q_{t+\Delta t}, \dot{q}_{t+\Delta t})\} + \{F(t)\} \quad (16)$$

where Δt is integration time step; t is time.

By applying Newmark and Newton–Raphson methods, the total system for linear algebraic equations is solved in each of k-th iteration:

$$[A_{t+\Delta t,k}]\{\Delta q_k\} = -\{P_{t+\Delta t,k}\}, \text{ or } \{\Phi_{t+\Delta t,k}\} = [A_{t+\Delta t,k}]\{\Delta q_k\} + \{P_{t+\Delta t,k}\} = 0, \quad (17)$$

where:

$$[A_{t+\Delta t,k}] = \left(\frac{1}{\beta\Delta t^2}[M] + \frac{\gamma}{\beta\Delta t}\left([C]-[C_{T,k}]\right) + \left([K]-[K_{T,k}]\right)\right),$$

$$\{P_{t+\Delta t,k}\} = [M]\{\ddot{q}_{t+\Delta t,k}\} + [C]\{\dot{q}_{t+\Delta t,k}\} + [K]\{q_{t+\Delta t,k}\} - \{F_{NL,k}\} - \{F(t+\Delta t)\},$$

where: β, γ are the Newmark coefficients ($\gamma = 1/2$, $\beta = 1/4$).

The computational algorithm of nonlinear system "Vehicle – Track" is presented in Fig. 5.

An iterative process is performed in each time step, until the accuracy of solution is reached. Furthermore, the rate of displacement vector $\{\Delta q_k\}$ must be lower than the entered accuracy $\{\Delta q_k\} \leq toler$. Parameters of ballast, rail, wheel and the contact between them, speed, accelerations and other parameters must be saved at each time step.

Each of the computational parameters, given in the mathematical model, are averaged depending on the length of the wheel-rail contact. The computational process continues until the condition of $t \leq T_{max}$ is true. Such a solution is suitable for examination of system "Vehicle – Track", when the wheel is scaled, because the created wheelset scale model is three-dimensional.

3. Results and discussion

3.1. Initial data of research of the system "Vehicle – Track", when the wheel has a metal scale

The purpose of the research is to determine the interaction between wheel and rail, show how the wheel and rail movement changes and introduce the impact of geometrical parameters of metal scale on the dynamical loads occurring during the wheel-rail contact, by using the mathematical model of the system "Vehicle – Track", when the wheel has a metal scale. The system "Vehicle – Track" is analysed, when the vehicle wheel has radius R_W = 0.495 m and has a metal scale, is moving on the rail (R-65) at different speeds (V = 40, 60, 80,

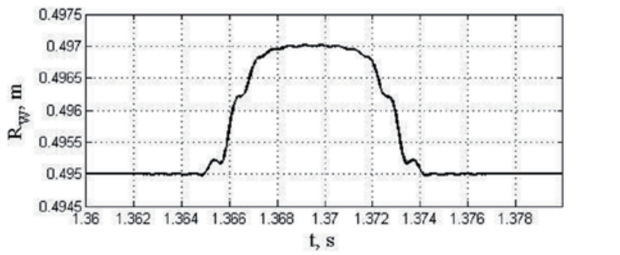


Fig. 6. Alteration of wheel radius, when the wagon is moving at speed $V = 40$ km/h, metal scale width $L_P = 100$ and at different heigths of metal scale h_{max} = 2 mm, at time interval from 1.36 s to 1.38 s

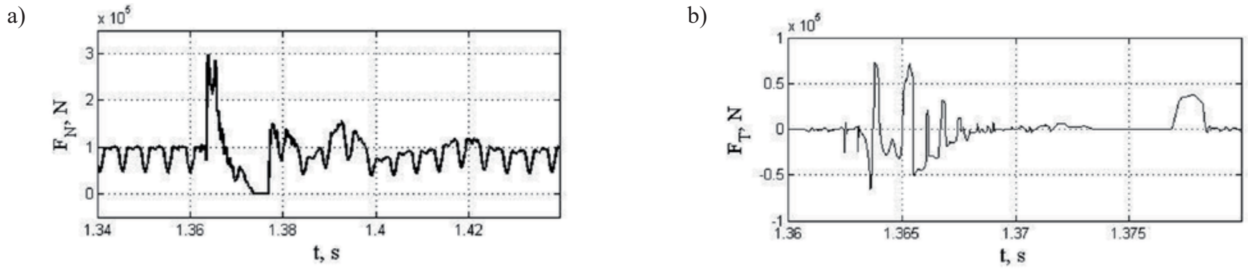


Fig. 7. Dependency of dynamic characteristics on time, when height of vehicle wheel scale is 2 mm and when the vehicle is moving at the speed of 40 km/h: a) normal force F_N at time interval from 1.34 s to 1.44 s; b) friction force F_T at time interval from 1.36 s to 1.38 s

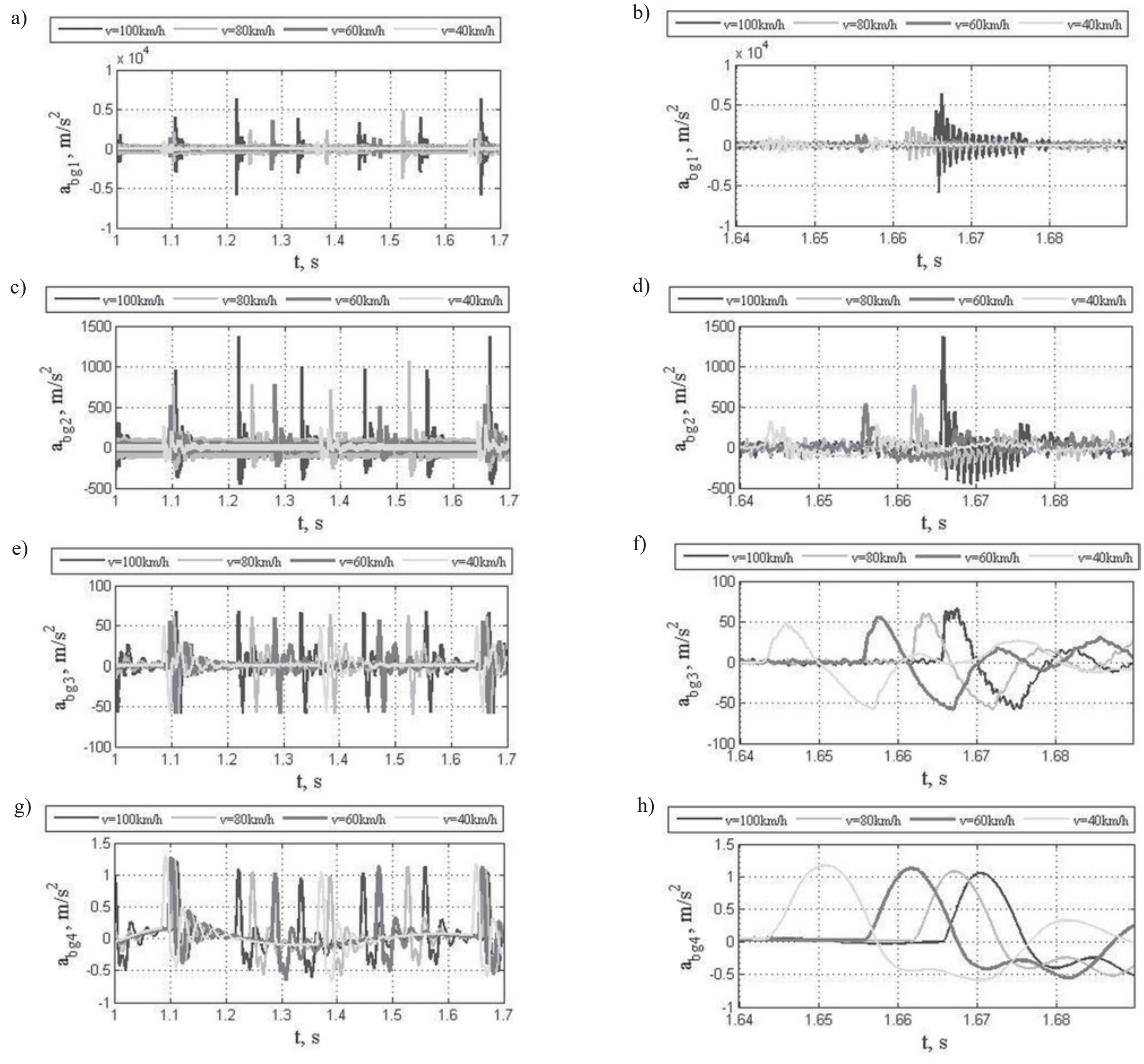


Fig. 8. Dependency of vehicle mass accelerations a_{bgi} on time, when the wheel is moving at different speeds and the wheel has a metal scale, which height is $h_{max} = 2$ mm: a) m_{bg1} in time interval from 1.0 s to 1.7 s; b) m_{bg1} in time interval from 1.635 s to 1.69 s; c) m_{bg2} in time interval from 1.0 s to 1.7 s; d) m_{bg2} in time interval from 1.635 s to 1.69 s; e) m_{bg3} in time interval from 1.0 s to 1.7 s; f) m_{bg3} in time interval from 1.635 s to 1.69 s; g) m_{bg4} in time interval from o 1.0 s to 1.7 s; h) m_{bg4} in time interval from 1.635 s to 1.69 s

100 km/h), when the length of scale is $L = 100$ mm and the maximum height of scale is $h_{max} = 2$ mm.

The data of system "Vehicle – Track", used in the calculations is published in previous author's works [5, 27] and shown in Appendix A. Integration time step is $\Delta t = 5 \cdot 10^{-6}$ s. A profile of a vehicle with scaled wheel is described using Fourier transformation, number of harmonics is $NH = 401$. The calculations assume, that average value of friction coefficient is $\mu = 0.135$, obtained from experiments carried out by the authors [5, 31]. Friction coefficient between the vehicle wheels in relation to the rail must not be lesser than 0.09 - 0.12, otherwise a wheel slip may occur, due to its sticking.

3.2. Results and discussion of research of the system "Vehicle – Track", when the wheel has a metal scale

Dynamical characteristics of the wheel may alter due to damages in vehicle wheel. The developed model [5], allows a detailed analysis of kinematical and dynamical characteristics of system "Vehicle – Track". All parameters of calculations that are shown below are averaged according to the length of the contact.

The dependence of vehicle wheel radius R_W on time and height of metal scale of the wheel, when the movement speed is $V =$ 40 km/h, is shown in Figure 6.

Wheel sliding on the deformed rail causes friction forces in contact zone. Dependency of normal forces F_N, friction forces F_T around the wheel longitudinal axis *Y* and on time, when vehicle wheel has $h_{max} = 2$ mm metal scales height and the vehicle is moving at the speed of 40 km/h, are shown in Figure 7.

Normal force F_N increases in the contact zone of wheel with metal scale and rail, when the metal scale is in the contact zone.

The value of this force depends on the size of the metal scale (Fig. 7 a) and the movement speed. When the movement speed of wheel is $V =$ 40 km/h and the maximum height of metal scale is $h_{max} =$ 2 mm, the maximum normal force F_N is equal to 300 kN.

Therefore, due to the wheel slip that appears in the contact zone (Fig. 8) and the normal force F_N acting in the contact, the friction force F_T must appear.

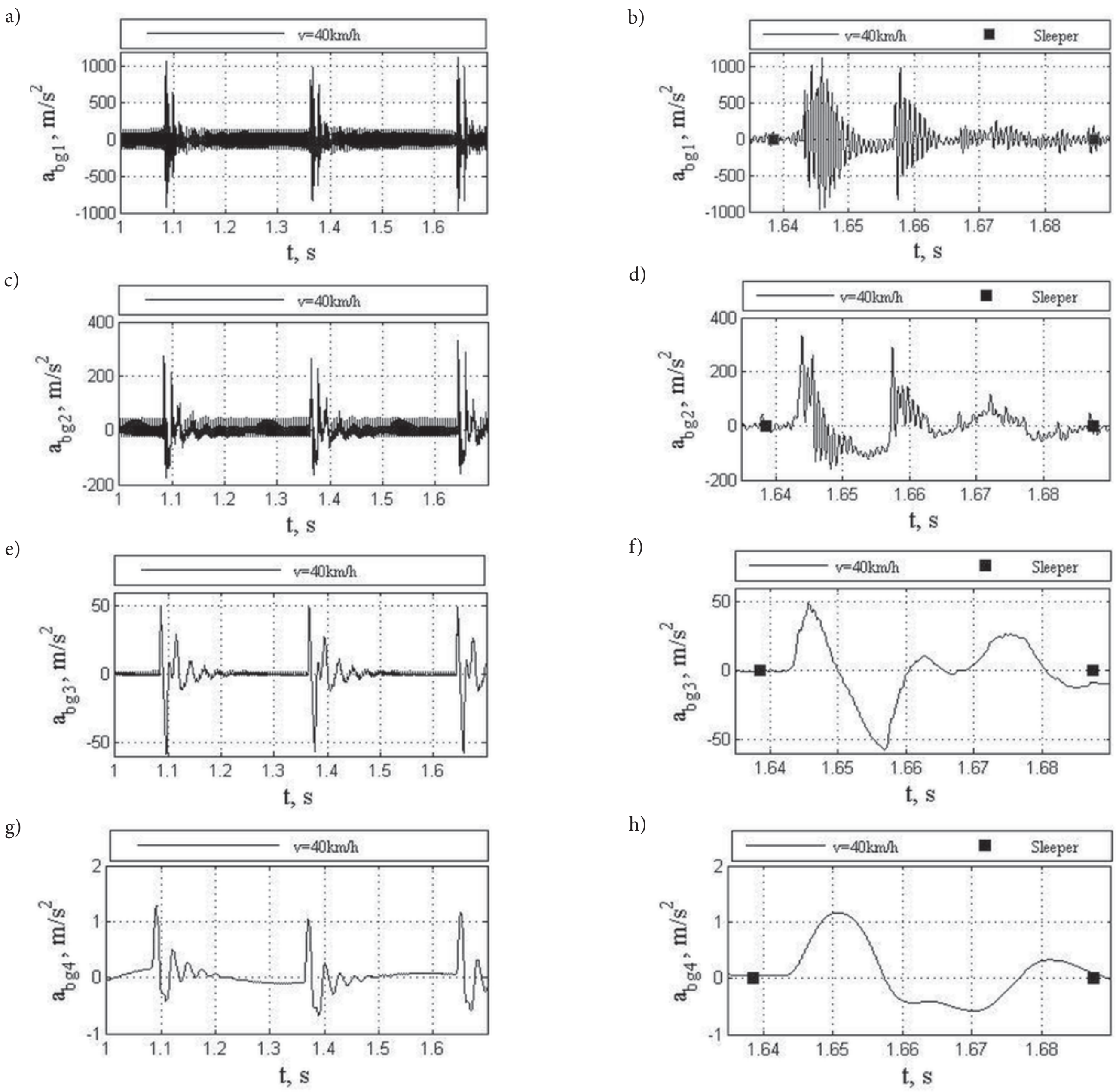


Fig. 9. Vehicle mass acceleration dependency a_{bgi} on time, when the wheel is with a metal scale (when $h_{max} = 2$ mm) moving at speed of 40 km/h: a) m_{bg1} in time interval from 1.0 s to 1.7 s; b) m_{bg1} in time interval from 1.635 s to 1.69 s; c) m_{bg2} in time interval from 1.0 s to 1.7 s; d) m_{bg2} in time interval from 1.635 to 1.69 s; e) m_{bg3} in time interval from 1.0 s to 1.7 s; f) m_{bg3} in time interval from 1.635 s to 1.69 s, g) m_{bg4} in time interval from 1.0 s to 1.7 s; h) m_{bg4} in time interval from 1.635 s to 1.69 s

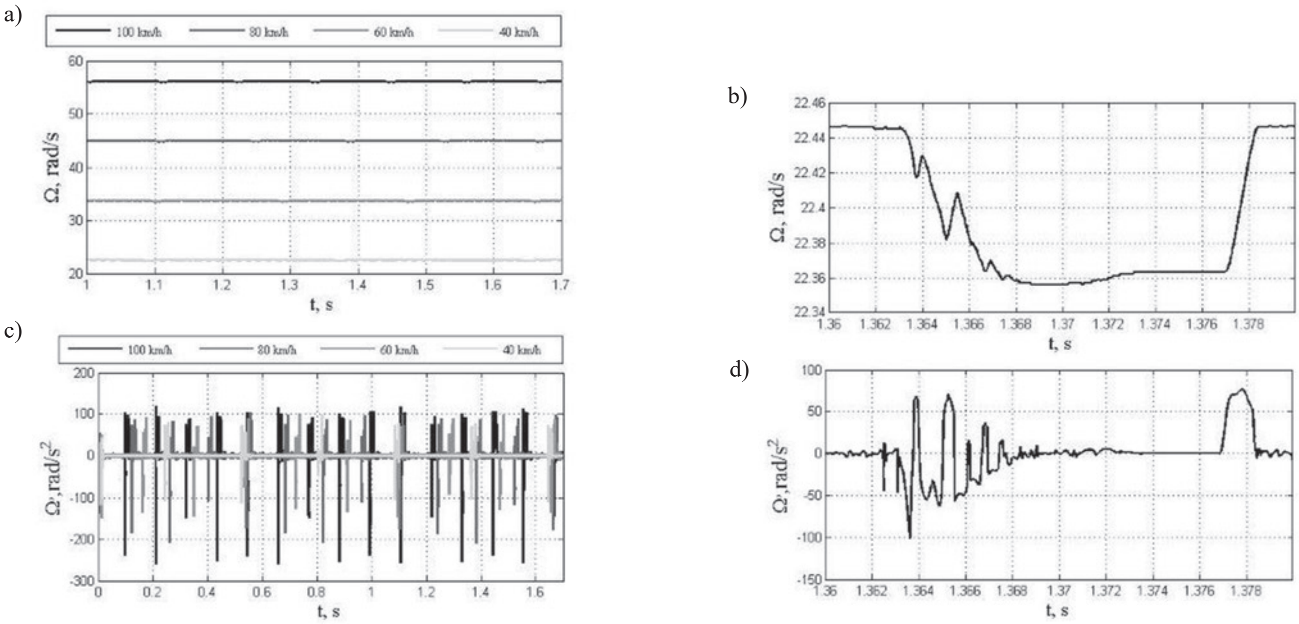


Fig. 10.Dependencies of the angular velocity of wheel Ω *and angular acceleration* $\dot{\Omega}$ *on time, when the vehicle wheel has a metal scales, with heights (*$h_{max} = 2\,\text{mm}$*): a) angular velocity of wheel* Ω*, when vehicle is moving at different speeds (*$V = 40, 60, 80, 100\,km/h$*), at time interval from 1.0 s to 1.7 s; b) angular velocity of wheel* Ω *at time interval from 1.36 s to 1.38 s, when the vehicle moves at speed of 40 km/h; c) wheel angular acceleration* $\dot{\Omega}$*, when vehicle is moving at different speeds (*$V = 40, 60, 80, 100\,km/h$*), at time interval from 1.0 s to 1.7 s; d) wheel angular acceleration* $\dot{\Omega}$ *at time interval from 1.36 s to 1.38 s, when the vehicle moves at speed of 40 km/h*

In Figure 8 it can be seen that acceleration of wheel (mass m_{bg2}) depends on the movement speed of the wheel. When the movement speed V changes (40 km/h, 60 km/h, 80 km/h, 100 km/h), the maximum acceleration of the wheel a_{bg2} alters to 350 m/s^2, 510 m/s^2, 750 m/s^2, 1400 m/s^2 and the minimum acceleration alters to 80 m/s^2, 150 m/s^2, 250 m/s^2, 490 m/s^2, respectively.

Parameters of interaction between wheel with metal scale and rail are dependent on contact zone of the rail. The closer the contact zone is to the sleeper, the bigger the rail stiffness and parameters of interaction are.

The vehicle mass m_{bgi} (when i = 1, 2) acceleration a_{bgi} dependencies on time and contact zone, when vehicle moves at speed $V = 40$ km/h and wheel is with metal scale, which's height is $h_{max} = 2$ mm are shown in Figure 9.

After a comparison of mass m_{bgi} accelerations (when i = 1, 2) a_{bgi} in time $t = 0.08$ s, 1.37 s, 1.66 s, it can be seen that, contact between wheel and rail is closer to sleeper at time $t = 1.66$ s. Wheel mass accelerations acquire highest values at this time. Contact between wheel and rail is most distant from the sleeper at time $t =$ 1.37 s and has the smallest mass acceleration. Dependency of vehicle mass m_{bgi} accelerations a_{bgi} (when i = 1, 2) on time, when vehicle wheel has a metal scale, which height is $h_{max} = 2$ mm and the vehicle movement speed is $V = 40$ km/h.

Variations in time of the angular velocity of wheel Ω and angular acceleration $\dot{\Omega}$, when the vehicle wheel has a metal scale, which height is $h_{max} = 2$ mm, at different moving speeds are shown in Figure 10 a, c and moving at speed $V = 40$ km/h are shown in Figure 10 b, d.

In Figure 10, it can be seen that average angular velocity Ω and acceleration $\dot{\Omega}$ of wheel with a metal scale decreases in the contact zone and its decrease depends on the movement speed.

Loads on sleepers of forces occurring in the contact zone of wheel with metal scale and rails, on sleeper when wheel movement speed is $V =$ 40 km/h and metal scale ($h_{max} = 2$ mm), are shown in Figure 11.

In Figure 11, it is possible to see, that the impact on rail occurs between sleepers 44 and 45, at the time 1.637 s. The impact force F_{pad} on rail loads up in to four sleepers in both directions from the most loaded sleeper (sleeper 45). Maximum force to sleepers, when wheel movement speed is $V = 40$ km/h and the height of metal scale is ($h_{max} = 2$ mm), is equal to $F_{pad} = 130\,kN$.

A more detailed method of interaction between the wheel with metal scale and rail has been developed, which allows determination of the forces and moments of the contact zone of wheel with a metal scale and rail. In most cases, scientists are investigating the problems of interaction between wheel with flat and rail, but in these studies, wheel geometry is simplified, resulting in a wide range of research results. Comprehensive studies of the forces acting on the contact of the wheel with metal scale and the rail contact area are not known to the authors of this article, which suggests that these studies are expanding the knowledge in this field.

The results obtained in this research determine the forces acting in the contact more accurately, as it details the geometry of wheel with metal scale and rail and the forces acting in their contact that arise due to the movement of the rail and wheel in relation to each other. The length of the contact area is divided into many elements (over 1000, element length is less than 0.100 mm), in whose points velocities and accelerations, slip speeds, friction forces and moments are determined at each time moment. A mathematical model of the system "Vehicle – Track" has been developed, which allows determination of the pressure distribution in the contact between wheel and rail, at any time moment.

Wheel's position on rail can be described in position phases. The first is the phase of landing, and the second is the phase of rising (Fig. 12). The first point of phase is on the rail, in the centre of the sleeper and the second one is in the middle of the rail that is located between two sleepers.

Vehicle wheel's impact force is distributed almost symmetrically from the contact zone to the right and left sides (Fig. 11), if the impact occurs between two sleepers, in the centre of the rail, i.e. at the second point of first phase. At the first point of the first phase, the pressure is

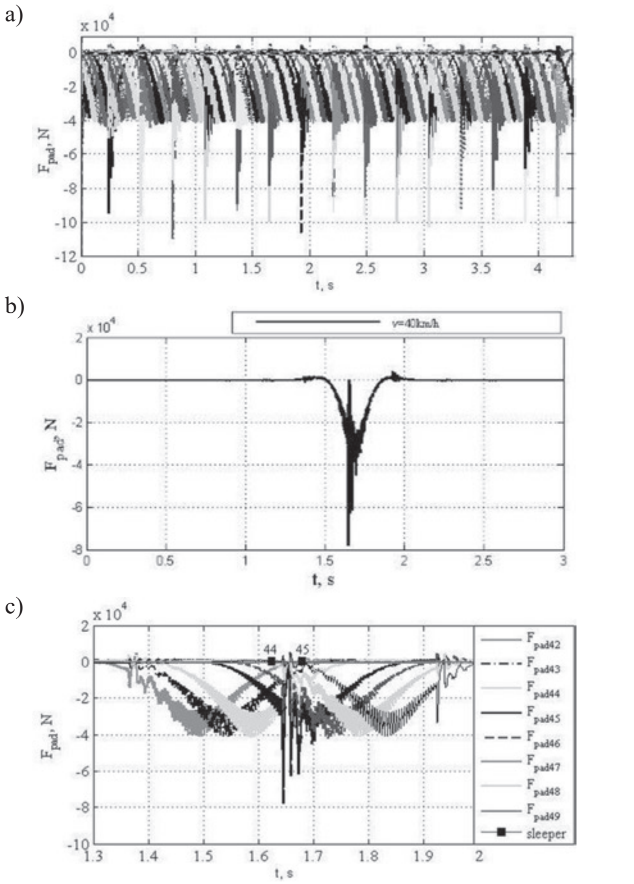


Fig. 11. Loads on sleepers of forces occurring in the contact zone of wheel with metal scale and rails, on sleeper when metal scale is $h_{max} = 2$ mm and wheel movement speed is $V = 40$ km/h

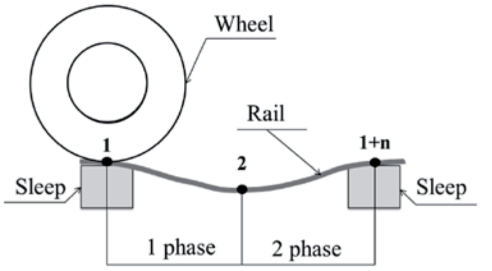


Fig. 12. Position phases of vehicle wheel

the highest. At the second phase, the friction forces increase, the soil and rail become deformed, when the wheel movement resistance is greater and that increases the weariness of wheel and rail. The developed mathematical model of the system „Vehicle – Track“ allows to determine the mass reduction of vehicle wheel with metal scale and rail, i.e. the weariness of profile of wheel with metal scale and rail.

Due to interaction between the wheel with the metal scale and the rail, at their points of contact increased forces and moments that overload the wheelset wheel and the rail appear. The values of these forces depend on the geometrical parameters of wheel, metal scale and rail, the movement speed of the wagon, the physical and mechanical properties of the roadbed, the dynamic characteristics of the wagon bogie, and the load on the wheelset wheel. The appearance of a metal scale on the wheel's surface causes technical and operational problems for the rolling stock. Railway standards limit the speed of movement for a certain size of metal scale.

The results of this work confirm the statement that a wheelset wheel with a metal scale increases the load on the wheelset, the axle box bearings, the rail, and reduces the durability and safe movement of the wheelset wheel and the rail. The developed method and obtained results of the research give a deeper insight into the problems of the wheel with metal scale and rail, as well as the analysis of the forces and moments involved in the contact, depending on the parameters of the system "Vehicle – Track" and the movement speed of the vehicle.

The developed method allows determination of the forces and moments occurring in the contact between wheel with metal scale and rail. It also allows to adjust the loads on wheel bearings, the permissible wagon speed, depending on the geometry of the wheel with metal scale, and determine the heat release in the contact, the speed and size of the wheel and rail wear, the reduction in ride comfort, etc. By using this method, it is possible to create monitoring systems for damaged wheels with metal scales.

The symmetrical profile of vehicle wheel changes during exploitation, due to changes in its radius. This mathematical method allows examining the changes in the wheel profile, if a flat, metal scale, other damages, or if other bodies appear on the surface of the wheel.

The main drawback of this research is that the zone of railway wheel with metal scale and rail is examined as a contact line, but the geometrical model of the metal scale is three-dimensional. The object of the research is the system ‚Vehicle-Track“, so the main focus was on description of contact geometry of rail and wheel with metal scale and on determination of contact forces and moments occurring in the contact zone. Research of interaction between surfaces of wheel and rail are provided later.

The release of heat is studied in the presence of wheel and rail interactions at contact points, the weariness of contacting bodies is studied with and without lubricants (fluid), and also the effect of the rail and wheel wheels on the friction forces and the wearing of the bodies are studied in the research. The increasing speeds of the railway transport show the importance of the problem of contact between the wheel with scale and the rail.

To ensure effective maintenance loads handing over to rail, forces that influence the rail are needed to establish. This method and research results can be used to calculate the interaction between track and other transport vehicles with solid (metal) wheels that have metal scale.

4. Conclusion

The developed spatial mathematical model of wheel with a metal scale (2-13) allows to divide the vehicle wheel surface perimeter into three zones during the generation of scaled wheel profile, also it is used to generate the shape of metal scale and if the geometry of scale is known – to generate the wheel profile with a metal scale.

The mathematical model allows evaluation of the wheel rotation around its longitudinal axis Y, in order to evaluate the rotation of the vehicle wheel and determine its slip on the rails.

It is determined, that when the height of metal scale is $h_{max} = 2$ mm and the length of scale is $L = 100$ mm, the maximum normal contact forces F_N is 300 kN, when the speed of moving wheelset alters $V =$ 40 km/h, static wheel load is 100 kN and average wheel radius is $R_W = 0.495$ m.

When there is a metal scale on the wheel, which has a height of that $h_{max} = 2$ mm and length of $L = 100$ mm, the slipping of wheel on rail increases, maximum slipping speed alters from 0.0437 m/s ($V = 40$ km/h). Wheel slipping on rail causes friction forces in the contact, which increase the weariness of wheel and rail. The friction force in the contact F_T is 75 kN ($h_{max} = 2$ mm, $V = 40$ km/h). Part of mechanical power is converted to heat per unit of time, due to that, the temperature of wheel scale gets higher and the metal scale can be heavily worn.

Rail is loaded with short-term distributed load in the contact zone of wheel with metal scale and rail. The deforming rail transfers this load to the sleepers. It is determined, that the load on the wheel-rail contact zone loads up to four sleepers, when the distance between sleepers is 0.54 m.

The distributed load that acts on the contact zone of wheel with metal scale and rail, loads the wheel, therefore kinematical parameters of wheel, bogie and wagon alter. Because of metal scale that has formed in vehicle wheel, maximum acceleration of wheel is 336 m/s^2 (h_{max} 2 mm, $V =$ 40 km/h), maximum acceleration of bogie is 50 m/s^2 and maximum acceleration of wagon alter is 1.3 m/s^2. The wheelset with scaled wheel not only increases wheelset wheel loads, but also increases loads of bogie, wagon and the wagon's cargo.

The method allows examination of dynamical processes occurring in the system „Vehicle – Track", when the wheel is with a metal scale and the rail profiles are constantly changing.

It has been established, that the developed mathematical models of system "Vehicle – Track" and wheel is with a metal scale allow determination of the resistance forces between rail and wheel and determination of the energy loss, electricity and fuel consumptions. All of this is very important for the operation of the railway rolling stock, rail and road constructions.

Acknowledgement

This work has been supported by the European Social Fund within the project "Development and application of innovative research methods and solutions for traffic structures, vehicles and their flows", project code VP1-3.1-ŠMM-08-K-01-020.

References

1. Bian J, Gu Y, Murray M H. A dynamic wheel–rail impact analysis of railway track under wheel flat by finite element analysis. Vehicle System Dynamics 2013; 1(6): 784-797, https://doi.org/10.1080/00423114.2013.774031.
2. Bogdevicius M, Zygienė R, Dailydka S, Bartulis V, Skrickij V, Pukalskas S. The dynamic behaviour of a wheel flat of a railway vehicle and rail irregularities. Transport 2015; 30(2): 217-232, https://doi.org/10.3846/16484142.2015.1051108.
3. Bogdevicius M, Zygiene R, Skrickij V, Methodology for the determination of maximum contact vertical wheel loads. Transbaltica 2015: Proceedings of the 9th International Scientific Conference, May 7–8, 2015.Vilnius Gediminas Technical University, 2016; 134: 348-352, https://doi.org/10.1016/j.proeng.2016.01.018.
4. Bogdevicius M, Zygiene R. Simulation dynamic processes of rail vehicle and rail with irregularities. Journal KONES. 2014; 21(2): 48-51, https://doi.org/10.5604/12314005.1133858.
5. Bogdevicius M., Zygiene R., Bureika G., Dailydka S. An analytical mathematical method for calculation of the dynamic wheel–rail impact force caused by wheel flat. Vehicle System Dynamics 2016; 54(5): 689-705, https://doi.org/10.1080/00423114.2016.1153114.
6. Bureika G, Bekintis G, Liudvinavičius L, Vaičiūnas G. Applying analytic hierarchy process to assess traffic safety risk of railway infrastructure. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2013; 15(4): 376-383.
7. Chudzikiewicz A, Bogacz R St, Kostrzewski M, Konowrocki R. Condition Monitoring of Railway Track Systems by Using Acceleration Signals on Wheelset Axle-Boxes. Transport 2018; 33(2): 30-42, https://doi.org/10.3846/16484142.2017.1342101.
8. Domin R, Domin I, Cherniak G. Estimation of dynamic performances of the safe operation of highspeed electric train. Archives of Transport 2017; 41(1): 7-16, https://doi.org/10.5604/01.3001.0009.7374.
9. Ferrara R, Leonardi G, Jourdan F. Numerical Modelling of Train Induced Vibrations. Procedia– Social and Behavioral Sciences (Sciencedirect) 2012; 53: 155-165. https://doi.org/10.1016/j.sbspro.2012.09.869
10. Jazykov Vl. Priminenie modeli negercevskogo kontakta kolesa s relsom dlia ocenki dinamicheskix kachestv gruzovogo teplovoza. Dissertacija na soiskanije stepeni kandidata technicheskih nauk. Briansk: BGTU, 2004. (in Russian).
11. Keropyan A M. Determination of Rational Values of the Coefficients of Friction and Roughness of the Surfaces of Rails and Ties Locomotives. Nedelia gorniaka, 2014.
12. Keršys R, Bazaras Ž, Griškevičius P. Vagono vertikalios dinamikos modeliavimas. Transport engineering 1999; 5: 40-42.
13. Konop J, Konowrocki R. On Evaluation of the Wheelsets-Track Interaction Quality in Railway Engineering. Machine Dynamics Research 2013; 37(4): 61-70.
14. Kostrzewski M. Analysis of selected vibroacoustic signals recorded on EMU vehicle running on chosen routes under supervised operating conditions JVE International LTD. Vibroengineering Procedia. Sep. 2017; 13: 2345-0533, https://doi.org/10.21595/vp.2017.18958.
15. Kouroussis G, Gazetas G, Anastasopoulos I, Conti C, Verlinden O. Discrete modelling of vertical track–soil coupling for vehicle–track dynamics. Soil Dynamics and Earthquake Engineering 2011; 31(12): 1711-1723, https://doi.org/10.1016/j.soildyn.2011.07.007.
16. Kumar V, Rastogi V. Investigation of vertical dynamic behaviour and modelling of a typical Indian rail road vehicle through bond graph. World Journal of Modelling and Simulation 2009; 5(2): 130-138.
17. Nielsen J C O. High-frequency vertical wheel-rail contact forces – Validation of a prediction model by field testing. Wear 2008; 265: 1465-1471, https://doi.org/10.1016/j.wear.2008.02.038.
18. Nielsen J, Igeland A. Vertical dynamic interaction between train and track influence of wheel and track imperfections. Journal of Sound and Vibration 1995; 187(5): 825-839, https://doi.org/10.1006/jsvi.1995.0566.
19. Pieringer A, Kropp W, Nielsen J C O. The influence of contact modelling on simulated wheel/rail interaction due to wheel flats. Wear 2014; 314(1-2): 273-281, https://doi.org/10.1016/j.wear.2013.12.005.
20. Sackfield A, Dini D, Hills D A. Contact of a rotating wheel with a flat. International Journal of Solids and Structures 2007; 44(10): 3304-3316, https://doi.org/10.1016/j.ijsolstr.2006.09.025.
21. Sackfield A, Dini D, Hills D A. The contact problem for a wheel having a 'flat'. Wear 2006; 261(11-12): 1265-1270, https://doi.org/10.1016/j.wear.2006.03.009.
22. Shabana A A, Rathod C. Geometric coupling in the wheel/rail contact formulations: a comparative study. Journal of Multi–body Dynamics 2007; 221(2): 147-160, https://doi.org/10.1243/14644193JMBD103.
23. Shabana A A, El–Ghandour A I, Zaazaa K E. Study of the effect of the spiral geometry on wheel/rail contact forces. Journal of Multi–body Dynamics 2011; 225(2): 111-125, https://doi.org/10.1177/1464419311406626.

24. Sivilevičius H, Maskeliūnaitė L. The numerical example for evaluating the criteria describing the quality of the trip by international. E&M Economics and Management 2014; 2: 73-86, https://doi.org/10.15240/tul/001/2014-2-006.
25. Sladkowski A, Sitarz M. Analysis of wheel–rail interaction using FE software. Wear 2005; 258(7–8), 1217-1223, https://doi.org/10.1016/j.wear.2004.03.032.
26. TNN. Techninio gelezinkelių naudojimo nuostatai. Lietuvos Respublikos susisiekimo ministro 1996 m. rugsėjo 20 d. įsakymas Nr. 297.
27. Turskis Z, Zavadskas E K. A new fuzzy additive ratio assessment method (ARAS-F). Case study: The analysis of fuzzy multiple criteria in order to select the logistic centers location. Transport 2010; 25: 423-432, https://doi.org/10.3846/transport.2010.52.
28. Vasauskas V, Bazaras Ž, Čapas V. Strength anisotropy of railway wheels under contact load. Mechanika 2005; 1(51): 31-38.
29. Wang K, Liu P F, Zhai W, Huang Ch, Chen Z, Gao J. Wheel/rail dynamic interaction due to excitation of rail corrugation in high–speed railway. Special Topic: High–speed Railway Infrastructure 2015; 58(2): 226-235, https://doi.org/10.1007/s11431-014-5633-y.
30. Wasiwitono U, Zheng D, Chiu W K. How useful is track acceleration for monitoring impact loads generated by wheel defects? In Proc. of the 5th Australasian Congress on Applied Mechanics, ACAM 2007, 10–12 December 2007, Brisbane, Australia, 2007: 502-507.
31. Zygienė R, Bogdevicius M, Dabulevicienė L. A mathematical model and simulation results of the dynamic system railway vehicle wheel–track with a wheel flat. Science – The future of Lithuania: Transport engineering, 2014; 6(5): 95-101, https://doi.org/10.3846/mla.2014.696.

Appendix A

Table A1. Data calculations of the system "Vehicle-Track"

Definition	Notation	Definition	Notation
Masses of ballast:	$m_{s1} = 800$ kg	Rail mass per meter	$m_R = 65$ kg/m
	$m_{s2} = 465$ kg	Static load	$F_x = 100$ kN
	$m_{s3} = 200$ kg	1/8 car body mass	$m_{bg4} = 8743$ kg
Damping coefficients of ballast:	$c_{s01} = 90$ kNs/m	1/4 bogie mass	$m_{bg3} = 700$ kg
	$c_{s12} = 70$ kNs/m	1/2 wheel set mass	$m_{bg2} = 640$ kg
	$c_{s23} = 60$ kNs/m	Mass in contact	$m_{bg1} = 110$ kg
	$c_{s34} = 50$ kNs/m	Stiffness coefficient of the car body	$k_{bg34} = 2.55$ MN/m
	$c_{s11,i,j} = 10$ kNs/m	Stiffness coefficient of the bogie	$k_{bg23} = 6.5$ MN/m
	$c_{s22,i,j} = 13$ kNs/m	Stiffness coefficient of the wheel set	$k_{bg12} = 5$ GN/m
	$c_{s33,i,j} = 15$ kNs/m	Damping coefficient of the car body	$c_{bg4} = 10$ kNs/m
	$k_{s01} = 180$ MN/m	Damping coefficient of the bogie	$c_{bg3} = 100$ kNs/m
	$k_{s12} = 170$ MN/m	Damping coefficient of the wheel set	$c_{bg2} = 50$ kNs/m
Stiffness coefficients of ballast:	$k_{s23} = 160$ MN/m	Damping coefficient of mass in contact	$c_{bg1} = 44.2$ kNs/m
	$k_{s34} = 150$ MN/m	Wheel radius	$R_W = 0.495$ m
	$k_{s11,i,j} = 15$ MN/m	Elastic modulus of the wheel	$E_W = 210$ GPa
	$k_{s22,i,j} = 16$ MN/m	Exponent	n = 3 / 2
	$k_{s33,i,j} = 17$ MN/m	Maximal penetration velocity	$\dot{\delta}_{max} = 10$ m/s
Spacing between the sleepers centres	$L_{sl} = 0.5435$ m	Mass inertia moment of wheelset	$I_{WY} = 65$ kNm
Mass of the sleeper	$m_{sl} = 140$ kg	Restitution coefficient	$e = 0.65$
Width of a *railway* sleeper	$b_{sl} = 0.15$ m	Poisson's coefficient of the wheel	$\upsilon_W = 0.30$
Height of a *railway* sleeper	$h_{sl} = 0.12$ m	Friction parameters:	$\mu_{X0} = 0.14$
Number of finite elements between two sleepers	10		$\mu_X = 0.11$
Pad damping coefficient	$c_{pad} = 45$ kNs/m		$\gamma_v = -2.50\,s/m$
Pad stiffness coefficient	$k_{pad} = 140$ MN/m		$k_s = 800\,s/m$
The Second moment of the area of the rail about Z axis	$J_{RZ} = 5.69 \cdot 10^{-6}$ m4	Contact length of wheel with rail	$L_{contact} = 100$ mm
The Second moment of the area of the rail about Y axis	$J_{RY} = 3.54 \cdot 10^{-5}$ m4	Length of metal scale	$2b_{P1} = L_P = 0.1$m
Elastic modulus of the rail	$E_R = 206$ GPa	Maximum heights of metal scale	$h_{max1} = 0.001$ m

Poisson's coefficient of the rail	$\upsilon_R = 0.30$	$h_{max2} = 0{,}003$ m
Cross-sectional area of the rail	$A_R = 82.9 \cdot 10^{-4}$ m2	$h_{max3} = 0{,}005$ m
Rail density	$\rho_R = 7850$ kg/m3	

Marijonas BOGDEVICIUS
Department of Mobile Machinery and Railway Transport
Vilnius Gediminas Technical University
Plytines g. 27, 10105 Vilnius, Lithuania

Rasa ZYGIENE
Department of study program of Electronics and Electrical Engineering
Kaunas University of Applied Engineering Sciences,
Tvirtoves al. 35, 50155 Kaunas, Lithuania

E-mails: marijonas.bogdevicius@vgtu.lt, rasazz@gmail.com

Article citation info:
JASIULEWICZ-KACZMAREK M, ŻYWICA P. The concept of maintenance sustainability performance assessment by integrating balanced scorecard with non-additive fuzzy integral. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 650–661, http://dx.doi.org/10.17531/ein.2018.4.16.

Małgorzata JASIULEWICZ-KACZMAREK
Patryk ŻYWICA

THE CONCEPT OF MAINTENANCE SUSTAINABILITY PERFORMANCE ASSESSMENT BY INTEGRATING BALANCED SCORECARD WITH NON-ADDITIVE FUZZY INTEGRAL

KONCEPCJA OCENY ZRÓWNOWAŻONEGO UTRZYMANIA RUCHU Z ZASTOSOWANIEM ZRÓWNOWAŻONEJ KARTY WYNIKÓW I NIE-ADDYTYWNEJ CAŁKI ROZMYTEJ

In response to the growing sustainability concerns, manufacturing companies have to formulate measures to assess sustainable manufacturing performance, aiming at integration of sustainability aspects. Although various models and methods to assess the sustainability of production processes, and point the role of maintenance have been developed in recent years, contribution of all the elements of the maintenance to the results of sustainable production has not been comprehensively considered, since mostly financial aspects were analyzed. Taking into account this research gap, the article presents the concept of a model and procedure for assessing maintenance from the perspective of sustainable manufacturing requirements. Authors integrate three sustainability dimensions (economic, social and environmental) with Kaplan and Norton's balance scorecard perspectives as a basis to develop the model of maintenance sustainability performance assessment. For the model developed, the assessment procedure based on the paradigm of aggregate assessment was designed. The Choquet integral, based on the so-called λ – measure, was implemented to aggregate the measures. Then, the results of research on determining the importance and interactions between the perspectives and criteria for assessing sustainable maintenance in enterprises representing the automotive and food industries are presented.

***Keywords**: maintenance performance**,** sustainable maintenance, balanced scorecard, fuzzy integral.*

W odpowiedzi na wyzwania zrównoważonego rozwoju (SD), przedsiębiorstwa produkcyjne włączają ekonomiczne, środowiskowe i społeczne wymagania SD do swoich praktyk produkcyjnych i formułują miary do oceny skuteczności podjętych działań. Pomimo, iż w ostatnich latach opracowano wiele modeli i metod oceny zrównoważonej produkcji i wskazywano w nich na rolę utrzymania ruchu, to jednak poza aspektem finansowym nie rozważano w sposób kompleksowy wszystkich elementów wkładu utrzymania ruchu w wyniki zrównoważonej produkcji. Biorąc pod uwagę tę lukę badawczą, w artykule przedstawiono koncepcję modelu i procedury oceny utrzymania ruchu z perspektywy wymagań zrównoważonej produkcji. Autorzy integrują trzy wymiary zrównoważonego rozwoju (ekonomiczny, społeczny i środowiskowy) z perspektywami zrównoważonej karty wyników Kaplana i Nortona, jako podstawę do skonstruowania modelu oceny wyników zrównoważonego utrzymania ruchu. Dla tak opracowanego modelu zaprojektowano opartą na paradygmacie oceny agregatowej procedurę oceniania. Do agregacji składników oceny zastosowano całkę Choqueta, opartą na tzw. mierze λ. Następnie przedstawiono wyniki badań pilotażowych dotyczących określenia ważności i interakcji między perspektywami i kryteriami oceny zrównoważonego utrzymania ruchu w przedsiębiorstwach branży motoryzacyjnej i spożywczej.

***Słowa kluczowe**: wyniki utrzymania ruchu, zrównoważone utrzymanie ruchu, zrównoważona karta wyników, całka rozmyta.*

1. Introduction

About thirty years ago, sustainable development was defined as a 'development that meets the needs of the present without compromising the ability of future generations to meet their own needs'. One of the most important issues regarding sustainable development is sustainable manufacturing. Manufacturing operations are accompanied by various environmental and social concerns at different stages of the production processes [46]. Consequently, in the last few years, research is focusing on this new paradigm, which aims to develop sustainable production processes, innovative technologies, and new tools for assessing economic, environmental, and social impacts of industrial assets. In this context, according to [15, 32, 42, 44], maintenance function, necessary to ensure availability, reliability, and safety of industrial assets, could become one of the main pillars for sustainable manufacturing.

In response to the growing sustainability concerns, manufacturing companies have to formulate measures to assess sustainable manufacturing performance, aiming at integration of sustainability aspects. Although various models and methods to assess the sustainability of production processes and point the role of maintenance have been developed in recent years, contribution of all the elements of the maintenance to the results of sustainable manufacturing has not been comprehensively considered, since mostly financial aspects were analyzed. Taking into account this research gap, the article presents the concept of a model and procedure for assessing maintenance from the perspective of sustainable manufacturing requirements. The goal of maintenance sustainability assessment is to provide decision makers with information about current maintenance results and support them in the decision-making process regarding future directions of maintenance activities.

The authors integrate three sustainability dimensions (economic, social and environmental) with four Kaplan and Norton's balance scorecard perspectives (financial, client, internal processes and learning & growth) as a basis to develop the model of maintenance sustainability performance assessment. The assessment model developed has a two-tier hierarchical structure, the first level of which encompasses the assessment perspectives, while the second level includes the assessment criteria. The assessment procedure was developed for the model constructed as explained. The goal is to develop a synthetic indicator of sustainable maintenance performance. The assessment procedure is based on the paradigm of aggregate assessment, which stresses determination of the synthetic efficiency of an organization and company's operation on the basis of merging the particular assessment criteria into the one. The main steps of the procedure are: determining importance of the assessment criteria, defining method of assessing for each of them, and then aggregating them. Since the assessment process should also take into account a number of interrelated environmental, social and economic issues, the determination of importance and aggregation of criteria are the critical steps [14].

Hence, the second goal of the article is to select the appropriate aggregate function. In general, the properties of an aggregate function can be related to mathematical properties and behavioral properties. Mathematical properties point to formally correct aggregation of criteria, while behavioral properties express relationships between criteria including, for example, synergy and redundancy. Most of the aggregation methods presented in the literature in relation to the assessment of the sustainable development of the system have some drawbacks, namely they do not reflect the interaction between the criteria [14]. The solution to this problem in the model of sustainability assessment is to use fuzzy integrals for aggregation. Due to its characteristics, the fuzzy integral has been widely applied to the multiple attribute decision-making [8, 53]. In our application, we consider a particular case of Choquet integral, based on the so-called λ – measure.

The article is divided into three parts. The first one includes: the role of maintenance function in sustainable manufacturing environment (chapter 2); maintenance sustainability assessment framework and the fuzzy integrals with fuzzy measure as a tool for aggregation of multiple criteria (chapter 3). The second part presents the results of pilot studies on the perception of the importance and interactions between the perspectives and criteria for assessing maintenance sustainability in enterprises representing the automotive and food industries (chapter 4). The third part is the conclusions and presentation of directions for further research (chapter 5).

2. Sustainable maintenance

From a pragmatic point of view, the main goal of maintenance is to optimize the overall lifecycle of an object. In other words, ensuring maximum availability and reliability of production equipment, at minimum cost and in accordance with binding legal requirements (concerning environment, occupational safety, etc.). This goal is indicated in various maintenance definitions, e.g. [10].

Over the years, along with the production process, maintenance has evolved from the reactive function, through preventive, lean (Lean Maintenance), green (Green Maintenance), to the modern approach in which it is considered a process that should be managed in a sustainable perspective [25]. In the literature, the role of maintenance in the implementation of sustainable manufacturing goals is emphasized many times, e.g. [6, 24]. The contribution of maintenance to the economic dimension of sustainable manufacturing concerns the reduction of both maintenance costs (e.g. costs of spare parts and consumables, labor costs, etc.) and total operating costs (e.g. environmental fees, technological media costs) [5, 28, 47, 49, 55], whereas concerning the environmental dimension of the maintenance operation, i.e. planning service activities, monitoring the condition of machines and devices, it enables reduction of consumption of technological media, raw materials and materials, as well as their more efficient use [1, 3, 27, 29, 30, 52]. The social dimension concerns the relationship between the maintenance function and its stakeholders both within the company and outside of it, with particular emphasis on the employees of the technical department, production, suppliers of spare parts, service providers, etc., competence and satisfaction of employees, work environment [4, 48], satisfaction. Moreover, by preventing emergency events and limiting their consequences, maintenance is related to the safety of processes and people [43].

Because the maintenance processes in an enterprise depend on the context of the enterprise, its objectives, structure, internal constraints and external conditions, it is not possible to provide a standard formula for achieving sustainable development in terms of maintenance, given its complexity in various systems. Nevertheless, it is possible to clarify some of the common aspects of sustainable development in different industries and to indicate the general procedure to be followed when maintaining maintenance to the next level of sustainable maintenance. This procedure can be interpreted as an improvement cycle plan-do-check-act (P-D-C-A). In this way, the transition to sustainable maintenance is defined as a set of activities and processes that turn maintenance into sustainable one through the continuous process that consists of: (1) an assessment of current sustainability maintenance performance; (2) an identification of improvement areas; (3) a suggestion of specific actions across the company; and (4) an implementation of these actions. Therefore, the first step should be to assess the current maintenance performance from the perspective of the requirements of sustainable manufacturing.

Over the last twenty years, many measures and indicators, models and frameworks for measuring and monitoring the maintenance systems performance have been developed [23, 36]. Among them, the most popular assessment tool is the scorecard, in particular the balanced scorecard (BSC) by Kaplan and Norton [26]. BSC is a holistic approach that transfers the company's strategy (or strategy for any function in the enterprise) to clearly predefined goals related to four different perspectives, namely: financial, clients, internal processes, and learning and development. This diversity of perspectives is important for the measurement of performance because intangible assets, such as customer relations, employee skills, knowledge and innovations, are currently treated as the main source of competitive advantage for enterprises. Balanced scorecard clearly transfers focus from current financial results to building or strengthening internal potential and investing in people, systems, procedures and internal processes in order to improve future achievements. Kaplan and Norton also pointed out that the four perspectives presented in the overall concept are just an example and may change depending on the purpose of the assessment. Hence, it is possible to add perspectives and change their names. This approach was used to design models for assessment of maintenance performance [2, 9, 11, 12, 31, 34, 41, 47, 54]. The literature review of models of maintenance assessment based on the Kaplan and Norton's scorecard shows that: (1) every author has a unique way to classify maintenance assessment perspectives. The different categories of perspectives show different areas of interest in maintenance performance. They also differ in their choice of criteria describing each perspective; (2) some categories of perspectives and criteria are recognized by all authors as vital for management of the maintenance function, for example, much emphasis has been placed on maintenance cost-related measures, skills and competencies of maintenance workers, safety and environment, maintenance work management, and customer satisfaction.

However, in any of the literature models, according to the authors (the model developed by [47] is an exception), criteria important from sustainable manufacturing perspective, such as waste requirements, energy consumption, water, consumables, work environment etc., were not taken into account comprehensively. Although the model developed by [47] should be considered pioneering in the assessment of maintenance sustainability, further research is needed, in particular with regard to how to assess and use the assessment results for improvement purposes.

3. The framework of maintenance sustainability performance assessment

3.1. Assumptions, goal and scope of assessment

Considering [35, 40] it was assumed that 'maintenance as a subsystem of the manufacturing system is sustainable if it contributes to the sustainability of the large system (of which it is a part) while maintaining its own sustainability'. This definition results in an important implication for the design of the maintenance assessment model from the perspective of sustainable manufacturing. The model should simultaneously include two elements, firstly the assessment of results in relation to the superior system represented by, for example, stakeholders and their requirements, secondly, the maintenance potential. Moreover, because the recipients of the result of the assessment are decision-makers, the use of too many indicators and measures could lead to the 'dilution' of the information they contain. Therefore, it is necessary to design a composite indicator to assess sustainable maintenance. Composite indicators are used in many areas of research, including maintenance e.g. [13, 20].

Considering the above, it was assumed that the concept of maintenance sustainability assessment should fulfil the following criteria: (1) it should integrate the three factors of sustainability – economic, environmental and social, (2) it should take into account the links between maintenance and its stakeholders, (3) it should be based on financial and non-financial measures, (4) it should be easy to interpret, as the composite maintenance performance is a model in the mathematical sense. The first three requirements refer to the criteria being assessed, while the fourth requirement concerns the method of presenting the result of the assessment.

One of the methods of developing performance measurement models most frequently mentioned in the literature from the perspective of sustainable development, which: (1) combines the strategy with the objectives and measures of their implementation, (2) includes and links financial and non-financial measures, (3) takes into account the links between internal effectiveness of processes and their external efficiency, and in addition (4) enables inclusion of dimensions of sustainable development, is a scorecard developed by Kaplan and Norton. This method is also used by many authors to develop models for assessment of maintenance performance (see chapter 2). However, it has not yet been applied to a comprehensive assessment of maintenance performance from the perspective of sustainable manufacturing requirements. Considering the above, the general framework of the developed model for measuring maintenance performance from the perspective of sustainable manufacturing was embedded in the general assumptions of the balanced scorecard.

The goal of the of maintenance sustainability assessment is to provide information on current maintenance performance and support decision-makers in the decision-making process regarding future directions of operations. This information should be synthetic, and thus show the result of the assessment in an aggregated way, and at the same time enable decomposition to lower levels showing the impact of each of the assessed criteria to the final result. However, the BSC method does not include any techniques for quantifying the contribution of each perspective, or criteria/indicators within the same perspective. The developed concept of assessment of maintenance from the perspective of sustainable manufacturing solves this problem based on the paradigm of aggregate assessment. The general scheme of the methodology for aggregate maintenance assessment includes three main stages: (1) Assessment criteria selection, (2) Selection of methods of criteria assessment, and (3) Development of Composite Maintenance Sustainability Index.

The model of maintenance performance assessment from sustainable manufacturing perspective developed according to the three stages scheme should help the maintenance managers transfer the strategy into action and offer predictive measures for future performance. In particular, it should answer the four important questions: (1) What is the impact of the perspectives and criteria on the sustainable maintenance performance? (2) How can importance of these perspectives and criteria be determined? (3) How can maintenance performance be measured? (4) What is the actual level of maintenance performance from sustainable manufacturing point of view?

3.2. Identification of assessment criteria

Initial four classic perspectives of the BSC model suggested by [26] were modified (e.g. as suggested by [39] the customer perspective was replaced by the stakeholders' perspective), criteria for the assessment of each perspective and the corresponding specific issues were defined, thus extending the traditional BSC model to the hierarchical structure of HBSC (Fig. 1).

To ensure relevance and credibility of the developed model of measurement of maintenance performance and the corresponding criteria, two types of information sources were used. First of all, on the basis of the literature analysis, the assessment criteria and detailed issues for each of the criteria were distinguished. Secondly, experts were consulted (researchers and practitioners from enterprises). Experts representing the scientific community assessed completeness and indicated potential gaps in the model. On the other hand, experts representing business practice, apart from completeness, assessed the scope, usefulness and the possibility of conducting an assessment in real business conditions.

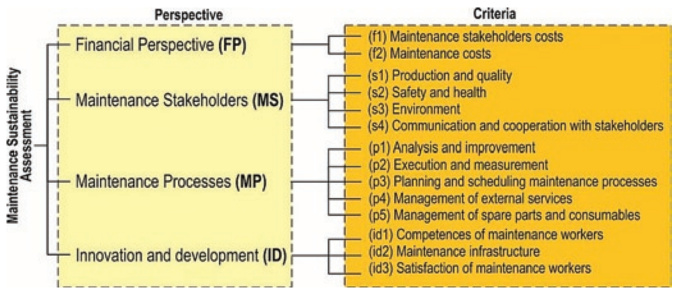


Fig. 1. Hierarchical model for maintenance sustainability assessment

The first of the perspectives in the assessment model is the 'Financial Perspective' (FP). It reflects the maintenance function that can be defined as ensuring human and equipment safety, respect for the environment and accessibility at the lowest possible cost. The assessment of this perspective includes two criteria: (f1) costs of maintenance stakeholders and (f2) maintenance costs. The second perspective is the perspective of 'Maintenance Stakeholders' (MS). Maintenance stakeholders are various organizational units of the company that receive services, in particular the production department and its employees. The key aspects of this perspective include, among others: satisfaction with services rendered, response time, availability of technical equipment, information obtained from technical support services, cooperation between departments and the safety of employees and the environment. Under this perspective, four criteria will be assessed: (s1) production and quality, (s2) safety and health, (s3) environment, (s4)

Table 1. Criteria of sustainable maintenance assessment

Perspective	Criterion	Description
Financial	f1	costs related to production downtime, quality discrepancies associated with failures / unstable machinery operation, unplanned costs of purchase of spare parts and sub-contractors services, costs of non-compliance with environmental legal requirements, environmental costs due to malfunctions (e.g. waste treatment, penalties), costs of disposal of wastes generated during service work, overtime costs, costs of non-compliance with work safety requirements, costs related to injuries and accidents of maintenance workers, operators and third parties that occurred during maintenance works, as a result of failures, etc.,
	f2	principles of determining the budget of the maintenance department, costs of labor, training, purchase and maintenance of inventory of spare parts and consumables, cost of sub-contractors, cost of media consumed by maintenance department (electricity, compressed air, water)
Maintenance stakeholders'	s1	availability and reliability of machines, response to the request for service, quality of training provided by maintenance department for production employees, availability and quality of procedures and instructions for operators, machines capability, product non-compliance due to unstable machine operation and failures
	s2	injuries and accidents while performing maintenance work by operators, sub-contractors and abandonment or non-performance of such work resulting from them, system for identification of non-compliance with OHS principles, improvement actions undertaken by maintenance staff to eliminate health and safety hazards, limiting criticality of accidents/ failures
	s3	environmental incidents during works performed by operators, sub-contractors and resulting from the abandonment or non-performance of such works, waste monitoring system related to them, product waste caused by unstable machinery and failures, improvement actions taken by maintenance staff to eliminate environmental hazards related to machinery failures, limiting criticality of accidents/ failures
	s4	communication channels used (e.g. mobile technologies, use of CMMS system), formal system of meetings, incompatibilities caused by lack of communication or untimely transfer of information, work of multidisciplinary teams to solve problems and improve
Maintenance processes'	p1	analysis of: technological limits of media consumption by machines (e.g. water, electricity), emergency events and their causes, repair time, delays in task execution, analysis of consumption of lubricants, hazardous substances used during servicing, formal improvement system and its effectiveness, modernization of machines and devices and their effectiveness, actions aimed at extending the life cycle of lubricants, and limiting the consumption of hazardous substances
	p2	the level of execution of the plan and schedule of maintenance, the method and scope of recording and documenting maintenance work (record standards), inspection system: e.g. noise, leakage or emission level, waste generated during maintenance works, their toxicity, waste segregation procedures, environmental incidents occurring during maintenance work, injuries and accidents concerning maintenance staff during service performance, compliance with applicable OHS procedures
	p3	availability of data and information on operational events, maintenance performed and their effectiveness, availability of information from diagnostic tests, formal methods and criteria for identifying critical equipment (including safety and environmental issues in machine criticality assessment), criteria for selection of maintenance strategies for individual machines and devices, procedures and work instructions that take into account OSH and environmental hazards
	p4	methods and criteria for selection of sub-contractors, defining the scope of services, cooperation rules and procedures (including, for example, safety rules, waste management procedures), how to document the activities completed, monitoring the execution
	p5	methods and criteria implemented for selection and assessment of suppliers of spare parts and consumables, demand planning, inventory monitoring, use of environmentally friendly consumables, storage methods (safety and environmental requirements), use of regenerated/remanufactured parts
Innovation and development	id1	planning of training for employees, new employee introducing programs, methods for verifying knowledge and skills of employees, training topics (e.g. improvement methods and techniques, new technologies, environmental management and OSH issues)
	id2	adequacy of quality and quantity of equipment in relation to the scope of performed maintenance works, including diagnostic tests, planning and implementation of investments in this area, delays in service work resulting from lack of availability or quality of equipment, the scope of IT tools for planning, supervising, monitoring and analysis work of technical facilities and maintenance processes,
	id3	working hours, working environment, rotation of maintenance staff, motivation systems

communication and cooperation with stakeholders. Another perspective (third) is 'Maintenance Processes' (MP) which concerns all aspects related to maintenance operations, including planning and scheduling, monitoring and control as well as improvement. The overall objective is to ensure efficiency and effectiveness of operations, through proper organization of maintenance work, improvement of spare parts and consumables management, improvement of service reliability by own technical services and external units (service providers), improvement of safety of services provided, etc. Within this perspective, five criteria are to be assessed: (p1) analysis and improvement, (p2) execution and measurement, (p3) planning and scheduling maintenance processes, (p4) management of external services, (p5) management of spare parts and consumables. The fourth perspective, 'Innovation and development' (ID), indicates that achieving efficiency from the financial, stakeholders and internal processes perspectives depends to a large extent on the competencies of employees, resources that they can have during the implementation of tasks and the level of job satisfaction. Within this perspective, three areas are assessed: (id1) competences of maintenance workers, (id2) maintenance infrastructure, (id3) satisfaction of maintenance workers. The detailed scopes of each criterion of assessment are presented in the table 1.

The scope of the assessment of individual criteria presented in Table 1 in authors' opinion covers all the key maintenance issues from the perspective of sustainable manufacturing. Each of the issues describing a specific criterion will be assessed with a point scale by a team of experts from the company, and the final assessment of the criterion will be calculated as the ratio of the number of points obtained to total number of points available. Due to the scope of the article, a detailed description of the method of criteria assessment has not been presented. The presented model (Fig. 1 and table 1) is designed for the company's internal purposes. It can be implemented as a self-assessment tool and at the same time a 'road map' of activities that should be considered when striving for improvement. The model does not impose or suggest ready-made solutions but presents many approaches to achieving success in meeting the challenges of sustainable manufacturing.

3.3. Structure of the composite index

One of the assumptions of the assessment of maintenance sustainability is the scheme of presentation of the result of the assessment. This result is to be presented in the form of an interpretation-friendly composite index, i.e. it is a function in the mathematical sense and at the same time it needs to present the contribution of each of the perspectives and assessment criteria to the final result of the assessment. Based on the literature review [8, 16, 18, 53] to solve the problem of aggregating perspectives and criteria that are interdependent, a non-additive fuzzy integral was selected. Fuzzy integral method applies fuzzy measures to deal with the problems of human subjective perception and uncertainty as well as to address the level of interdependency effects among criteria [51]. Fuzzy measure is non-additive hence the total importance of individual features may not be equal to the combined importance of the features. In this research, we are motivated to implement the theory of fuzzy measures to model the importance and interaction between features in the Choquet integral

The principle and construction process of the non-additive fuzzy integral for composite maintenance sustainability index (CMSI) can be described in three steps: (1) Determine the importance of decision criteria with linguistic variables; (2) Determine the decision factors importance using fuzzy measure; (3) Calculate the aggregate value (CMSI).

Step 1. *Determine the importance of decision criteria with linguistic variables_*

The first step in developing a maintenance sustainability index focuses on weighting the individual elements (perspectives and criteria). Weighting is to assign importance for each element based on their relative importance. It is a very sensitive process which can lead to different results due to different importance assigned. Therefore, it affects the accuracy of the assessment. The determination of the importance of each of the perspectives and assessment criteria is carried out by a team of experts. Because, experts have different levels of cognitive vagueness (different experience and knowledge), linguistic variables are used to determine the importance degree of maintenance assessment perspectives and criteria. Those importance weights are aggregated by applying the fuzzy arithmetic. During construction of CMSI FN-OWA (Fuzzy Number Ordered Weighted Average) [7] aggregation operator was applied. The main reason why FN-OWA was selected is that it has ability to aggregate not only the quantitative data, but it also can handle linguistic as well as crisp data. Moreover, it is idempotent operator which means that operator retains the same linguistic state if all input criteria have equal values [45]. The aggregated fuzzy importance weights of decision factors need to be converted into crisp numbers for subsequent utilization in fuzzy measure construction.

Step 2. *Determine the decision factors weight using fuzzy measure*

One important issue in sustainable assessment is the need to express not only the importance of individual features but also interactions between them. There are normally three kinds of interactions between two criteria A and B (μ denotes importance/weight)

a) Synergetic interaction, which can be represented by $\mu(A \cup B) > \mu(A) + \mu(B)$,
b) Inhibitory interaction, which can be represented by $\mu(A \cup B) < \mu(A) + \mu(B)$,
c) Non-interaction, which can be represented by $\mu(A \cup B) = \mu(A) + \mu(B)$.

Classical probability theory can only be applied to the third situation when there is no interaction between two criteria, while fuzzy measure theory can describe any of the three situations [16]. The fuzzy measure is a measure for representing the membership degree (importance) of an object (a criterion) in candidate sets (set of criteria). Fuzzy measures can be considered as the generalization of the probability measure. In this case, the additive property is replaced with a weaker requirement. Mathematically, a fuzzy measure is defined as follows [17]:

Definition 1

A discrete fuzzy measure on X is a set function $\mu : 2^X \rightarrow [0,1]$ satisfying:

1. $\mu(\emptyset)=0$, $\mu(X)=1$,
2. $A \subset B$ implies that $\mu(A) \leq \mu(B)$ (monotonicity).

However, constructing suitable fuzzy measure is not trivial; because the number of coefficients increases exponentially as the number of features grows (in general such measure requires $2^{|X|}$ values to be defined). Moreover, fuzzy measures also need to meet the monotonicity and continuity requirements. To address this problem Sugeno [50] proposed the λ-measure (also called Sugeno measure, or λ-addtive measure). To restrict number of required coefficients Sugeno added additional axiom to fuzzy measure definition:

Definition 2 (Discrete Sugeno λ-measure):

A discrete fuzzy measure is called Sugeno λ-measure if it satisfies:

3. If $A \cap B = \emptyset$, then $\mu_\lambda(A \cup B) = \mu_\lambda(A) + \mu_\lambda(B) + \lambda\mu_\lambda(A)\mu_\lambda(B)$.

Sugeno [50] proved that given those 3 axioms, fuzzy measure can be uniquely determined using only $n = |X|$ coefficients μ_i that are often called fuzzy densities that represent the degree of importance of criteria i-th and can be calculated by parametric or nonparametric methods. The λ-measure can be calculated using following formula:

$$\mu_\lambda(A) = \frac{1}{\lambda}\left[\prod_{x_i \in A}(1+\lambda\mu_i) - 1\right] \quad (1)$$

where $\lambda > -1$ is solely determined by the equation:

$$\lambda + 1 = \prod_{i=1}^{|X|}(\lambda\mu_i + 1) \quad (2)$$

λ-measure is constrained by a parameter λ, which describes the degree of additivity the attributes hold. According to [21, 22]:

- If $\lambda < 0$, then it implies that the attributes are sharing sub-additive (redundancy) effect.
- If $\lambda > 0$, then it interprets that the attributes are sharing super-additive (synergy) effect.
- If $\lambda = 0$, then it indicates that the attributes are non-interactive.

It is important to note that in fuzzy measure importance of the single criterion or a pair of criteria it not solely determined by $\mu\left(\{x_i\}\right)$ or $\mu\left(\left\{x_i, x_j\right\}\right)$. One need to consider all $\mu(A)$ such that $x_i \in A$ or $\left\{x_i, x_j\right\} \subseteq A$. Murofushi et al. [37, 38] proposed solution to this problem based on game theory for single criterion and utility theory for pairs for criteria. Based on fuzzy measure, the importance index (Shapley value) and interaction indices of different perspectives and criteria were defined.

Definition 3 (Shapley importance index, Shapley value) [37]:

Let μ be a fuzzy measure on X. The Shapley value (or the importance index) for every element $x_i \in X$ is defined by following formula:

$$\Lambda(x_i) = \sum_{A \subset X - \{x_i\}} \gamma_X(A)\left[\mu\left(A \cup \{x_i\}\right) - \mu(A)\right], \tag{3}$$

where:

$$\gamma_X(A) = \frac{\left(|X| - |A| - 1\right)!|A|!}{|X|!}. \tag{4}$$

The Shapley value with respect to the measure μ is a vector $v = \left[\Lambda(x_1), \Lambda(x_2), \ldots, \Lambda(x_n)\right]$. It describes the global importance of every element by taking into account the effects of all subsets with and without the given element. According to the definition, the Shapley value has the property that the sum of all its components is '1', which can be formulated as $\sum_{i=1}^{n} \Lambda(x_i) = 1$. Scaled by a factor n, Shapley values greater than '1' indicate that the given element (criterion) is more important than the average.

Another important topic is the notion of interaction among two criteria, as proposed by [38].

Definition 4 (Interaction Index) [38]:

Let μ be a fuzzy measure on X. Interaction index of criteria x_i and x_j is defined by:

$$I_\mu\left(x_i, x_j\right) = \sum_{A \subset X - \left\{x_j, x_j\right\}} \xi_X(A) I_\mu\left(x_i, x_j \mid A\right), \tag{5}$$

where:

$$\xi_{\mathrm{X}} = \frac{\left(|\mathrm{X}| - |A| - 2\right)!\,|A|!}{\left(|\mathrm{X}| - 1\right)!}, \tag{6}$$

and

$$I_\mu\left(x_i, x_j | A\right) = \mu\left(A \cup \left\{x_i, x_j\right\}\right) - \mu\left(A \cup \{x_i\}\right) - \mu\left(A \cup \left\{x_j\right\}\right) + \mu(A). \tag{7}$$

Interaction index takes values from $[-1,1]$ interval, where negative (positive) values indicate negative (positive, synergic) interaction.

Based on the concept of Shapley importance index and interaction index, the contribution of one or more elements to a whole fuzzy measure can be described as follows:

a) Element x_i is said to be more important than x_j if the Shapley value $\Lambda(x_i) > \Lambda\left(x_j\right)$;

b) Elements x_i and x_j are redundant if the interaction index $I_\mu\left(x_i, x_j\right) < 0$;

c) Elements x_i and x_j are complementary if the interaction index $I_\mu\left(x_i, x_j\right) > 0$

When fuzzy measure is constructed the next step is to apply it in Choquet integral to obtain value of CMSI.

Step 3. *Calculate the aggregate value (CMSI).*

The fuzzy integral is used in the sustainable maintenance assessment to combine assessments primarily because this model does not need to assume independence among the criteria. A formal definition of the Choquet integral is as follows:

Definition 5 (Discrete Choquet integral):

Let μ be a fuzzy measure on X. Choquet integral of function $f : X \to [0,1]$ with respect to fuzzy measure μ is defined by:

$$C_\mu\left(f_1, f_2, \cdots, f_n\right) = \sum_{i=1}^{n} \left(f_{(i)} - f_{(i-1)}\right) \mu\left(A_{(i)}\right), \tag{8}$$

where $f_{(i)}$ indicates that the indices have been permutated so that $0 \le f_{(1)} \le \cdots \le f_{(n)} \le 1$, $A_{(i)} = \left\{x_{(i)}, \cdots, x_{(n)}\right\}$ and $f_i = f(x_i)$.

Main properties of Choquet integral important from perspective of multi-criteria decision making, are [16, 19]: idempotence, continuity, monotonicity (non-decreasingness) with respect to each argument, stability under the same positive linear transformation. Because of stability of Choquet integral, the exact numerical scale in relation to which the calculations are made is not important. This significantly simplifies the collection of data from experts, enabling the assessment on a linguistic scale without the need to establish specific interpretations. However, the interest in the fuzzy integral is mainly due to its ability to represent the interactions between criteria. This is due to the fact that weights in fuzzy measure are assigned to every subset of all criteria. The above properties show that fuzzy integrals have the potential for sustainability maintenance assessment.

Calculation procedure of Composite Maintenance Sustainability Index (CMSI) requires fuzzy measure (μ) and actual values of criteria obtained from company assessment team (h). The outline of this procedure is given on Fig. 2.

CMSI value measures organization's maintenance sustainability and lie in the range between 0 and 1. Value of CMSI closer to '0' indicates that the maintenance is unsustainable; whereas a value closer to '1', means that the maintenance structure is sustainable and contributes to the sustainability of the manufacturing system while maintaining its own sustainability.

4. Empirical studies of the importance of perspectives and criteria assessment for assessing maintenance sustainability performance

The determination of the importance of both perspectives and individual criteria was carried out by an expert method according to the scheme: (1) developing linguistic scale; (2) appointing experts; (3) assessment of importance of perspectives and criteria by experts; (4) calculating the average value for perspectives and criteria; (5) developing λ measure; (6) determining the importance and interactions between perspectives and criteria. In the assessment model, linguistic variables represent a subjective assessment of the importance of the perspectives, as well as their criteria. A five-level linguistic scale was used (Table 2). Only descriptive definitions of extreme elements of the scale were presented, i.e.: irrelevant - the perspective/criterion is

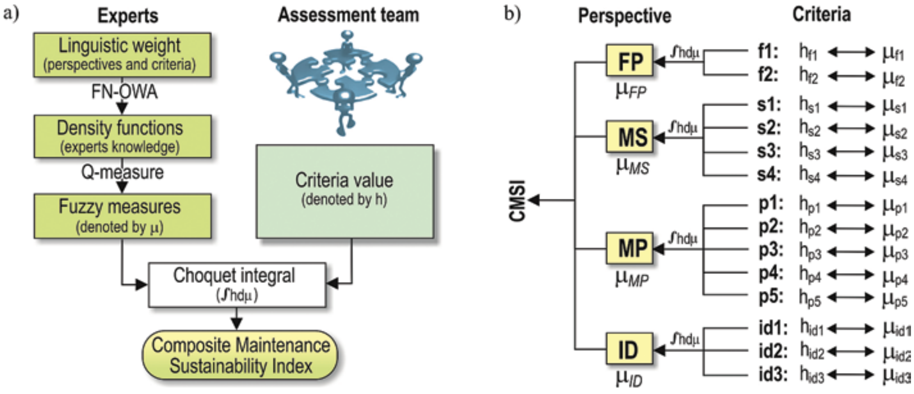
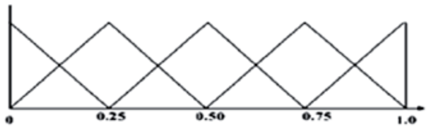


Fig. 2. Calculation procedure of Composite Maintenance Sustainability Index (CMSI). Outline of CMSI construction process. (b) Scheme of actual hierarchical calculation of CMSI value

Table 2. Linguistic scales

Linguistic terms	Linguistic values	Triangular fuzzy membership functions
Very important	(0.75, 1.0, 1.0)	
Important	(0.5, 0.75, 1.0)	
Moderately important	(0.25, 0.5, 0.75)	
Equal important	(0, 0.25, 0.5)	
Irrelevant	(0, 0, 0.25)	

almost irrelevant to the level assessment; very important - the perspective/criterion can be used alone to assess the entire level.

Based on the model developed (fig. 1) and the adopted scale of linguistic assessments, a form was designed for the experts to determine the importance of perspectives and criteria. The questionnaire contained the definition and meaning of each of the perspectives. After the questionnaire was designed, it was distributed among 3 experts from enterprises to verify the wording, its intelligibility and unambiguity.

In order to determine the importance of perspectives and assessment criteria, empirical studies were carried out in production companies. For research purposes, companies that have sustainable development or corporate social responsibility policies defined and have implemented quality management system (e.g. ISO 9001, ISO 22000, BRC, IFS, ISO / TS 16949), are implementing or currently operating in accordance with the Lean Manufacturing concept and perform maintenance operations by themselves were selected. Eight companies from the automotive industry (A) and seven food industry enterprises (F) were invited to the research. The research was carried out among the heads of maintenance departments and heads of production separately in each of the companies.

The assessment of importance of perspectives and criteria was carried out individually in each of the companies during meetings with the head of the production department and the head of the maintenance department, at which the model of sustainable maintenance assessment was described at the outset, and then the principles for assessing individual perspectives and criteria were presented. Experts (heads of Maintenance department (M) and heads of Production department (P)) were requested to assess the importance of perspectives and criteria depending on the hierarchical system using the five linguistic variables. The experts were not limited with the imposed numerical interpretation of the applied linguistic variables.

The average importance for each perspective was calculated using FN-OWA operator with weight vector that discards extreme values. Input values were arranged with respect to natural order of linguistic terms (irrelevant < less important < moderately important < important < very important). The same method was used to aggregate expert assessments for criteria inside perspectives (second level criteria). The last step was to calculate the single numerical values for each of the fuzzy numbers. These values can be treated as an average assessment of the importance of individual perspectives/criteria. Application of Center of Gravity defuzzification method led to the results presented in Table 3 and 4.

Calculated values of μ_i (Table 3 and 4) were used to develop fuzzy measure. Algorithm presented in [33] was implemented in R 3.4.4 Statistical Computing Platform and applied without fixing λ to averaged importance values to construct Sugeno λ-measure (Definition 2). The main benefit of this procedure is its ability to automatically rescale input values so that obtained measure optimally reproduces expert data. Obtained measures as well as λ-values for automotive industry and food industry are presented in Table 5.

Table 3. Fuzzy densities μ_i obtained by experts for perspective

Perspective	Automotive industry (A)			Food industry (F)		
	MA	PA	MA &PA	MF	PF	MF &PF
FP	0.2917	0.3333	0.3173	0.2500	0.2500	0.2500
MS	0.3542	0.3750	0.3558	0.3750	0.3750	0.3750
MP	0.3333	0.3125	0.3173	0.3500	0.3500	0.3438
ID	0.2292	0.2083	0.2115	0.2500	0.1750	0.2187

Table 4. Fuzzy densities μ_i obtained by experts for criteria

Perspective	Criteria	Automotive industry (A)			Food industry (F)		
		MA	PA	MA &PA	MF	PF	MF &PF
FP	f1	0.3608	0.4165	0.3972	0.3998	0.3998	0.4026
	f2	0.3887	0.2495	0.3330	0.3330	0.2328	0.29132
MS	s1	0.2833	0.3000	0.2923	0.3000	0.3000	0.3000
	s2	0.3000	0.2833	0.2923	0.2200	0.3000	0.2583
	s3	0.2167	0.2000	0.2154	0.2600	0.2000	0.2333
	s4	0.1833	0.1333	0.1615	0.1800	0.1400	0.1583
MP	p1	0.1800	0.1662	0.1726	0.1660	0.1494	0.1591
	p2	0.25000	0.2500	0.2500	0.2500	0.2500	0.2500
	p3	0.2220	0.2080	0.2112	0.1996	0.1828	0.1940
	p4	0.1940	0.1802	0.1855	0.2332	0.2332	0.2290
	p5	0.1660	0.1245	0.1468	0.1660	0.1162	0.1453
ID	id1	0.3333	0.3542	0.3365	0.3000	0.3000	0.3021
	id2	0.3333	0.2708	0.3077	0.3500	0.3000	0.3229
	id3	0.2708	0.2500	0.2692	0.2250	0.1250	0.1771

Table 5. Fuzzy measure λ value for criteria

Perspective	Criteria	Automotive industry (A)			Food industry (F)		
		λ_{MA}	λ_{PA}	$\lambda_{MA\,\&PA}$	λ_{MF}	λ_{PF}	$\lambda_{MF\,\&\,PF}$
FP	f1, f2	1,7862	3,2141	2,0394	2,0070	3,9474	2,6112
MS	s1, s2, s3, s4	0,0462	0,2622	0,1110	0,1150	0,1814	0,1468
MP	p1,p2,p3,p4,p5	-0,0296	0,2014	0,0896	-0,0365	0,1938	0,0593
ID	id1, d2, id3	0,2097	0,4743	0,3032	0,4762	1,5120	0,8855

Table 6. Interaction index for criteria

Perspective	Criteria	Automotive industry (A)			Food industry (F)		
		MA	PA	MA &PA	MF	PF	MF &PF
FP	f1, f2	0.2505	0.3340	0.2698	0.2672	0.3674	0.3062
MS	s1, s2	0.0040	0.0233	0.0097	0.0078	0.0168	0.0117
	s1, s3	0.0029	0.0166	0.0072	0.0092	0.0113	0.0106
	s1, s4	0.0030	0.0157	0.0072	0.0068	0.0113	0.0091
	s2, s3	0.0024	0.0112	0.0054	0.0064	0.0080	0.0072
	s2, s4	0.0026	0.0106	0.0054	0.0047	0.0080	0.0062
	s3, s4	0.0019	0.0075	0.0040	0.0055	0.0054	0.0056
MP	p1, p2,	-0.0013	0.0088	0.0040	-0.0015	0.0076	0.0024
	p1,p3	-0.0012	0.0074	0.0034	-0.0012	0.0056	0.0019
	p1,p4	-0.0016	0.0110	0.0048	-0.0018	0.0093	0.0029
	p1,p5	-0.0010	0.0064	0.0030	-0.0014	0.0071	0.0022
	p2,p3	-0.0014	0.0095	0.0043	-0.0021	0.0118	0.0034
	p2,p4	-0.0013	0.0080	0.0036	-0.0017	0.0087	0.0027
	p2,p5	-0.0009	0.0044	0.0023	-0.0010	0.0036	0.0014
	p3,p4	-0.0012	0.0066	0.0034	-0.0015	0.0059	0.0022
	p3,p5	-0.0011	0.0055	0.0029	-0.0012	0.0044	0.0017
	p4,p5	-0.0009	0.0048	0.0025	-0.0014	0.0056	0.0020
ID	id1, id2	0.0240	0.0482	0.0327	0.0527	0.1488	0.0931
	id1, id3	0.0196	0.0447	0.0287	0.0348	0.0695	0.0541
	id2, id3	0.0196	0.0348	0.0264	0.0402	0.0695	0.0574

Table 7. Shapley value for perspective

Perspectives	Automotive industry			Food industry		
	MA	PA	MA &PA	MF	PF	MF &PF
FP	0.9616	1.0880	1.0576	0.8040	0.8608	0.8320
MS	1.1836	1.2360	1.1952	1.2404	1.3184	1.2792
MP	1.1088	1.0152	1.0576	1.1512	1.2256	1.1652
ID	0.7460	0.6612	0.6900	0.8040	0.5952	0.7236

Once the fuzzy measures for perspectives and criteria are identified, the next step is to compute the interaction index using formula 5, 6 and 7 (Table 6.).

The interaction index allows to identify criteria that are synergistic (positive value of the indicator) or redundant (negative value of the indicator). According to the assessment of the managers of the maintenance departments of both industries (MA and MF assessment), the criteria describing the perspective of 'Maintenance processes' (MP) are redundant, which means that some criteria should be rejected (see also λ_{MA} and λ_{MF} in Table 5), however, according to other experts both the criteria describing this perspective and the criteria describing the remaining perspectives are synergistic. Nevertheless, since the values of interaction ratios are close to zero, it is difficult to draw binding conclusions. Therefore, only changes to the description of the criteria have been introduced without rejecting any of them. Nevertheless, the analysis presented above shows that in the assessment of importance of the criteria 1) the opinions of various groups of stakeholders should be considered, because the opinion of only one group may lead to wrong conclusions; 2) aggregated results of all stakeholder groups, as well as partial ones assigned individually to each stakeholder group should be presented, because they can provide information on potential directions of model improvement.

Once the fuzzy measures for perspectives and criteria are identified, and interaction index is computed the next step is to compute the Shapley value using formula 3 and 4 (Tab. 7 and 8). The goal is to determine relative importance between perspectives and between criteria.

The results presented in Table 7 indicate that regardless of the industry or department represented by experts, the most important is the 'Maintenance Stakeholders' (MS) perspective, and the 'Innovation and Development' (ID) perspective is the least important (Fig. 3a, 3b, 3c). Recognition of the MS as the most important perspective is understandable and results from the role played by maintenance in the enterprise. Analyzing the importance of the other two perspectives in most indications, the 'Maintenance processes' (MP) perspective is more important than the 'Financial perspective' (FP).

This means that the assumptions on the need to include non-financial criteria taken for the concept development are confirmed by the experts of both industries. Comparing the results of experts' assessment from the point of view of the industry (Fig. 3a), they are consistent only for extreme indications (the most important MS, the least important ID), while the other ones differ. In the food industry, the MP perspective is definitely more important than FP (compared to the automotive industry) in the opinion of both the heads of the maintenance department (Fig. 3b) and production managers (Fig. 3c). This difference may result from the specificity of the food industry. Inconsistencies caused by emergency events or incorrect performance of technical service affect the health safety of the food product, which may result, for example, in the disposal of the entire batch of the product, the need to carry out cleaning and disinfection of the machine, which is associated with a financial loss. Thus, assigning a higher importance

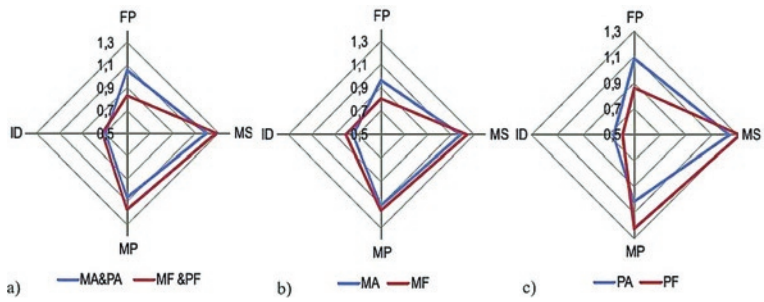


Fig. 3. Experts' perception of importance of perspectives

Table 8. Shapley value for criteria

Perspectives	Criteria	Automotive industry (A)			Food industry (F)		
		MA	PA	MA &PA	MF	PF	MF &PF
FP	f1	0.9722	1.1670	1.0642	1.0668	1.1670	1.1114
	f2	1.0278	0.8330	0.9358	0.9332	0.8330	0.8886
MS	s1	1.1516	1.3004	1.2132	1.2464	1.2716	1.2584
	s2	1.2192	1.2308	1.2132	0.9180	1.2716	1.0868
	s3	0.8820	0.8784	0.8980	1.0828	0.8552	0.9836
	s4	0.7468	0.5908	0.6756	0.7528	0.6020	0.6712
MP	p1	0.8890	0.8970	0.8945	0.8175	0.8060	0.8150
	p2	1.2360	1.3385	1.2910	1.2325	1.3355	1.2770
	p3	1.0970	1.1180	1.0925	0.9835	0.9825	0.9930
	p4	0.9585	0.9715	0.9605	1.1495	1.2475	1.1705
	p5	0.8195	0.6750	0.7615	0.8175	0.6285	0.7445
ID	id1	1.0647	1.1991	1.1004	1.0287	1.2147	1.1205
	id2	1.0647	0.9342	1.0104	1.1865	1.2147	1.1877
	id3	0.8706	0.8667	0.8892	0.7848	0.5706	0.6918

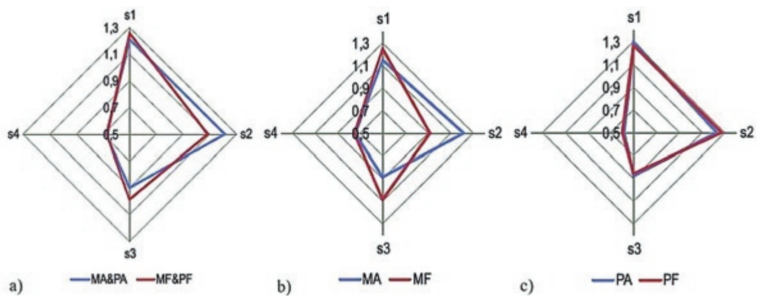


Fig. 4. Perception of importance of criteria from 'Maintenance Stakeholders' perspective

to the MP perspective is a kind of prevention against the increase of costs.

The results presented in Table 8 apply to the value of the Shapley index for the criteria describing particular perspectives. Analyzing the 'Financial perspective' (FP), all experts except the heads of the automotive industry maintenance department (MA) indicate that 'The costs of maintenance stakeholders - f1' is more important than 'maintenance costs - f2'. However, regardless of the industry (MA&PA and MF&PF), criterion f1 is more important than f2.

Assessing the perspective of 'Maintenance Stakeholders' (MS) (Fig. 4a), experts from both industries indicated that the first two criteria ('Production and quality - s1' and 'Safety and health - s2') are most important, followed by 'Environment - s3 'and' Communication and cooperation with stakeholders - s4 '. The assessments of experts representing production departments in both industries do not show any difference in preferences (Fig. 4c), while the differences are recognizable in the assessment of experts representing maintenance departments (Fig. 4b) and concern s2 and s3 criteria. According to MF experts, the criterion 'Environment - s3' is more important than 'Safety and health - s2', which may be related to the perception of the impact of emergency events primarily on the product (e.g. product disposal and related environmental impact).

Another perspective analyzed is 'Maintenance processes' (MP). The distribution of the importance of the assessment criteria in this perspective indicates (Fig. 5a) that the most important is the criterion 'Execution and measurement - p2', while the least important is the criterion 'Management of spare parts and consumables - p5'. For experts in both industries (Fig. 5a), the criterion of 'Management of external service - p4' is more important than 'Management of spare parts and consumables - p5'. This may indicate an increasing participation of third parties in the execution of maintenance work, as well as the transfer of responsibility for maintaining spare parts stocks. In the food industry (MF & PF (Fig. 5a), MF (Fig. 5b), PF (Fig. 5c)), this is the second, after the most important criterion 'Execution and measurement - p2'.

The fourth perspective is 'Innovation and development' (ID). According to the assessment of experts from the automotive industry (MA & PA) the most important criterion is 'Competence of maintenance workers - id1', while according to experts in the food industry (MF & PF) the most important criterion is 'Maintenance infrastructure - id2' (Table 8). Such a difference in perception of criteria importance may result from the production process and the product. The repeatable continuous production was the prevailing solution in companies representing the food industry. Hence, the use of, for example, technical diagnostic tools, their availability and quality is particularly important for the prevention of failures. Moreover, taking into account the assessment of food industry experts for the perspective (MP), where criterion (p4) is more important than (p5), indication that (id2) is more important than (id1) would be justified.

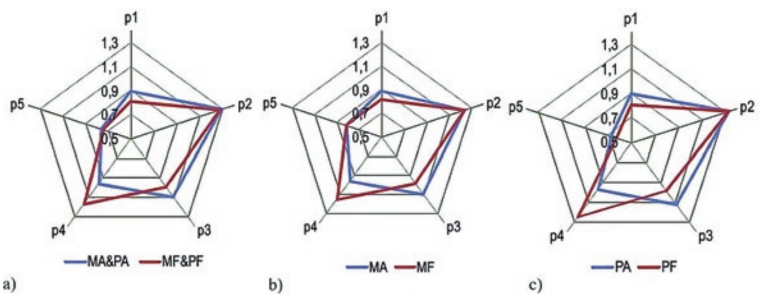


Fig. 5. Perception of importance of criteria from 'Maintenance Processes' perspective

5. Conclusion

To change the conventional vision of maintenance as a cost generator, it is necessary to integrate sustainable perspective in maintenance decision-making process. This requires including economic, environmental and social sustainability requirements in the maintenance management system in order to reduce maintenance-related impacts and their consequences and develop a new system for assessing maintenance performance taking into account this impact and its consequences. According to these challenges, in this paper the authors present an original model for maintenance sustainability assessment to integrate the sustainability-related aspects into the conventional maintenance management.

The internal multidimensional complexity of sustainable maintenance resulting both from the context of the enterprise and its objectives as well as from the scope of the impact of the maintenance processes implemented, makes it very difficult to assess maintenance activities in the context of sustainable production requirements. What is needed is both a synthetic evaluation of the entire maintenance area and individual aspects of its operation, so that it is possible to assess both the overall progress in performance and indicate areas for improvement. The presented assessment method meets these requirements. In addition, implementation of the assessment model will provide a systematic and gradual approach to structuring information (from whom, what, when, ...) that will enable policy makers to deal with the complexity of maintenance problems.

The model extends recent research work on Maintenance Performance Measurement by introducing the requirements of the sustainable manufacturing concept as the basis for performance measurement. Moreover, the model incorporates fuzzy integrals with fuzzy measure methodologies as the basis for construction CSMI. Fuzzy integral method applies fuzzy measures to deal with the problems of human subjective perception and uncertainty as well as to address the level of interdependency effects among criteria. To the best of the author's knowledge, such a framework of maintenance sustainability assessment is missing in previous studies.

When developing performance measurement models, the important issue is whether the presented model is universal and independent of the industry in which the company operates. The second part of the article presents preliminary results of the research on perception of importance of perspectives and criteria for assessing sustainable maintenance performance among experts from two industries. Results obtained from pilot studies show that in order to obtain reliable results, the research must be carried among representatives of different areas, not only heads of maintenance departments. Moreover, they indicate that industry conditions affect the perception of importance of assessment criteria, and thus the assessment model should be customized to the industry specifics.

Of course, there are several limitations to the research conducted and presented above, which require in-depth analysis and indicate directions for further research. First of all, the research was conducted on an expert sample represented by two industries. A larger sample would allow for more sophisticated assessment analysis of the importance and interactions between the assessment perspectives and the criteria describing them. It would be advisable to carry out research also in other industries and to include other company's functions in the research, for example, the HSE departments. Secondly, to provide more objective information on the applicability of the developed assessment model, further research should be carried out, using case studies of specific companies, thus confirming the practicality of the assessment procedure. Finally, because the CMSI calculation procedure is mathematically complex, which limits its application in practice, the intention of the authors is also to build an IT tool in the future. This tool would allow on the basis of an assessment of the criteria carried out by the evaluation team from the company to automatically generate CMSI value, and simultaneously through the built-in analytical module it would be possible to develop scenarios and assess their impact on the CMSI value, and thus identify directions for improvement.

References

1. Ajukumar V N, Gandhi O P. Evaluation of green maintenance initiatives in design and development of mechanical systems using an integrated approach. Journal of Cleaner Production 2013; 1-13, https://doi.org/10.1016/j.jclepro.2013.01.010.
2. 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.
3. Al-Turki U M, Ayar T, Yilbas B S, Sahin A Z. Integrated Maintenance Planning in Manufacturing Systems. Springer Briefs in Manufacturing and Surface Engineering 2014; 25-57, https://doi.org/10.1007/978-3-319-06290-7_3.
4. Antosz K. Maintenance - identification and analysis of the competency gap. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (3): 484-494, https://doi.org/10.17531/ein.2018.3.19.
5. Antosz K, Ratnayake R M C. Classification of spare parts as the element of a proper realization of the machine maintenance process and logistics - case study. IFAC-PapersOnLine 2016; 49(12): 1389-1393, https://doi.org/10.1016/j.ifacol.2016.07.760.
6. Bokrantz J, Skoogh A, Berlin C, Stahre J. Maintenance in digitalised manufacturing: Delphi-based scenarios for 2030. International Journal of Production Economics 2017; 191: 154-169, https://doi.org/10.1016/j.ijpe.2017.06.010.
7. Chen S J, Chen S M. A new method for handling multi-criteria fuzzy decision making problems using FN-IOWA operators. Cybernetics and Systems 2003; 34: 109-137, https://doi.org/10.1080/01969720302866.
8. Chiou H K, Tzeng G H. Fuzzy multicriteria decision-making approach to analysis and evaluation of green engineering for industry. Environmental Management 2002; 30(6): 816-830, https://doi.org/10.1007/s00267-002-2673-z.
9. Chopu-inwai R, Diaotrakun R, Thaiupathump T. Key indicators for maintenance performance measurement: The aircraft galley and associated equipment manufacturer case study. in: 10th International conference Service systems and service management (ICSSSM), 17-19 July, 2013; 844-849.
10. EN 13306 : 2018-01 Maintenance - Maintenance terminology
11. Galante G M, Inghilleri R, La Fata C M. A hierarchical framework for the measurement of maintenance efficacy and efficiency using performance indicators, in: Proceedings of the European Safety and Reliability Conference, ESREL 2014, 2015; 1109-1117.
12. Galar D, Parida A, Kumar U, Stenström C, Berger L. Maintenance metrics: a hierarchical model of balanced scorecard. IEEE International Conference on Quality and Reliability (IEEQR), Piscataway 2011; 67-74, https://doi.org/10.1109/ICQR.2011.6031683.
13. Galar D, Berges L, Sandborn P, Kumar U. The need for aggregated indicators in performance asset management. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2014; 16 (1): 120-127.
14. Gan X, Fernandez I C, Guo J, Wilson M, Zhao Y, Zhou B, Wu J. When to use what: Methods for weighting and aggregating sustainability indicators. Ecological Indicators 2017; 81: 491-502, https://doi.org/10.1016/j.ecolind.2017.05.068.
15. Garetti M, Taisch M. Sustainable manufacturing: trends and research challenges. Production Planning and Control 2012; 23(2-3): 83-104, https://doi.org/10.1080/09537287.2011.591619.
16. Grabisch M. The Application of Fuzzy Integrals in Multicriteria Decision Making. European Journal of Operational Research 1996; 89: 445-456, https://doi.org/10.1016/0377-2217(95)00176-X.
17. Grabisch M. K-order additive discrete fuzzy measures and their representation. Fuzzy Sets and Systems 1997; 92(2): 167-189, https://doi.org/10.1016/S0165-0114(97)00168-1.
18. Grabisch M, Labreuche C A. Decade of Application of the Choquet and Sugeno Integrals in Multicriteria Decision Making. Annals of Operations Research 2010; 175: 247-286, https://doi.org/10.1007/s10479-009-0655-8.
19. Grabisch M. Fuzzy Measures and Integrals: Recent Developments. Fifty years of fuzzy logic and its applications 2015; 125-151.

20. Gupta S, Gupta P, Parida A. Modeling lean maintenance metric using incidence matrix approach. International Journal of Systems Assurance Engineering and Management 2017; 8(4): 799-816, https://doi.org/10.1007/s13198-017-0671-z.
21. Gürbüz T, Alptekin S E, Alptekin G I. A hybrid MCDM methodology for ERP selection problem with interacting criteria. Decision Support Systems 2012; 54: 206-214, https://doi.org/10.1016/j.dss.2012.05.006.
22. Hu Y C, Chen H. Choquet Integral-based hierarchical networks for evaluating customer service perceptions on fast food stores. Expert Systems with Applications 2010; 37: 7880-7887, https://doi.org/10.1016/j.eswa.2010.04.049.
23. Ighravwe D E, Oke S A. Ranking maintenance strategies for sustainable maintenance plan in manufacturing systems using fuzzy axiomatic design principle and fuzzy-TOPSIS. Journal of Manufacturing Technology Management 2017; 28(7): 961-992, https://doi.org/10.1108/JMTM-01-2017-0007.
24. Iung B, Levrat E. Advanced Maintenance Services for Promoting Sustainability. Procedia CIRP 2014; 22: 15-22, https://doi.org/10.1016/j.procir.2014.07.018.
25. Jasiulewicz-Kaczmarek M. Role and contribution of maintenance in sustainable manufacturing. Manufacturing Modelling, Management, and Control 2013; 7(1): 1146-1151, https://doi.org/10.3182/20130619-3-RU-3018.00511.
26. Kaplan R S, Norton D P. The balanced scorecard: translating strategy into action. Harvard Business Press 1996.
27. Kłosowski G, Kozłowski E, Gola A. Integer linear programming in optimization of waste after cutting in the furniture manufacturing. Advances in Intelligent Systems and Computing 2018; 637: 260-270, https://doi.org/10.1007/978-3-319-64465-3_26.
28. Kosicka E, Kozłowski E, Mazurkiewicz D. The use of stationary tests for analysis of monitored residual processes. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2015; 17 (4): 604-609, https://doi.org/10.17531/ein.2015.4.17.
29. Kozłowski E, Optimal route determining for LQ problem with optimally stopped horizon, In 20th International Conference on Methods and Models in Automation and Robotics (MMAR 2015), Międzyzdroje, Poland, 24-27 August 2015; 553-557, https://doi.org/10.1109/MMAR.2015.7283935.
30. Kozłowski E, Optimal stopping of controlled linear stochastic systems, In 21th International Conference on Methods and Models in Automation and Robotics (MMAR 2016), Międzyzdroje, Poland, 29 August - 1 September 2016; 272-277, https://doi.org/10.1109/MMAR.2016.7575146.
31. Kumar U, Ellingsen H P. Design and development of maintenance performance indicators for the Norwegian oil and gas industry. Proceedings of the 15th European Maintenance Congress: Euromaintenance 2000, Gothenburg, Sweden 2000: 224-228.
32. Liyanage J P. Operations and maintenance performance in production and manufacturing assets: The sustainability perspective. Journal of Manufacturing Technology Management 2007; 18(3): 304-314, https://doi.org/10.1108/17410380710730639.
33. Magdi A M, Xiao W. Q-measures: an efficient extension of the Sugeno lambda-measure. IEEE Transactions on Fuzzy Systems 2003; 11(3): 419-426, https://doi.org/10.1109/TFUZZ.2003.812701.
34. Mather D. The Maintenance Scorecard: Creating Strategic Advantage. Industrial Press: New York 2005.
35. Moldavska A. Model-based Sustainability Assessment - an enabler for Transition to Sustainable Manufacturing. Procedia CIRP 2016; 48: 413-418, https://doi.org/10.1016/j.procir.2016.04.059.
36. Muchiri P, Pintelon L, Gelders L, Martin H. Development of maintenance function performance measurement framework and indicators. International Journal of Production Economics 2011; 131: 295-302, https://doi.org/10.1016/j.ijpe.2010.04.039.
37. Murofushi T. Techniques for reading fuzzy measures (I): The Shapley value with respect to a fuzzy measure. in: 2nd Fuzzy Workshop, Nagaoka, Japan, 1992; 39-48, in Japanese.
38. Murofushi T, Soneda S. Techniques for reading fuzzy measures (III): interaction index. in: 9th Fuzzy System Symposium, 693-696, Sapporo, Japan, May 1993. In Japanese
39. Nikolaou I E, Tsalis T. Development of a sustainable balanced scorecard framework. Ecological Indicators 2013; 34: 76-86, https://doi.org/10.1016/j.ecolind.2013.04.005.
40. Nooteboom S. Impact assessment procedures for sustainable development: A complexity theory perspective. Environmental Impact Assessment Review 2007; 27(7): 645-65, https://doi.org/10.1016/j.eiar.2007.05.006.
41. Parida A, Chattopadhyay G. Development of multi-criteria hierarchical framework for maintenance performance measurement (MPM). Journal of Quality in Maintenance Engineering 2007; 13(3): 241-258, https://doi.org/10.1108/13552510710780276.
42. Pires S P, Sénéchal O, Loures E F R, Jimenez J F. An approach to the prioritization of sustainable maintenance drivers in the TBL framework. IFAC-PapersOnLine 2016; 49-28: 150-155.
43. Raouf A. Productivity enhancement using safety and maintenance integration. An overview. Kybernetes 2004; 33(7): 1116-1126, https://doi.org/10.1108/03684920410534452.
44. Roy R, Stark R, Tracht K, Takata S, Mori M. Continuous maintenance and the future - Foundations and technological challenges. CIRP Annals - Manufacturing Technology 2016; 65: 667-688, https://doi.org/10.1016/j.cirp.2016.06.006.
45. Sadiq R, Tesfamariam S. Developing environmental indices using fuzzy numbers ordered weighted averaging (FN-OWA) operators, Stochastic Environmental Research & Risk Assessment 2008; 22(1): 494-505, https://doi.org/10.1007/s00477-007-0151-0.
46. Saniuk A, Saniuk S, Caganova D, Cambal M. Control of strategy realization in metallurgical production. in: 23rd International Conference on Metallurgy and Materials - Metals 2014; 1876-1881.
47. Sari E, Shaharoun A M, Maaram A, Yazid A M. Sustainable maintenance performance measures: a pilot survey in Malaysian automotive companies. Procedia CIRP 2015; 26: 443-448, https://doi.org/10.1016/j.procir.2014.07.16.
48. Savino M M, Macchi M, Mazza A. Investigating the impact of social sustainability within maintenance operations: An action research in heavy industry, Journal of Quality in Maintenance Engineering 2015; 21(3): 310-331, https://doi.org/10.1108/JQME-06-2014-0038.
49. Sobaszek Ł, Gola A, Kozłowski E, Application of survival function in robust scheduling of production jobs, Proceedings of the 2017 Federated Conference on Computer Science and Information Systems (FEDCSIS 2017); 575-578, https://doi.org/10.15439/2017F276.
50. Sugeno M. Theory of fuzzy integrals and its applications, Ph.D. theses, Tokyo Institute of Technology, Japan 1974.
51. Sugeno M, Terano T. A model of learning based on fuzzy information. Kybernetes (1997); 6(2): 157-166.
52. Takata S, Kimura F, van Houten F J A M, Westkämper E, Shpitalni M, Ceglarek D, Lee J. Maintenance: Changing role in life cycle management. Annals of the CIRP 2004; 53(2): 643-656, https://doi.org/10.1016/S0007-8506(07)60033-X.

53. Tsai H H, Lu I Y. The evaluation of service quality using generalized Choquet integral. Information Sciences 2006; 176(6): 640-663, https://doi.org/10.1016/j.ins.2005.01.015.
54. Tsang A H C. A strategic approach to managing maintenance performance. Journal of Quality in Maintenance Engineering 1998; 4(2): 87-94, https://doi.org/10.1108/13552519810213581.
55. Van Horenbeek A, Kellens K, Pintelon L, Duflou J R. Economic and environmental aware maintenance optimization. Procedia CIRP 2014; 5: 343-348, https://doi.org/10.1016/j.procir.2014.06.048.

Małgorzata JASIULEWICZ-KACZMAREK
Chair of Ergonomics and Quality Management
Faculty of Management Engineering
Poznan University of Technology
ul. 11 Strzelecka, 60-965 Poznań, Poland

Patryk ŻYWICA
Department of Imprecise Information Processing Methods
Faculty of Mathematics and Computer Science
Adam Mickiewicz University in Poznań
87 Umultowska Str. 61-614 Poznań, Poland

E-mails: malgorzata.jasiulewicz-kaczmarek@put.poznan.pl, bikol@amu.edu.pl

Article citation info:
WANG C, XU J, WANG H, ZHANG Z. A criticality importance-based spare ordering policy for multi-component degraded systems. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 662–670, http://dx.doi.org/10.17531/ein.2018.4.17.

Cheng WANG
Jianxin XU
Hongjun WANG
Zhenming ZHANG

A CRITICALITY IMPORTANCE-BASED SPARE ORDERING POLICY FOR MULTI-COMPONENT DEGRADED SYSTEMS

OPARTA NA KRYTERIUM KRYTYCZNOŚCI POLITYKA ZAMAWIANIA CZĘŚCI ZAMIENNYCH DO ZDEGRADOWANYCH SYSTEMÓW WIELOELEMENTOWYCH

With the increasing complexity and variety of production systems, more attention is being paid to preventive replacement on multi-component systems. Each component is non-identical and has its own degradation process. In this paper, we propose a criticality importance-based spare ordering policy for a complex system, which consists of multiple series-parallel degrading components. Replacement action is triggered whenever the system reliability drops below a lower threshold and spares for replacement are available. Our policy mainly consists of two steps: (1) determine which components to be replaced; (2) determine when to order spares for components selected. In step 1, when the replacement action is triggered, we select components that most need to be replaced within the system in accordance with the optimum ranking of components until the system meets an upper reliability threshold. In step 2, a spare ordering policy for components selected is made and the optimal spare ordering time is obtained by minimizing the expected replacement cost during the once replacement cycle. Finally, a numerical example is given to illustrate the proposed multi-spare ordering policy. Moreover, the proposed policy is of significance for safety-critical systems such as sub-station automation system, bridge system, nuclear power plants and aerospace equipment.

Keywords: *Multi-component, spare ordering, criticality importance, random lead-time, reliability threshold.*

Wraz ze wzrostem złożoności i różnorodności systemów produkcyjnych, coraz większą uwagę zwraca się na wymianę zapobiegawczą w systemach wieloelementowych. Każdy element takiego systemu jest nieidentyczny z pozostałymi elementami i charakteryzuje się własnym procesem degradacji. W niniejszym artykule proponujemy strategię zamawiania elementów zamiennych dla systemu złożonego składającego się z wielu ulegających degradacji komponentów tworzących strukturę szeregowo-równoległą. Omawiana strategia wymiany opiera się na kryterium krytyczności elementów. Akcja wymiany uruchamiana jest za każdym razem, gdy niezawodność systemu spada poniżej dolnego progu i dostępne są części zamienne. Na proponowaną strategię składają się zasadniczo dwa etapy: (1) określenie elementów wymagających wymiany oraz (2) określenie terminu zamówienia części zamiennych do wybranych elementów. W 1. etapie, po uruchomieniu akcji wymiany, wybiera się komponenty systemu, które najpilniej wymagają wymiany, kierując się optymalnym rankingiem komponentów, do momentu aż system osiągnie górny próg niezawodności. W 2. etapie, opracowuje się politykę zamawiania części zamiennych dla wybranych komponentów oraz określa się optymalny czas zamawiania części zamiennych poprzez minimalizację oczekiwanego kosztu wymiany podczas jednego cyklu wymiany. W artykule przedstawiono przykład numeryczny, który ilustruje proponowaną strategię jednoczesnego zamawiania wielu części zamiennych. Proponowana strategia może znaleźć zastosowanie w systemach o kluczowym znaczeniu dla bezpieczeństwa, takich jak systemy automatyki podstacji, systemy mostowe, elektrownie jądrowe i sprzęt lotniczy.

Słowa kluczowe: *system wieloelementowy, zamawianie części zamiennych, krytyczność, losowy czas realizacji procesu produkcyjnego, próg niezawodności.*

1. Introduction

Maintenance plays an important role in industrial production and system safety, especially in areas where the loss of system failure is large. Various maintenance policies have been developed to improve system safety (or system reliability), reduce system failure and manufacturing cost [1]. In non-repairable systems, preventive replacement (PR) [33, 12, 13, 6, 15, 17] is a policy that occurs when a system is still operating, aiming to renew the system or components. In PR policies, condition-based maintenance (CBM) is a more promising maintenance policy since it emphasizes on combining data-driven reliability models with condition monitoring data. Therefore, CBM has received considerable attention in both academia and industry [2,8]. Most CBM policies are developed under the implicit assumption that at any time there is an unlimited supply of available spares for replacement. However, this assumption is generally unrealistic and unpractical when available spares are limited and/or delivery lead times are much longer. When spares are expensive, scarce, and with higher and random lead times, it is important to consider shortage cost and holding cost. Therefore, proper supply of spares is essential for maintenance [16].

In practice, the performance of a PR policy depends not only on the operating condition of a system but also on the availability of spares. In order to order spares on demand and achieve the minimum maintenance cost, the joint of a system condition and spares ordering is very necessary. Motivated by the idea of joint optimization of maintenance and spare ordering, some spare ordering policies have been extensively researched. Wang, L., et al. proposed a joint optimization of condition-based maintenance and spare ordering management for a single-component system [34]. Chien, Y.H. proposed a spare ordering policy based on the optimal number of minimal repairs with regular lead-time [7]. Louit, D., et al. presented an order policy based on remaining useful life of a component [23]. Godoy, D.R., et al. presented an order policy through graphic technique, which depended on condition-based reliability function and lead-time [11]. Panagiotidou, S. proposed a joint optimization of spares ordering and maintenance policies for multiple identical items [27]. Wang, Z.Q., et al. proposed a condition-based spare ordering policy with random lead-time for a deteriorating system [35]. Chen, X., et al. proposed a joint optimization of replacement and spare ordering for critical rotary component based on collected condition monitoring signals [5]. Cai, J., et al. proposed an appointment policy of spares based on (s,S) policy [4]. Lin, X., et al. proposed a condition based spare parts supply policy that is more efficient on average than a standard, state-independent base stock policy [21]. In the literature, spare ordering policies mainly focus on a single-component system or multiple identical components.

Importance measures (IM) are used in various fields to evaluate the relative importance of various objects such as components in a system [19]. IM would be capable of the needs of selecting components within a complex system. IM is widely used in systems engineering to identify components within a system that more significantly influence the system behavior with respect to reliability, risk and/or safety. The information gathered by the use of IM provides management with useful insights for the safe and efficient operation of a system. IM is valuable in suggesting the most effective way to operation and maintain system status. In general, IM is used to quantify the contribution of individual components of a system to the overall system performance (e.g., reliability, risk, availability) [24,9]. Several IM such as Birnbaum's measure [3], Fussell-Vesely's measure [31,10], risk achievement worth [32], risk reduction worth [20] and criticality importance (CI) [18] for components have been proposed in the past. For more applications, see [19] for an overview about recent advances on IM. Recently, IM provides an efficient tool to solve multi-component maintenance problems [36]. More recently, IM have been applied for maintenance optimization of a multi-component system with complex structure [25,26]. Especially, in the work of Nguyen, K-A., et al. (2017), the authors developed a joint predictive maintenance and inventory strategy for multi-component systems using Birnbaum's structural importance [26]. Whereas Nguyen, K-A., et al. (2017) focused on PR threshold and ordering threshold of each component and IM (Birnbaum's structural importance) is used to reduce the number of decision parameters, we proceed to solve a multi-spare optimal ordering problem for components that most need to be replaced based on system reliability threshold and spares random lead time, and IM is used to select components that most need to be replaced within the system. On the other hand, from the perspective of overall system reliability, multi-spare ordering and replacement can be used as a complementary method to [26]. To the best of our knowledge, studies that investigate multi-spare ordering and replacement for multi-component complex systems are relatively rare. Therefore, it is of great importance to study the multi-spare ordering policy based on system reliability threshold and IM for solving the multi-component replacement of complex systems. After investigating IM, we select two appropriate measures of the importance including CI measure and Birnbaum's measure to quantify the component importance in a complex degraded system.

In recent years, due to the increasing complexity and variety of production systems, more attention should be paid to spares ordering on a complex degraded system composed of many non-identical components. Since each component has its own contribution to the system, it is essential to select the most important components as the replaced objects to ensure the safe and reliable operation of the system. In view of the advantages of IM, this paper attempts to find a policy to solve the issue of multi-spare ordering. Therefore, this paper aims to propose a criticality importance-based spare ordering policy for the system with continuously degrading components. Each component has its own degradation process. Replacement action is triggered whenever the system reliability drops below a lower threshold and spares for replacement are available. Our policy mainly consists of two steps: (1) determine which components to replace though the optimum ranking of components; (2) determine when to order spares for components selected to minimize the expected replacement cost during the once replacement cycle. The main contribution of this paper is to propose a novel multi-spare ordering policy based on CI for the complex degraded system. Under the condition of overall system reliability constraint, the problem of how to select the most needed spares and when to place an order with minimized maintenance cost is solved.

The remainder of this paper is organized as follows. Section 2 describes the problem statement. Section 3 constructs a system reliability model and develops the policy of components selection. Section 4 develops a novel multi-spare ordering policy for a complex system. Section 5 gives a numerical example and performs sensitivity analysis on critical parameters. Finally, Section 6 concludes the study.

Notations and Nomenclatures

PR	Preventive replacement	σ_2	System upper threshold
CBM	Condition-based maintenance	R_i^{π}	Reliability of component i after one maintenance action
IM	Importance measures	R_i^{γ}	Reliability of component i after one replacement action
CI	Criticality importance	$\left\{I_1^{CR}, I_2^{CR}, \ldots, I_m^{CR}\right\}$	Optimum ranking
BI	Birnbaum importance	$\left\{c_{I_1^{CR}}, c_{I_2^{CR}}, \ldots, c_{I_{m_s}^{CR}}\right\}$	Components selected sequence
CDF	Cumulative distribution function	T_r	System PR time
MEMS	micro-electro-mechanical systems	T	Ordering time
X_i	Degradation level of component i	L	Spare lead-time
μ_i	Degradation rate of component i	P_{S1}	CDF of State 1
ε_i	Error term of component i	P_{S2}	CDF of State 2
L_i	Degradation threshold value of component i	$W(t)$	CDF of the lead-time
$R_i(t)$	Reliability of component i	ρ_h	Holding cost per unit time
$F_i(t)$	Unreliability of component i	ρ_s	Shortage cost per unit time
$R(t)$	Reliability of system	C	spares cost
$F(t)$	Unreliability of system	EH	Expected holding time
$I_i^B(t)$	BI of component i	ES	Expected shortage time

$I_i^{CR}(t)$	CI of component i	EV	Expected replacement cost
σ_1	System lower threshold	$T^{\bullet}$	Optimal ordering time

2. Problem statement

Consider a complex system consisted of n different components. Each component has its own degradation process. The system reliability is determined by component reliability. In a mission, the system may be maintained many times. The system reliability variation during a mission timespan is depicted in Fig. 1. Under the premise of ensuring safe and reliable operation, how to minimize maintenance costs is a problem that must be solved. The lower and upper threshold values of the system reliability directly affect the execution reliability and maintenance cost of the entire mission. Under normal conditions, the lower threshold is a constant and provided by domain experts. It mainly affects the system safety. However, the upper threshold is a variate. It mainly affects the PR times and the entire maintenance cost in a mission. Considering that each maintenance process is similar, this paper only investigates the first maintenance process. To facilitate the study of a maintenance process, we preset the upper limit as a constant in the first PR action. Replacement action is triggered whenever the system reliability drops below a lower threshold and spares for replacement are available. Therefore, the challenge is to identify an optimal spare ordering policy for most needed components, in order to meet both system reliability constraints and minimum maintenance cost in engineering practice.

To ensure the effectiveness of this study, the following assumptions are used:

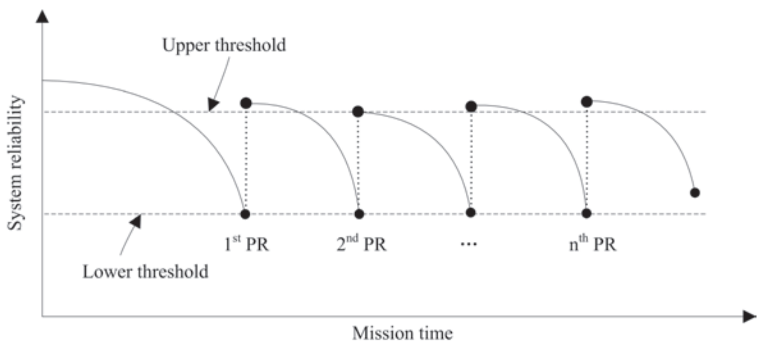


Fig. 1. System reliability variation during a mission timespan

(1) Components are mutually independent and each component is continuously monitored.

(2) Degradation is the only cause of each component failure. The impact of the external environment on the component is not considered, such as artificial destruction and natural disaster.

(3) The system does not degenerate when it is suspended during operation.

(4) Shortage cost per unit time is bigger than holding cost per unit time due to the system shutdown affecting the production progress and custom service negatively.

(5) Spares are supplied by the identical manufacturer.

3. Model statement

3.1. System reliability modeling

The system considered here is a series-parallel system with many different components. In addition, each component has its own contribution to the system and the reliability of the system is measured by the reliability of these components. The failure degree of component is measured by its degradation level. Each component's degradation mechanism follows its own degradation path. Denote the degradation level of component i over time t as $X_i(t;\mu_i,\varepsilon_i)$, where μ_i is degradation rate, and ε_i is error term, i.e., $\varepsilon_i \sim N(0,\sigma_i^2)$. In most cases, $X_i(t;\mu_i,\varepsilon_i)$ is a monotonic function over time t [22,29].

For each component, it fails whenever the degradation level X_i exceeds threshold value L_i. The set of failure threshold values, $L=\{L_1,L_2,...,L_n\}$, is assumed to be pre-set. Without loss of generality, degradation level is assumed to be monotonically increasing, and reliability of component i at time t is represented by the probability that X_i stays below threshold L_i, that is:

$$R_i(t)=\Pr\{X_i(t;\mu_i,\varepsilon_i)<L_i\} \tag{1}$$

Let $X_i(t;\mu_i,\varepsilon_i)=\mu_i t+\varepsilon_i$, and $R_i(t)$ can be obtained as:

$$R_i(t)=\Pr\{X_i(t;\mu_i,\varepsilon_i)<L_i\}=\Pr\{\mu_i t+\varepsilon_i<L_i\}=\Pr\{\varepsilon_i<L_i-\mu_i t\}=\Phi\left(\frac{L_i-\mu_i t}{\sigma_i}\right), \tag{2}$$

where $\Phi(\bullet)$ is the cumulative distribution function (CDF) of standard normal distribution.

For a series-parallel system composed of m subsystems, with each subsystem containing n_i components, its reliability can be obtained as:

$$R(t)=\prod_{i=1}^{m}\left(1-\prod_{j=1}^{n_i}\left(1-R_{ij}(t)\right)\right), \tag{3}$$

where $R_{ij}(t)$ is the reliability of each component. It is easy to know that the system contains a total of $n=\sum_{i=1}^{m} n_i$ components.

3.2. Components selection

In view of the advantages of IM, we proposed a method on components selection based on IM. For parallel subsystems, each component within subsystem has the same value of CI. Therefore, for a series-parallel system, we select CI measure and Birnbaum's measure for optimum ranking of components. On the basis of optimum ranking of components, we select components that most need. Next, we first introduce the definitions of Birnbaum's measure and CI.

Birnbaum [3] first introduced the concept of importance in 1969 and it is one of the most widely used reliability importance measures. Analytically, it is defined by:

$$I_i^B(t)=\frac{\partial R(t)}{\partial R_i(t)}=R(t;R_i(t)=1)-R(t;R_i(t)=0), \tag{4}$$

where $I_i^B(t)$ is the Birnbaum importance (BI) of component i; $R(t)$ is the system reliability at time t; $R_i(t)$ is the reliability of component i at time t; $R(t;R_i(t)=1)$ is the system reliability at time t when the component i functions; $R(t;R_i(t)=0)$ is the system reliability at time t when the component i fails.

From the definition, the Birnbaum's measure may serve as a good indicator for selecting components that are the best candidates for efforts leading to improving system reliability. However, $I_i^B(t)$ does not depend on the component reliability $R_i(t)$. This is a weakness of Birnbaum's measure. To solve this weakness of Birnbaum's measure, the CI is proposed. The CI includes the component unreliability $F_i(t)$. The CI includes the component unreliability $F_i(t)$. The CI can be defined by:

$$I_i^{CR}(t)=I_i^B(t)\frac{F_i(t)}{F(t)}, \tag{5}$$

where $F(t)$ is the system unreliability at time *t* and $F_i(t)$ is the unreliability of component *i* at time *t*.

As compared with Birnbaum's measure, CI is more suitable for prioritizing maintenance action in complicated systems. However, for parallel subsystems, CI cannot identify the importance of each component. Therefore, to make up for the lack of CI, we adopt CI and BI for optimum ranking for a series-parallel system. The rank of components mainly consists of three steps: (1) we compute CI of each component and rank components according to CI. (2) We compute BI of parallel components within subsystems, and rank components within subsystems according to BI. (3) Only keep the maximum BI of components in each subsystem composed of parallel components and form the optimum ranking. The optimum ranking set of components can be expressed as $\left\{I_1^{CR},I_2^{CR},\cdots,I_m^{CR}\right\}$, where $\left\{I_m^{CR}\right\}$ denotes the CI value of component with the maximum BI in mth subsystem.

Secondly, we develop a policy for components selection based on optimum ranking of components. We define σ_1 as reliability threshold of system PR, that is, lower threshold. And we define σ_2 as system reliability upper threshold after replacement. When system reliability drops below a lower threshold, we compute the optimum ranking of components. We select the components that most need to be replaced within the system in accordance with the optimum ranking of components until the system reliability is improved above the upper threshold.

To describe the effect of a replacement action on component reliability, we first assume that a small maintenance action will improve component reliability by an infinitesimal positive shift, that is:

$$R_i^{\pi}=R_i(t-\varepsilon), \tag{6}$$

where $\varepsilon>0$ denotes infinitesimal reliability transposition due to maintenance. Considering a replacement action is equivalent to performing infinite maintenance until component reliability will be improved to one. On the basis of Eq. (6), a replacement action will improve component reliability by a positive shift of system PR time, that is:

$$R_i^{\gamma}=R_i(t-T_r) \tag{7}$$

Therefore, for a series-parallel system, we denote $\left\{c_{I_1^{CR}},c_{I_2^{CR}},\ldots,c_{I_{m_s}^{CR}}\right\}$ as components selected sequence set, that is:

$$\begin{aligned}
&\left\{c_{I_1^{CR}},c_{I_2^{CR}},\ldots,c_{I_{m_s}^{CR}}\right\}\\
&=\inf_{\left\{c_{I_1^{CR}},c_{I_2^{CR}},\ldots,c_{I_{m_s}^{CR}}\right\}}\left\{\prod_{i=1}^{m_s}\left(1-\frac{\left(1-R_{I_i^{CR}}^{\gamma}(t)\right)\cdot\prod_{j=1}^{n_i}\left(1-R_{ij}(t)\right)}{\left(1-R_{I_i^{CR}}(t)\right)}\right)\cdot\prod_{i=m_s+1}^{m}\left(1-\prod_{j=1}^{n_i}\left(1-R_{ij}(t)\right)\right)\geq\sigma_2\left|\left\{I_1^{CR},I_2^{CR},\cdots,I_{m_s}^{CR},\ldots,I_m^{CR}\right\},m_s\leq m\right.\right\}\\
&=\inf_{\left\{c_{I_1^{CR}},c_{I_2^{CR}},\ldots,c_{I_{m_s}^{CR}}\right\}}\left\{\prod_{i=1}^{m_s}\left(1-\frac{\left(1-R_{I_i^{CR}}(t-T_r)\right)\cdot\prod_{j=1}^{n_i}\left(1-R_{ij}(t)\right)}{\left(1-R_{I_i^{CR}}(t)\right)}\right)\cdot\prod_{i=m_s+1}^{m}\left(1-\prod_{j=1}^{n_i}\left(1-R_{ij}(t)\right)\right)\geq\sigma_2\left|\left\{I_1^{CR},I_2^{CR},\cdots,I_{m_s}^{CR},\ldots,I_m^{CR}\right\},m_s\leq m\right.\right\}
\end{aligned} \tag{8}$$

where m_s denotes subsystems selected, and $\left\{I_1^{CR},I_2^{CR},\cdots,I_{m_s}^{CR},\ldots,I_m^{CR}\right\}$ denotes the optimum ranking set of components.

4. Multi-spare ordering policy

In this section, we will build replacement cost model to find the optimal ordering time. To this end, we first propose two cases of spare ordering when ordering time occurs before the PR time, that is, the ordered spares are delivered before PR time or after PR time. The detailed spare ordering policy for a system is depicted in Fig. 2 and some parameters are explained as follows, where T_r is system PR time, T is ordering time, and L is spare lead-time. Specifically, the implications of these two cases are summarized as follows.

case1: If spares arrive before system PR time, replacement action is triggered on system PR time and the components selected are replaced by the spares in stock. Let P_1 denote the probability of the current state and corresponding CDF can be shown as:

$$P_1=\Pr\{T+L<T_r\}=\int_0^{T_r-T}\mathrm{d}W(t) \tag{9}$$

case2: If spares arrive after system PR time, replacement action is triggered as soon as the spares arrive. Let P_2 denote the probability of the current state and corresponding CDF can be shown as:

$$P_2=\Pr\{T+L>T_r\}=\int_{T_r-T}^{\infty}\mathrm{d}W(t) \tag{10}$$

Second, we will build objective function, that is, replacement cost model. The most important task is to express the expected holding time and expected shortage time:

Since holding time occurs in the case1, the expected holding time EH during the once replacement cycle is expressed as:

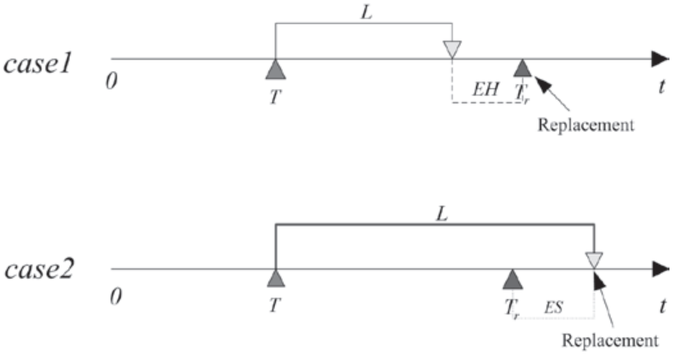


Fig. 2. Possible order-replacement states of one cycle

$$EH = E\left[T_r - T - L\right] = \int_0^{T_r - T} \left(T_r - T - t\right) \mathrm{d}W\left(t\right) \quad (11)$$

Similarly, since the shortage time occurs in the case2, the expected shortage time ES during the once replacement cycle is expressed as:

$$ES = E\left[T + L - T_r\right] = \int_{T_r - T}^{\infty} \left(T + t - T_r\right) \mathrm{d}W\left(t\right) \quad (12)$$

The replacement cost during the once replacement cycle mainly includes spares cost $\sum_{i=1}^{m_s} C_{I_i^{CR}}$, ordering cost C_o , holding cost $\rho_h \cdot EH$ and shortage cost $\rho_s \cdot ES$, that is,

$$EV\left(T\right) = \sum_{i=1}^{m_s} C_{I_i^{CR}} + C_o + \rho_h \cdot EH + \rho_s \cdot ES, \; 0 < T < T_r \quad (13)$$

What we aim is to seek an optimal ordering time $T^{\bullet}$ by minimizing the replacement cost, that is:

$$T^{\bullet} = \min_{T^{\bullet}} \left\{EV\left(T\right)\right\} \quad (14)$$

To facilitate the implementation, the detailed process of spare ordering policy is depicted in Fig. 3. The following content presents detailed steps.

Step1: Building system reliability model. The system reliability is determined by the reliability of each component and can be computed by Eq.(3).

Step2: Selecting components we want to order. We select the most important components using the components selection method in Section 3.2, and replace components selected with spares until the system reliability meets its upper threshold.

Step3: Making spare ordering policy. According to the distribution of the system reliability, components selected and spare lead-time, we consider two cases of spare ordering policy.

Step4: Finding the optimal ordering time. We first build the replacement cost model during the once replacement cycle, and then find the optimal ordering time by minimizing the replacement cost. The optimal ordering time is obtained by search algorithm with given step length.

5. A numerical example

A complex electromechanical system, a typical multi-component system, is the lifeline of the national economy and security. With the advancement of science and technology, and the modern large-scale production, the complex electromechanical system regarded as the key element in manufacturing industry, is developing toward large scale, automation, integration with mechanic, electric, hydraulic and computer technology, while the updating cycle is shorter and shorter. The failure rate is increasing, failure modes are various, and even the disastrous accident happens frequently, which are resulting from multi-function, improved performance, and heavy load of the complex system. Therefore, how to effectively improve the quality and reliability of the complex electromechanical system has become a key proposition and cannot be ignored in the national development strategy. Most complex electromechanical systems can be converted into an equivalent series-parallel system. Therefore, a numerical example of a series-parallel system can provide reference values for the maintenance and reliability of complex systems.

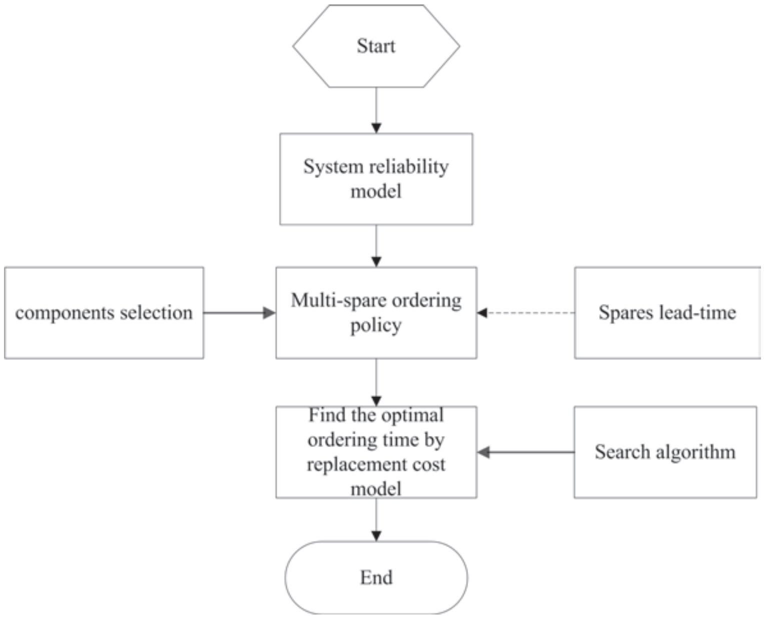


Fig. 3. The detailed process of multi-spare ordering policy

5.1. Specifications of the system and components

To show the implementation procedure of the modeling and analysis proposed in this paper, an illustrative example of a 6-component series-parallel system is used, as show in Fig. 4.

The Fig. 4 shows that the system reliability is:

$$R\left(t\right) = R_1\left(t\right) \cdot \left(1 - \left(1 - R_2\left(t\right)\right)\left(1 - R_3\left(t\right)\right)\right) \cdot \left(1 - \left(1 - R_4\left(t\right)\right)\left(1 - R_5\left(t\right)\right)\left(1 - R_6\left(t\right)\right)\right) \quad (15)$$

Many failures can be traced to underlying degradation, such as the wear on rubbing surfaces of a micro-electro-mechanical systems (MEMS) system composed of many non-identical components. In our study, each component follows a linear degradation path. This linear model has been used to characterize the failure mechanism of the wearing process in MEMS [29,30]. According to [29,30], the value of component-specific parameters is summarized in Table 1.

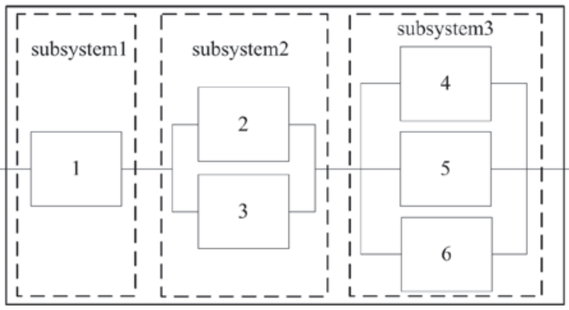


Fig. 4. 6-component series-parallel system

Table 1. Component-specific parameters

c_i	μ_i	ε_i	L_i	C_i
1	1.06	$N(0,1)$	10	0.40
2	1.16	$N(0,1)$	8	0.20
3	1.20	$N(0,2)$	10	0.30
4	1.18	$N(0,2)$	9	0.18
5	1.20	$N(0,3)$	9	0.15
6	1.10	$N(0,3)$	10	0.35

5.2. Ordering decision for the first PR action

5.2.1. Components selection

From the definition of CI, the CI of each component can be obtained as:

$$I_1^{CR}(t)=\frac{\left(1-\left(1-R_2(t)\right)\left(1-R_3(t)\right)\right)\cdot\left(1-\left(1-R_4(t)\right)\left(1-R_5(t)\right)\left(1-R_6(t)\right)\right)\cdot\left(1-R_1(t)\right)}{\left(1-R(t)\right)}$$
$$I_2^{CR}(t)=\frac{R_1(t)\cdot\left(1-R_3(t)\right)\cdot\left(1-\left(1-R_4(t)\right)\left(1-R_5(t)\right)\left(1-R_6(t)\right)\right)\cdot\left(1-R_2(t)\right)}{\left(1-R(t)\right)}$$
$$I_3^{CR}(t)=\frac{R_1(t)\cdot\left(1-R_2(t)\right)\cdot\left(1-\left(1-R_4(t)\right)\left(1-R_5(t)\right)\left(1-R_6(t)\right)\right)\cdot\left(1-R_3(t)\right)}{\left(1-R(t)\right)}$$
$$I_4^{CR}(t)=\frac{R_1(t)\cdot\left(1-\left(1-R_2(t)\right)\left(1-R_3(t)\right)\right)\cdot\left(\left(1-R_5(t)\right)\left(1-R_6(t)\right)\right)\cdot\left(1-R_4(t)\right)}{\left(1-R(t)\right)}$$
$$I_5^{CR}(t)=\frac{R_1(t)\cdot\left(1-\left(1-R_2(t)\right)\left(1-R_3(t)\right)\right)\cdot\left(\left(1-R_4(t)\right)\left(1-R_6(t)\right)\right)\cdot\left(1-R_5(t)\right)}{\left(1-R(t)\right)}$$
$$I_6^{CR}(t)=\frac{R_1(t)\cdot\left(1-\left(1-R_2(t)\right)\left(1-R_3(t)\right)\right)\cdot\left(\left(1-R_4(t)\right)\left(1-R_5(t)\right)\right)\cdot\left(1-R_6(t)\right)}{\left(1-R(t)\right)} \quad (16)$$

From Eq. (16), it can be seen that $I_2^{CR}(t)=I_3^{CR}(t), I_4^{CR}(t)=I_5^{CR}(t)=I_6^{CR}(t)$. It is because that parallel components has the same value of CI in subsystems. In order to increase the degree of differentiation, we add the BIs of components, that is,:

$$I_2^B(t)=R_1(t)\cdot\left(1-R_3(t)\right)\cdot\left(1-\left(1-R_4(t)\right)\left(1-R_5(t)\right)\left(1-R_6(t)\right)\right)$$
$$I_3^B(t)=R_1(t)\cdot\left(1-R_2(t)\right)\cdot\left(1-\left(1-R_4(t)\right)\left(1-R_5(t)\right)\left(1-R_6(t)\right)\right)$$
$$I_4^B(t)=R_1(t)\cdot\left(1-\left(1-R_2(t)\right)\left(1-R_3(t)\right)\right)\cdot\left(\left(1-R_5(t)\right)\left(1-R_6(t)\right)\right)$$
$$I_5^B(t)=R_1(t)\cdot\left(1-\left(1-R_2(t)\right)\left(1-R_3(t)\right)\right)\cdot\left(\left(1-R_4(t)\right)\left(1-R_6(t)\right)\right)$$
$$I_6^B(t)=R_1(t)\cdot\left(1-\left(1-R_2(t)\right)\left(1-R_3(t)\right)\right)\cdot\left(\left(1-R_4(t)\right)\left(1-R_5(t)\right)\right) \quad (17)$$

PR action would be triggered when the system reliability reaches the lower threshold $\sigma_1=0.70$. Therefore, PR time can be expressed as $T_r=\arg\inf_{T_r}\left\{t:R(t)\le\sigma_1\right\}$. In the PR time point, we compute the CIs of components by Eq. (16), and compute the BIs of components with the same value of CI by Eq. (17). On the basis of components selection method proposed above, we would select the components that most need to be replaced within the system in accordance with the optimum ranking of components until the system reliability meets above upper threshold $\sigma_2=0.95$. The specific simulation results are shown in Fig. 5 and Table 2. Table 2 shows components selection sequence and prepares for replacement. Fig. 5 shows the system reliability variation within the first replacement cycle. At the time point of 7.68, the system reliability goes below $\sigma_1=0.70$, i.e., $R(7.68)=0.7000$. After replacement, the system reliability goes above $\sigma_2=0.95$, i.e., $R^{\gamma}(7.68)=0.9685$.

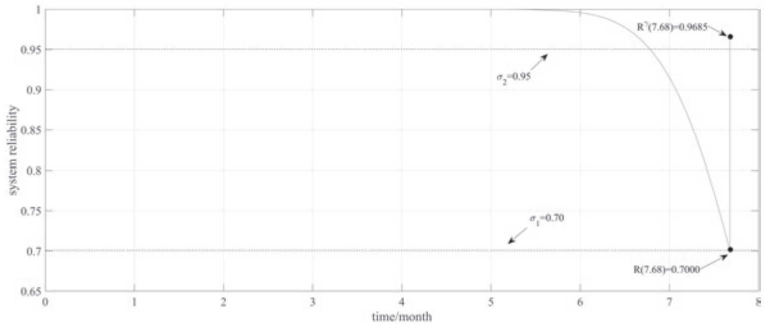


Fig. 5 System reliability variation within the first replacement cycle

Table 2. Components selection sequence

σ_1	0.70
T_r	7.68
Rank of CI	$I_2^{CR}=I_3^{CR}=0.7250, I_4^{CR}=I_5^{CR}=I_6^{CR}=0.1297, I_1^{CR}=0.0759$
Rank of BI with the same value of CI	$I_2^{CR}=I_3^{CR}\left(I_3^B=0.7508, I_2^B=0.2658\right)$, $I_4^{CR}=I_5^{CR}=I_6^{CR}\left(I_6^B=0.2102, I_4^B=0.0752, I_5^B=0.0708\right)$
Optimum Ranking set	$\left\{I_3^{CR}, I_6^{CR}, I_1^{CR}\right\}$
σ_2	0.95
Components selected	c_3, c_6

5.2.2. Multi-spare ordering

We need to clarify the values of cost parameters and lead-time parameter before spare ordering. Therefore, the values of cost parameters in spare ordering policy are assumed in Table 3. Assume further that the lead-time for delivering an ordered spares is Normal distributed, which is also one of the most common distributions in spare ordering investigation [28]. In our study, let random variable $X\sim N\left(\mu_w,\sigma_w^2\right)$, the CDF of an ordered lead-time is defined as $W(t)=P(X\le t\mid X\ge 0)$. Thus, the random lead-time will not be less than zero. For the purpose of illustration, the mean and standard deviation of the lead-time for delivering an ordered spares is $\mu_w = 2$ and $\sigma_w = 0.3$. Note that $W(t)$ used here are obtained by consulting with supplier of the spares.

As cost parameters and spare lead-time mentioned above, we utilize the proposed replacement cost model to find the optimal ordering time. The optimal ordering time is obtained by search algorithm with given step length as 0.1 and implemented computationally with MATLAB. The change law between ordering time and expected replacement cost is shown in Fig. 6. As seen in Fig. 6, the optimal ordering time is obtained by minimizing the expected replacement cost, i.e., $T^{\bullet} = 5.5, EV = 0.6817$. Considering factors like system preventive maintenance time and random lead-time error, the value of the optimal ordering time is reasonable.

Table 3. The values of cost parameters

C	C_o	ρ_s	ρ_h
$C_3 + C_6$	0.03	0.01	0.005

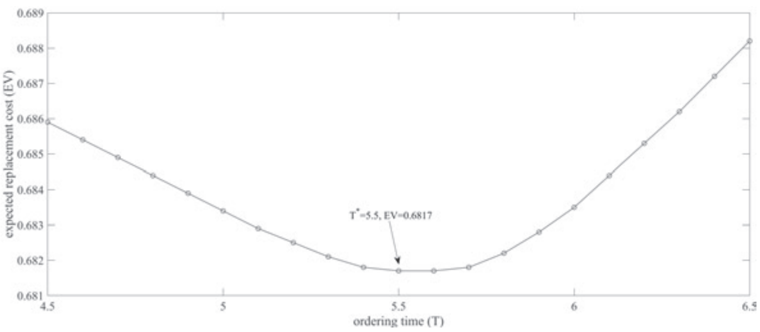


Fig. 6. The change law between ordering time and expected replacement cost

5.3. Sensitive analyses of critical parameters

In this section, we provide the sensitivity analysis on some critical parameters in order to verify the applicability of the proposed policy.

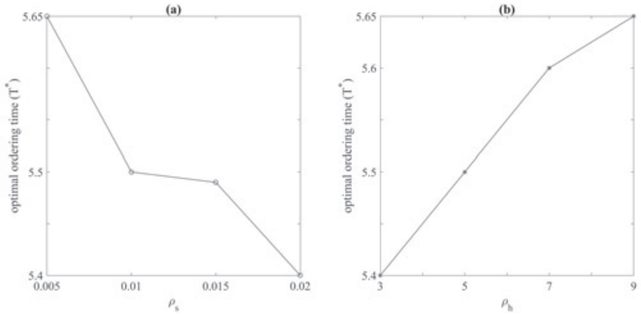


Fig. 7. (a) Sensitivity of optimal ordering time on ρ_s; (b) Sensitivity of optimal ordering time on ρ_h

We first provide critical parameters on components selection. In Table 4, the influence of σ_1 on the selecting components is studied, where σ_1 takes value from 0.60 to 0.80 with step size 0.10. We can see that the selecting components is affected by σ_1. An explanation for this is that the system reliability needs to replace more components up to the upper threshold as reliability lower threshold decreases. In Table 5, the influence of σ_2 on the selecting components is studied, where σ_2 takes value from 0.90 to 1.00 with step size 0.05. From Table 5, we can find that the number of components selected is growing as σ_2 increases. The reason for this is in that system reliability needs to replace more components up to the upper threshold.

In Table 6, we can find that the optimal ordering time $T^{\bullet}$ deceases as the standard deviation σ_w increases. The reason for this may lie in that the probabilities that the ordered spares can be delivered earlier or later than its mean lead-time will be along with the variation of the standard deviation and these probabilities are highly correlated with the expected shortage time and holding time among the once replacement cycle. In the present case, the optimal ordering time $T^{\bullet}$ decreases since the shortage cost is larger than holding cost. Therefore, the optimal ordering time moves backward.

Table 6. Sensitivity of optimal ordering time on σ_w

σ_w	$T^{\bullet}$	EV
0.1	5.6	0.6806
0.3	5.5	0.6817
0.5	5.4	0.6827

To further validate the applicability of the proposed policy, we conduct some sensitivity analysis on the two cost parameters, i.e., shortage cost per unit time and holding cost per unit time. In Fig. 7, they show the impact of ordering time on ρ_s and ρ_h, where ρ_s takes value from 0.005 to 0.02 with step size 0.005 and ρ_h takes value from 0.003 to 0.009 with step size 0.002. We can find that the optimal ordering time decreases as ρ_s increases and the optimal ordering time increases as ρ_h increases. The reason is simply that when the shortage cost is larger, one should place an order earlier. Similarly, when the holding cost is larger, one should postpone ordering.

Table 4. The influence of σ_1 on the components selected

σ_1	T_r	Optimum Ranking set	σ_2	Components selected
0.60	7.90	$\{I_3^{CR} = 0.6797, I_6^{CR} = 0.1311, I_1^{CR} = 0.0818\}$	0.95	c_3, c_6, c_1
0.70	7.68	$\{I_3^{CR} = 0.7250, I_6^{CR} = 0.1297, I_1^{CR} = 0.0759\}$	0.95	c_3, c_6
0.80	7.43	$\{I_3^{CR} = 0.7679, I_6^{CR} = 0.1245, I_1^{CR} = 0.0678\}$	0.95	c_3

Table 5. The influence of σ_2 on the components selected

σ_1	T_r	Optimum Ranking set	σ_2	Components selected
0.70	7.68	$\{I_3^{CR} = 0.7250, I_6^{CR} = 0.1297, I_1^{CR} = 0.0759\}$	0.90	c_3
0.70	7.68	$\{I_3^{CR} = 0.7250, I_6^{CR} = 0.1297, I_1^{CR} = 0.0759\}$	0.95	c_3, c_6
0.70	7.68	$\{I_3^{CR} = 0.7250, I_6^{CR} = 0.1297, I_1^{CR} = 0.0759\}$	1.00	c_3, c_6, c_1

5.4. Epilog

As mentioned earlier, most complex electromechanical systems can be converted into an equivalent series-parallel system. In practice, we validate the effectiveness of multi-spare ordering policy through a multi-component MEMS system. Further, the proposed method can be extended to more application fields, such as substation automation systems [14], and only needs to meet the following three conditions: (1). A complex system can be converted into an equivalent series-parallel system; (2). The reliability function of each component can be known; (3). Spares have the identical lead time distribution function. In practical application [26], a substation automation system is a complex system consisting of seven non-identical components in series-parallel. Such a system topology can also be found commonly in many industrial plants where various control and supervisory modes exist for different redundancy levels. Moreover, the reliability function of each component is known. Next, according to the multi-component selection method and spares ordering policy proposed in this paper, multi-component selection and multi-spare ordering and replacement can be performed to ensure the overall reliable operation of the system.

6. Conclusion

This study proposes a multi-spare ordering policy based on CI for a complex system with multiple continuously degrading components. According to the approach of components selection, we can select components that most need to be replaced within the system. The method of selecting components aids to recognize the bottleneck of the system and prevent the system from unexpected failure. In addition, the proposed multi-spare ordering policy cannot only identify the most needed components for replacement, but also minimize the expected replacement cost during the once system maintenance. A numerical example shows that the components selected is influenced by the lower threshold and upper threshold. In addition, the optimal ordering time is affected by the standard deviation of spare lead-time, shortage cost per unit time and holding cost per unit time. In one word, experimental results meet our expectations and the proposed multi-spare ordering approach is of significance for safety-critical systems such as substation automation system, bridge system, nuclear power plants and aerospace equipment.

Further work can be achieved by relaxing some assumptions. For example, spares are supplied by the identical manufacturer. In practical, spares are supplied by the different manufacturers. In other words, the lead-time of each spare should be different. In addition, it is significative to study a multi-component ordering policy for a mission. That is to say, in a mission, there may be multiple spare ordering and replacement actions.

Acknowledgment

The authors are grateful to the anonymous reviewers and the editor for their critical as well as constructive review of the manuscript. This research is supported by the National Natural Science Foundation of China (Grant No.51575055) and the National Science and Technology Major Project of China (Grant No.2015ZX04001002).

References

1. Ahmad R, Kamaruddin S. An overview of time-based and condition-based maintenance in industrial application. Computers & Industrial Engineering 2012; 63(1): 135-149, https://doi.org/10.1016/j.cie.2012.02.002.
2. Alaswad S, Xiang Y. A review on condition-based maintenance optimization models for stochastically deteriorating system. Reliability Engineering & System Safety 2017; 157: 54-63, https://doi.org/10.1016/j.ress.2016.08.009.
3. Birnbaum Z W. On the importance of different components in a multicomponent system. In multivariate analysis. San Diego: Academic Press, 1969.
4. Cai J, Yin Y B, Zhang L, Chen X. Joint optimization of preventive maintenance and spare parts inventory with appointment policy. Mathematical Problems in Engineering 2017; https://doi.org/10.1155/2017/3493687.
5. Chen X, Xu D, Xiao L. Joint optimization of replacement and spare ordering for critical rotary component based on condition signal to date. Eksploatacja i Niezawodnosc-Maintenance and Reliability 2017; 19(1): 76-85, https://doi.org/10.17531/ein.2017.1.11.
6. Chien Y H, Sheu S H, Zhang Z G, Love E. An extended optimal replacement model of systems subject to shocks. European Journal of Operational Research 2006; 175(1): 399-412, https://doi.org /10.1016/j.ejor.2005.04.042.
7. Chien Y H. Optimal number of minimal repairs before ordering spare for preventive replacement. Applied Mathematical Modelling 2010; 34(11): 3439-3450, https://doi.org/10.1016/j.apm.2010.02.033.
8. de Jonge B, Teunter R, Tinga T. The influence of practical factor on the benefits of condition-based maintenance over time-based maintenance. Reliability Engineering & System Safety 2017; 158: 21-30, https://doi.org/10.1016/j.ress.2016.10.002.
9. Elsayed E A. Reliability engineering. MA: Addison Wesley, 1996.
10. Fussell J B. How to hand-calculate system safety and reliability characteristics. IEEE Transactions on Reliability 1975; 24(3):169-174, https://doi.org / 10.1109/TR.1975.5215142.
11. Godoy D R, Pascual R, Knights P. Critical spare parts ordering decisions using conditional reliability and stochastic lead time. Reliability Engineering & System Safety 2013; 119: 199-206, https://doi.org/10.1016/j.ress.2013.05.026.
12. Grall A, Bérenguer C, Dieulle L. A condition-based maintenance policy for stochastically deteriorating systems. Reliability Engineering & System Safety 2002; 76(2): 167-180, https://doi.org/10.1016/S0951-8320(01)00148-X.
13. Grall A, Dieulle L, Bérenguer C, Roussignol M. Continuous-time predictive-maintenance scheduling for a deteriorating system. IEEE Transactions on Reliability 2002; 51(2): 141-150, https://doi.org/ 10.1109/TR.2002.1011518.
14. Hajian-Hoseinabadi H. Reliability and component importance analysis of substation automation systems. International Journal of Electrical Power & Energy Systems 2013; 49: 455-463, https://doi.org/10.1016/j.ijepes.2010.06.012.
15. Jardine A K S, Tsang A H C. Maintenance, replacement, and reliability: theory and applications. Boca Raton: Taylor & Francis Group, 2013.
16. Kennedy W J, Patterson J W, Fredendall L D. An overview of recent literature on spare parts inventories. International Journal of Production Economics 2002; 76(2): 201-215, https://doi.org/10.1016/S0925-5273(01)00174-8.
17. Kim J, Ahn Y, Yeo H. A comparative study of time-based maintenance and condition-based maintenance for optimal choice of maintenance policy. Structure and Infrastructure Engineering 2016; 12(12): 1525-1536, https://doi.org/10.1080/15732479.2016.1149871.
18. Kuo W, Zuo M J. Optimal reliability modeling: principles and application. New York: John Wiley & Sons, 2003.
19. Kuo W, Zhu X. Importance measures in reliability, risk, and optimization: principles and applications. Chichester: Wiley, 2012.

20. Levitin G, Podofillini L, Zio E. Generalised importance measures for multi-state elements based on performance level restrictions. Reliability Engineering & System Safety 2003; 82(3): 287-298, https://doi.org /10.1016/S0951-8320(03)00171-6.
21. Lin X, Basten R J I, Kranenburg A A, van Houtum G J. Condition based spare parts supply. Reliability Engineering & System Safety 2017; 168: 240-248, https://doi.org/10.1016/j.ress.2017.05.035.
22. Liu B, Liang Z, Parlikad A K, Xie M, Kuo W. Condition-based maintenance for systems with aging and cumulative damage based on proportional hazards model. Reliability Engineering & System Safety 2017; 168: 200-209. https://doi.org/10.1016/j.ress.2017.04.010.
23. Louit D, Pascual R., Banjevic D, Jardine A K S. Condition-based spares ordering for critical components. Mechanical Systems and Signal Processing 2011; 25(5): 1837-1848, https://doi.org/10.1016/j.ymssp.2011.01.004.
24. Meng F C. Some further results on ranking the importance of system components. Reliability Engineering & System Safety 1995; 47(2):97-101, https://doi.org/10.1016/0951-8320(94)00046-Q.
25. Nguyen K-A, Do P, Grall A. Condition-based maintenance for multi-component systems using importance measure and predictive information. International Journal of Systems Science 2014; 1(4): 228-245, https://doi.org/10.1080/23302674.2014.983582.
26. Nguyen K-A, Do P, Grall A. Joint predictive maintenance and inventory strategy for multi-component systems using Birnbaum's structural importance. Reliability Engineering & System Safety 2017; 168: 249-261, https://doi.org/10.1016/j.ress.2017.05.034.
27. Panagiotidou S. Joint optimization of spare parts ordering and maintenance policies for multiple identical items subject to silent failures. European Journal of Operational Research 2014; 235(1): 300-314, https://doi.org/10.1016/j.ejor.2013.10.065.
28. Park Y T, Park K S. Generalized spare ordering polices with random lead time. European Journal of Operational Research 1986; 23(3): 320-330, https://doi.org/10.1016/0377-2217(86)90298-5.
29. Peng H, Feng Q, Coit D W. Simultaneous quality and reliability optimization for microengines subject to degradation. IEEE Transactions on Reliability 2009; 58(1): 98-105, https://doi.org/10.1109/TR.2008.2011672.
30. Peng H, Coit D W, Feng Q. Component reliability criticality or importance measures for systems with degrading components. IEEE Transactions on Reliability 2012; 61(1): 4-12, https://doi.org/10.1109/TR.2011.2182256.
31. Vesely W E. A time-dependent methodology for fault tree evaluation. Nuclear Engineering and Design 1970; 13(2): 337-360, https://doi.org /10.1016/0029-5493(70)90167-6.
32. Vasseur D, Llory M. International survey on PSA figures of merit. Reliability Engineering & System Safety 1999; 66(3): 261-274, https:// doi.org / 10.1016/S0951-8320(99)00011-3.
33. Wang H Z. A survey of maintenance policies of deteriorating systems. European Journal of Operational Research 2002; 139(3): 469-489, https://doi.org/10.1016/S0377-2217(01)00197-7.
34. Wang L, Chu J, Mao W. A condition-based order-replacement policy for a single-unit system. Applied Mathematical Modelling 2008; 32(11): 2274-2289, https://doi.org/10.1016/j.apm.2007.07.016.
35. Wang Z Q, Hu C H, Wang W B, Kong X Y, Zhang W. A prognostics-based spare part ordering and system replacement policy for a deteriorating system subject to a random lead time. International Journal of Production Research 2015; 53(15): 4511-4527, https://doi.org/10.1080/00207543.2014.988892.
36. Wu S, Chen Y, Wu Q, Wang Z. Linking component importance to optimization of preventive maintenance policy. Reliability Engineering & System Safety 2016; 146: 26-32, https://doi.org/10.1016/j.ress.2015.10.008.

Cheng WANG
Jianxin XU
Mailbox No. 633, Northwestern Polytechnical University
West YOUYI Road 127, Xi'an 710072, P. R. China

Hongjun WANG
Beijing Information Science and Technology University
No. 12, Xiaoying East Road, Qinghe, Haidian District, Beijing 100192, P. R. China

Zhenming ZHANG
Mailbox No. 633, Northwestern Polytechnical University
West YOUYI Road 127, Xi'an 710072, P. R. China

E-mails: wangc1217@mail.nwpu.edu.cn, xujx@nwpu.edu.cn, wanghj86@163.com, zhangzm@nwpu.edu.cn

Article citation info:

IDZIASZEK Z. Method of analysis of productivity with an innovative model of the working capability of the object in the body (ℂ) for the new resource allocation on inherent and non-inherent. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 671–681, http://dx.doi.org/10.17531/ein.2018.4.18.

Zdzisław IDZIASZEK

METHOD OF ANALYSIS OF PRODUCTIVITY WITH AN INNOVATIVE MODEL OF THE WORKING CAPABILITY OF THE OBJECT IN THE BODY (ℂ) FOR THE NEW RESOURCE ALLOCATION ON INHERENT AND NON-INHERENT

METODA ANALIZY PRODUKTYWNOŚCI Z INNOWACYJNYM MODELEM POTENCJAŁU ROBOCZEGO OBIEKTU W CIELE ℂ DLA NOWEGO PODZIAŁU ZASOBÓW NA INHERENTNE I NIEINHERENTNE*

The aim of the article is to develop new methods of analysis, estimation and optimal selection of quantitative resources (inherent and non-inherent) in the planning of the product effect for specific environmental conditions. The required iterative approach in the construction of the mathematical model and analysis of its possible practical applications and search for how to figure those opportunities. As the testing method has been applied method intuitive, allowing you to use the experience of expert analysis from ongoing opportunities to make full use of the sustainability properties and customize to their processes. The results were presented in the form of mathematical models in the collection of complex numbers and graphically on the plane of complex numbers. Method to estimate changes inherent and non-inherent resources objects (machines, systems, organizations) on their productivity (P_o). The method uses the original, innovative, model potential workspace object (P_rO) in the form of a complex binding numerically inherent (Z_iO) and non-inherent ($Z_{ni}O$) resources objects. Evaluation of value Po it was proposed with the P_rO. The values of the Z_iO and $Z_{ni}O$ was adopted as two independent resources constituting the whole of resources in the required in the production (or in the service). Method evaluation Po illustrates for the resources object described model $R_o = \left| P_rO = f(Z_pO, Z_oO) \right|$, where Z_iO is a work resource (Z_pO), $Z_{ni}O$ is extracted from the operation of the resource service (Z_oO), and the generating capacity of the object P_o is described using a pointer named R object (R_o). Illustrated in the complex plane analysis results and the results obtained from the calculation P_rO and R_o, for contract values of the Z_pO and Z_oO, indicate the application capabilities developed method. Method allows a very clear description of the productivity changes objects (or processes, or production organization), in the context of the selection of manufacturing resource structure, through the separation of the factors causing these changes. Method can be adapted for optimal production costs (or services) through design changes object and/or design changes of the process exploitation. Developed the method brings new opportunities for theoretical and application in relation technical and economic sciences.

Keywords: *productivity, durability, reliability, operation and using and maintenance, maintenance of machinery, approved limit working time or approved number of working cycles, manufacturing resources, P_rO, Z_iO, $Z_{ni}O$.*

Celem artykułu jest opracowanie nowej metody analizy, szacowania i optymalnego doboru ilościowego zasobów (inherentnych i nieinherentnych) w planowaniu efektu produktowego w określonych warunkach środowiskowych. Realizacja celu wymagała iteracyjnego podejścia przy budowie modelu matematycznego i analizie możliwych jego zastosowań praktycznych oraz poszukiwaniu sposobu ilustracji tych możliwości. Jako metoda badawcza została zastosowana metoda intuicyjna, pozwalająca wykorzystać doświadczenie eksperckie z realizowanych analiz możliwości pełnego wykorzystania trwałości obiektów i dostosowywania do tego ich procesów eksploatacji. Wyniki zostały zaprezentowane w postaci modeli matematycznych w zbiorze liczb zespolonych i graficznie na płaszczyźnie liczb zespolonych. Metoda umożliwia szacowanie zmian inherentnych i nieinherentnych zasobów obiektów (maszyn, systemów, organizacji) na ich produktywność (P_o). W metodzie wykorzystano autorski, innowacyjny, model potencjału roboczego obiektu (P_rO) w postaci liczby zespolonej wiążącej liczbowo inherentne (Z_iO) i nieinherentne ($Z_{ni}O$) zasoby obiektu. Wyznaczanie wartości P_o zaproponowano z modułu P_rO. Wartości Z_iO i $Z_{ni}O$ przyjęto jako dwa niezależne od siebie zasoby stanowiące całość zasobów w realizacji danej produkcji lub usługi. Metodę oceny P_o zilustrowano dla zasobów obiektu opisanych modelem $R_o = \left| P_rO = f(Z_pO, Z_oO) \right|$, gdzie Z_iO to zasób pracy obiektu (Z_pO), $Z_{ni}O$ to wyodrębniony z eksploatacji zasób obsług (Z_oO), a zdolności wytwórcze obiektu P_o opisano za pomocą wskaźnika nazwanego resursem obiektu (R_o). Zilustrowane na płaszczyźnie zespolonej wyniki analiz i uzyskane wyniki z obliczeń P_rO i R_o, dla umownych wartości Z_pO i Z_oO, wskazują na duże możliwości aplikacyjne opracowanej metody. Metoda umożliwia bardzo czytelny opis zmian produktywności obiektów\procesów\organizacji, w kontekście doboru struktury zasobów wytwórczych, poprzez rozdzielenie czynników powodujących te zmiany. Metodę można adaptować na potrzeby optymalizacji kosztów produkcji\usług poprzez zmiany projektowe obiektu technicznego i\lub zmiany projektowe procesu jego eksploatacji. Opracowana metoda wnosi nowe możliwości teoretyczne oraz aplikacyjne w powiązaniu nauk technicznych i ekonomicznych

Słowa kluczowe: *produktywność, trwałość, niezawodność, eksploatacja, utrzymanie w ruchu maszyn, resurs, zasoby wytwórcze, P_rO, Z_iO, $Z_{ni}O$.*

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

1. Introduction

New challenges for the organization of production (high competitiveness and complexity of manufacturing processes) require the modern changes in the management of the working environment [15]. Necessary for this innovative method of collecting, organizing, analyzing and processing data. What counts here most of all the response time to changing conditions internals (staff turnover service quality, adapting to the aging machine park) and external (change consumer expectations, conditions of environmental standards) [10, 35].

Is part of a global trend to improve competitiveness through productivity growth (P_o) [17.21].

This in turn generates demand for modern methods and models to optimize the distribution of resources (material, human, financial, information, management) for the purpose of their efficient use [17, 19, 26, 30]. There are many publications devoted to measuring P_o and describing indicators P_o to the estimate of the production processes, which are explained in more detail in [17]. A lot of scientific studies is also about ways to improve P_o for example the application of philosophy LP [22, 29, 32]. There are many publications in which it moves issues full [18, 22] and secure [3, 6, 7] use of the technical resources and the improvement of the service organization [3, 4, 11, 12, 23], as well as optimizing the distribution of resources [27]. Maintenance of machinery in motion has developed a variety of ways to control the production process and optimization used in material resources, human and financial, eg. TPM, 5S, etc. [24], whether the indicators OEE [1, 14]. They have their use in monitoring the efficiency of a particular production system, but are not so the overall rate as productivity [17].

In the literature, however, was not a method enabling the binding in one mathematical model of productivity (P_o) all the resources [1, 8, 14, 17, 23], proposed divide it by inherent and non-inherent [11], you need to obtain a result production (or service) i.e. in the form of a number or quantity of the product. With increasing complexity and technical excellence object in maintaining proper P_o increasing importance to human resources. Maintenance quality schedule services and operations by the operators [4, 7, 9, 18, 23, 25], requires not only the appropriate management of these resources, but also for resources development customizations objects. Human resources identified are mainly from the non-inherent resources in the production process. In this sense, optimal productivity can be obtained only if the object resources (inherent) will be adjusted accordingly to the existing human resources in your environment (non-inherent).

You can measure the impact of non-inherent resources to the value of the final production, but more important is how to optimize their selection in the context of the inherent to obtain the optimum indicator P_o or, conversely, how human resources (operating, environment) to choose the object resources.

Based on its own expertise and on the results of analysis of the literature and by using intuitive method, it is considered that it will be necessary to divide production resources object and organizations on inherent production resources (related to technical, technological, reliability and durability capabilities of objects) and non-inherent production resources (resulting mainly from human decisions – policy of profitability, organization of work, operating strategies, activities and environmental conditions pro-quality-existing human potential, training, scientific, cultural, etc.). In the literature the author found no such allocation of resources (in the process of exploitation or maintenance of machinery) a mathematical model P_o. There is mentions only (in review the earlier publications), that their proper selection and behavior of the established quality determine the value of the P_o. In these publications, resources inherent and non-inherent, usually are analyzed separately or are not specifically divided. Hence, it was considered that in the evaluation P_o the object should be the method used for evaluation of the simultaneous impact of both of these resource groups i.e. inherent and non-inherent, which are all the resources for a given type of production (services). The initial discussion of this issue is outlined in the work [11].

Proposed in article method entering in a theory of resources of assessment P_o the object/ or organization, which replaces the traditional management, in which dominated the evolutionary approach. First of all, proposed in this work, method allow analysis and selection of optimal assignment of resources inherent resources non-inherent or vice versa (depending on what it is easier to fit) to achieve optimum value P_o the i.e. profit ratio of manufactured products in relation to invested in production funds. This means that you need to know, in which resources and how much you should allocate financial resources to achieve optimum productivity (P_o).

In English literature, there is no explicit descriptions of some terms and symbols (used in Polish literature) necessary for the understanding of the models described in this article. Hence the author introduced their English newly defined (fourteenth) meanings of words or symbols applied in this article. The first is P_o - *productivity* of technical object (or production process or production organization). The second is Z_iO - inherent resources in technical object (or production process or production organization). The third is $Z_{ni}O$ - non-inherent resources in technical object (or production process or production organization); The fourth is P_rO - potential workspace technical object (or production process or production organization) consequential to inherent and non-inherent resources. The fifth is *Object* - the technical object or production process or production organization (item technical, manufacturing plant, production facility, works technical device, assembly line, industrial organization etc.); everything what is producing technical products or products service. The sixth is *Exploitation - (in Polish - eksploatacja* [3, 4, 11]*):* servicing and uses i.e. organized or scheduled, in a rational way, exploiting the inherent potential of the technical object or production process or production organization for adopted criteria e.g. productivity, efficiency, durability, reliability, security, etc. The seventh is *Exploitation of object - operation, using* and *maintenance* the object, diagnostics, operational control and crew training, and continuous airworthiness management having an impact on safety. The eighth is *R - (in Polish - resurs* [3, 4, 11]*):* approved limit working time (or approved number of working cycles) of the technical object (or process or production organization or the number of made products or services) which guarantees the safety and efficiency of operation and support of object (or production process or production organization). The ninth is R_o - production index - *describes estimated* size output (or *production capacity* or *productivity* of object or production process or production organization). The tenth is Z_pO - production quality (with result of what are inherent resources - Z_iO). The eleventh is Z_oO - perfection of operating procedures (with result of what are none-inherent resources - $Z_{ni}O$). The twelfth is uZ_pO - hypothetical value Z_pO, for accepted contractual units – u. The thirteenth is uZ_oO - hypothetical value Z_oO, for accepted contractual units – u. The fourteenth is *c.u.* - contractual units *(in Polish - j.u.)*.

1.1. Allocation of resources used in the P_o

Based on the literature review, at the highest level of general allocation of resources to produce P_o it has been made in the work [9]. These are human resources (skills, knowledge, abilities and suitability of all employees in the enterprise), financial (the financial capital, which the organization uses to fund activities both current and long term), material (in the squad, which includes, among other raw materials, semi-finished products, office space and production and all kinds of equipment) and information (all kinds of useful information for effective decision making). By analyzing the binding capabilities of these resources for optimizing P_o encountered the difficulties arising from the diversity of the ways mathematical description of these resources. In the set of real numbers is not possible a simple bind these types of resources into a single mathematical model. Because these resources don't have mathematical common space, in which indicators could be described in one mathematical relation.

It was recognized that for this aim optimization, resources required is even more general allocation of resources. Therefore, the proposed allocation of resources necessary for the implementation of the object P_o not on four [9], and two independent of each other (however, one piece of resources in the production) resource group i.e. Z_iO (inherent) and $Z_{ni}O$ (non-inherent) resources of the object, as proposed in [11]. In relation to the [11] this article has developed a more general mathematical model P_o. Within the framework of the developed method for the analysis of productivity the object, that method may relate to the broader class of problems in manufacturing i.e.. for any products (machines, services, fuels, financial, human-training of specialists, etc.) allowing to meet all human needs.

To the resources of the inherent lists all material factors related to the technical and technological means of production (machinery, technological lines, software, database, and assigned to their creation of financial resources), and to non-inherent all the factors associated with the use of these factors inherent (human resources, organizational procedures, maintenance procedures of production and assigned to their maintenance of financial resources). Such the divide of resources to allow development of such a model to optimize them to get the not like the biggest effect, but optimal production effect, reference to the costs incurred in the manufacturing environment. One of the objectives of this modeling is the ability to minimize the planned costs for the type of production through selection of appropriate levels of resources inherent and non-inherent.

1.2. Characteristics of inherent (Z_iO) and non-inherent ($Z_{ni}O$) resources object

With the development of the theory of reliability [22, 25, 28, 33], the theory of maintenance objects in motion [6, 13, 19], the theory of exploitation [2, 3, 4, 12, 22, 25, 33, 34] and the development of theory, increasingly began to use the concepts in the form of: item durability [3:12], item operation [4] resource [11], the production potential [11, 123], resource techniques [3], an information resource [23]. PN-EN ISO 9000:2006 (now PN-EN ISO 9000:2015-10) emerged the concept of inherent in the definition of quality as "... the degree to which a set of inherent object ownership meets requirements". On this basis, the author suggested in [11] is a term to describe a work resource object (Z_pO), that in the spirit of the above definition describes the quality of the object. On these resources inherent consists of everything that follows from the inherent factors occurring in the manufacturing process of the goods (products). An example would be owned by the object its potential durability featured R since resources inherent object (R_{io}) [16, 20, 31, 32, 36], which was named in this method, the inherent resources object (Z_iO). You have the maximum value R object ($R_{i(max)o}$) of inherent resources (Z_iO) limits because the production capacity of the object (as well as utilities, or the ability of other tasks [31] for example, combat flight [36], removal of natural disasters, etc.). It can be generalized to the ability to perform the products in a general sense, i.e. both the material and the service (e.g. transportation). But the size of the degree of use of resources inherent (and hence the $R_{i(max)o}$) have a very strong impact your organization (assigned to the object and the resulting from the exploitation strategy [25, 34]) types of and the resources non-inherent [7], that can be described R since resources non-inherent object (R_{nio}) – consequential to assigned resources in exploitation system (facility management). The manner and quality of use of R_{nio} describe (according to the author) mainly non-inherent factors given to and dependent on the so-called the human factor, hence called them non-inherent resources object ($Z_{ni}O$). Selection of $Z_{ni}O$, in the framework of exploitation (machinery maintenance processes [13]), should enable the optimal use of the Z_iO with your object by taking into account optimizing global costs involved in the manufacture of products (described by P_o) for which it was intended. It's hard, and sometimes it is not possible to specify or extract the R_{io} and R_{nio} with R production or service object described here as R_o depending on (1). It is much easier to define the Z_iO and $Z_{ni}O$, and assess their impact on the value of P_rO. Hence the resources inherent and non-inherent create a larger resource that was called a potential workspace technical object – P_rO and described according to (1):

$$R_o = f(R_{io}, R_{nio}); \quad P_rO = f\,(Z_iO,\, Z_{ni}O), \tag{1}$$

where Z_iO can be described a numeric indicator R_{io} ($R_{io(max)}$), and $Z_{ni}O$ can be described a numeric indicator R_{nio} ($R_{nio(max)}$) from here P_o can be described a numeric indicator R_o. Indicator R_o can describe here the numerical value of the products (or services) get in object (or processes or organization) with their inherent and non-inherent resources. Ability to use P_o when depends on the functional value of therefore P_rO described dependency (2):

$$P_o\,(R_o) = f\,(P_rO). \tag{2}$$

It is assumed that the same value P_rO can be achieved as a result of the application of very different configurations of quantitative resources components Z_iO and $Z_{ni}O$. Therefore it is considered advisable to analyses their optimal allocation, in the strategy of exploitation within the limits of resources in the manufacturing environment aimed to optimize profits from executing (founded) the volume of production (services).

It is considered that only the optimum selection of Z_iO to $Z_{ni}O$ for the current exploitation conditions and objectives set production plans (services), enables the optimal use of the existing potential of the working organization or object (P_rO). Determination of optimal value productivity object P_o can be expressed by a numeric indicator R_o (R production technical object or organization service or production product). The results of such analyses can be used for example to adjust too ambitious production plans or justification offset excessive resources to other tasks.

2. Assumptions to the method

Developed method is the result of an intuitive process expertise author, obtained in the analysis of the exploitation processes within the framework of the multiannual research capabilities from strategy to operation according to the preventive work (R) on operation strategy according to the condition for complex technical systems. The development of these (published mainly in materials for business use) concerned the weapons systems of aircraft and helicopter, carried out surveys in terms of causes of damage to the weapon and pursued research reliability-durability aircraft cannons [31]. In addition, imposed on many years of experience in the analysis of reliability-durability objects in lectures, exercises and projects with the subject "Reliability, durability and exploitation of objects". The aim of all these studies was the search for answers to the question of how to model the use of the object in the event the placement in another exploitation system (than has planned for these objects their manufacturer), in order to use the whole work resource object while maintaining the required level of reliability of its activities during the periods of use [11.12]. The main conclusions, which have been obtained from these analyses is that the use of labor resources object depends on the adopted exploitation procedures, and they in turn depend on the assumptions made use of the object. In addition, that way the design objects is strongly dependent on the exploitation conditions to which had hit object. The next important conclusion was that comparing the quality of objects does not make sense without reference to (or comparison) of exploitation assumptions, in which it had to operate. On the basis of this the author came to the conclusion that optimizing the efficiency of the use of the object should be implemented simultaneously with the optimization of the efficiency of operation object. Hence, one step to the allocation of resources involved in productivity factors **inherent** (describing the applied technical solutions) and factors non-inherent (describing the service solutions used in the exploitation process). The consequence of this was to optimize the allocation of productive resources for the object, possible to recruit and assign him a service resources in its exploitation system. Hence, the allocation of resources needed to implement the effect of productive (or service) on the non-inherent and inherent requires indicate which ones belong to the production structures and that to the control structures separate from the exploitation system in which they are maintained. General illustration of this division is shown in Fig. 1 and Fig. 2 and described in section 2.1.

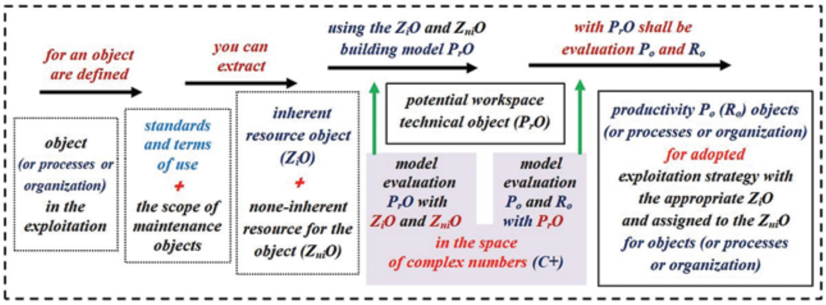


Fig. 1. Illustration of the links to the inherent work resources (Z_iO) and assigned him to non-inherent service resource ($Z_{ni}O$) from its exploitation system of its productivity (P_o) and R object (R_o) and illustration of the location of the proposed model evaluation P_rO with Z_iO and $Z_{ni}O$, and P_o and R_o with the P_rO, with the use of the space complex numbers (C^+)

It is assumed that the proposed article method should be so universal, that will be broadcast to both assess the productivity of a single machine or mechatronic system, as well as the production and the complex organization of production and services. What determines the distribution of the inherent and non-inherent resources and the quality of their use is directly related to the applied exploitation strategy, or in a narrower range of applied strategy of maintaining machinery in motion. Hence the change of conditions affecting the ability to implement, founded the exploitation strategy at the same time affects the quality level to change P_o. In addition, in the changed conditions exploitation often required is a different allocation of the inherent and non-inherent resource object or organization.

Proposed in section 2.1 model to allow analysis of the optimization of the inherent and non-inherent resource object (or manufacturing organizations) in different exploitation conditions for different classes and object size (or organization). The method described is at a very high level of generality. However, it has no restrictions on the amount of at issue constituent elements and types of resources provided to correctly qualify the resources to Z_iO or $Z_{ni}O$ (3):

$$P_rO = f(\Sigma Z_iO, \Sigma Z_{ni}O) \tag{3}$$

or (4):

$$P_rO = \Sigma P_rO_j \tag{4}$$

where *j* is a constituent working potentials i.e. for the same number of production lines.

2.1. Modeling of the relationship Z_iO and $Z_{ni}O$ with P_rO and P_o in the exploitation

In each exploitation system there is a limited resource non-inherent Z_{ni} depends on strong from conditioning environmental (especially educational level and technical culture of human potential – operators and support objects). That it can be fully handled ("well maintained" [22]) must be assigned and maintained [24] the competent $Z_{ni}O$. It determines the ability to use Z_iO in exploitation. For the purposes of the developed in article model P_rO assumes that each object in your exploitation system has a numerically specified value Z_iO and allocated him to the numerical value of $Z_{ni}O$ on the basis of which specifies its numeric value P_rO. With the value of the P_rO is determined for the object value P_o (R_o) (Fig. 1).

In the production process the value P_o it is determined on the basis of your Z_iO for specifically assigned to $Z_{ni}O$, taking into account the limitations of illustrated in Fig.2. In the new exploitation system conditions, it should be verify that P_o adopted in the existing exploitation of the object will not be changed due to the inability to secure appropriate values of $Z_{ni}O$.

A problem which was solved by the construction method, is how to adjust the links to the numeric resource Z_iO and $Z_{ni}O$ with P_rO and P_o through one mathematical relationship, would have been possible to calculate this depending on the numeric value having a meaningful unit of measuring. In Fig. 1 we read that the two models are needed. The first is the model for calculating the P_rO with numeric values Z_iO and $Z_{ni}O$, and the second is the model for calculating the P_o (R_o) with P_rO.

The model for calculating the $P_rO = f(Z_iO, Z_{ni}O)$ has been proposed in section 2.2 of this article. It is assumed that the full use of the Z_iO and allocated him to $Z_{ni}O$ is synonymous with getting the maximum value of the P_rO_{max} users achieve $P_{o\,max}$ (Fig. 3). However, because the other constraints (e.g. the number of repair or indicators reliability, the actual quality of service) the value of P_rO usually is less than the P_rO_{max} (Fig. 2).

Therefore, when you enter the object to a new exploitation system, with the appointment of its new value P_o [5], it should be refer to the initial value of the Z_iO and $Z_{ni}O$, and not only to the adopted, by the previous user (or administrator-using) it in other the conditions, value P_o. As the new exploitation conditions (operating and/or use) used by the manufacturer of the restrictions (on Fig. 2 - I and II type) may be different. For example, type I. increase reliability operation object, and for type II. this for example reduce operating costs, the reference exploitation economy until the more modern design objects.

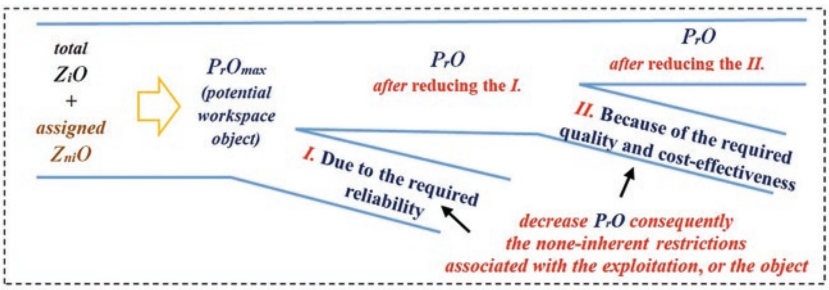


Fig. 2. Illustration of the relationship Z_iO and $Z_{ni}O$ with the potential of the working objects (P_rO) and illustration of the reduction of the P_rO as a result of the admission of non-inherent limitation (type I. i.e. to increase the reliability of the operation object or type II.- reduce operating costs, reference operating economy until the more modern design objects)

In the field of organization management of production (or of the service industry) to recognize factors that affect the processes can be difficult to identify when trying to deal with factors or resources reciprocally conjugated as independent. Hence the proposed in the article model assumes a breakdown of the factors or resources (inherent, non-inherent), which describes the regularity of exploitation and change the state of objects or processes affecting independently from each other on production capacity or service capacity organization. Used for the modeling of the space complex numbers combines these independence without complicated relationships and brings new opportunities in the analysis of their impact on the desired parameter assessment organization (or processes, or object), which represents a new quality in the formulation mathematical models that describe the simultaneous impact of environment and object to the manufacturing facilities capacity of the organization of any type (e.g. production, services). This type of modeling simplifies especially analysis of the actual causes of changes in manufacturing processes, cost and quality in technical systems (inherent - requiring redesign object and process or non-inherent - demanding taking into account the impact of the quality of work, level of culture or technical mentality in the country or the region, the corresponding organization of production and handling). Included in the work of the general model potential workspace technical object (P_rO), after appropriate changes in assumptions can be used to R analysis, cost, performance marketing, and so on, what the author intends to present in the next articles.

2.2. Mathematical model P_rO in a set of complex numbers

The mathematical form of model P_rO presented in the form of equation (5) is illustrated in Fig. 3. The mathematical model the numeric indicator P_rO allows you to link the inherent and non-inherent resources and exploitation system generating a result one numeric value.

$$P_rO = z = a_i + i\,b_j = \left(Z_iO\right)_i + i\left(Z_{ni}O\right)_j \tag{5}$$

where:

- P_rO – the potential workspace technical object
- Z_iO – inherent resource object
- $Z_{ni}O$ – non-inherent resource object
- i – contains in the range $i_{min} \div i_{max}$,
- j – contains in the range $j_{min} \div j_{max}$,
- $i_{min\backslash max}$ – minimum\maximum (limit) value inherent resource object
- $j_{min\backslash max}$ – minimum\maximum (limit) value non-inherent resource object

The potential workspace technical object (P_rO) described of equation (5) is based on the mathematical notation the complex number $z = a_i + ib_i$, where the real part describes the inherent resource object ($a_i = Z_iO$), and the imaginary part of the non-inherent resource object ($b_i = Z_{ni}O$). In Fig.3 the location of the complex numbers that describe the characteristic values P_rO. On it shown in the general case, numerical P_rO described the equation (6):

$$\left(P_rO\right)_{a_i,b_j} = \left(Z_iO\right)_{a_i} + i\left(Z_{ni}O\right)_{b_j} = z_{a_i,b_j} = a_i + i\,b_j, \tag{6}$$

as well as the distinctive position of the complex number P_rO on the complex plane such as:

a) $\left(P_rO\right)_{a_{max},b_{max}}$ – for adopted maximum values: $(Z_iO)_{max}$ and $(Z_{ni}O)_{max}$,

b) $\left(P_rO\right)_{a_{min},b_{min}}$ – for the adopted minimum values: $(Z_iO)_{min}$ and $(Z_{ni}O)_{min}$,

c) $\left(P_rO\right)_{a_{min},b_j}$ – for values: $(Z_iO)_{min}$ and $(Z_{ni}O)_{bj}$,

d) $\left(P_rO\right)_{a_i,b_{min}}$ – for the value of: $(Z_{ni}O)_{min}$ and $(Z_iO)_{ai}$.

With illustration, P_rO (Fig. 3) we conclude that when you change the value of the $b_i = Z_{ni}O$, or change the value of $a_i = Z_iO$ (or change them both at once) changes us value P_rO. This means that each change of exploitation conditions such as the pace of wear (*a* –changes to standards of use), whether the change in quality of service (*ib*) entails changing the value of P_rO, and thus changing the location of the complex number P_rO on the complex plane.

Comparison (inequality) of two P_rO represented by two complex numbers [29] is not possible, because the body of $\mathbb{C}$ (complex numbers) is the body of disordered. The lack of order in $\mathbb{C}$ makes the inequality between complex numbers, such as $z_1 > z_2$ (in our case, the $P_rO_1 > P_rO_2$) do not make sense, unless apply to the real numbers. Although there is a fine for two complex numbers such as (7):

$$a_1 + i\,b_1 \geq a_2 + i\,b_2 \Leftrightarrow a_1 \geq a_2 \quad or \quad a_1 = a_2 \;\; and \;\; b_1 \geq b_2\,, \tag{7}$$

however, it connect it with arithmetic and get a numeric value, that was to make sense of a volumetric units for the whole of the complex number, and not just for its components. This relationship describes how it changes the place of a complex number represented by described by the point on the complex numbers. Based on changes to this point on the plane of complex numbers can be assessed only, that of the basic types of resources you must change, or has changed since the last evaluation the value P_rO.

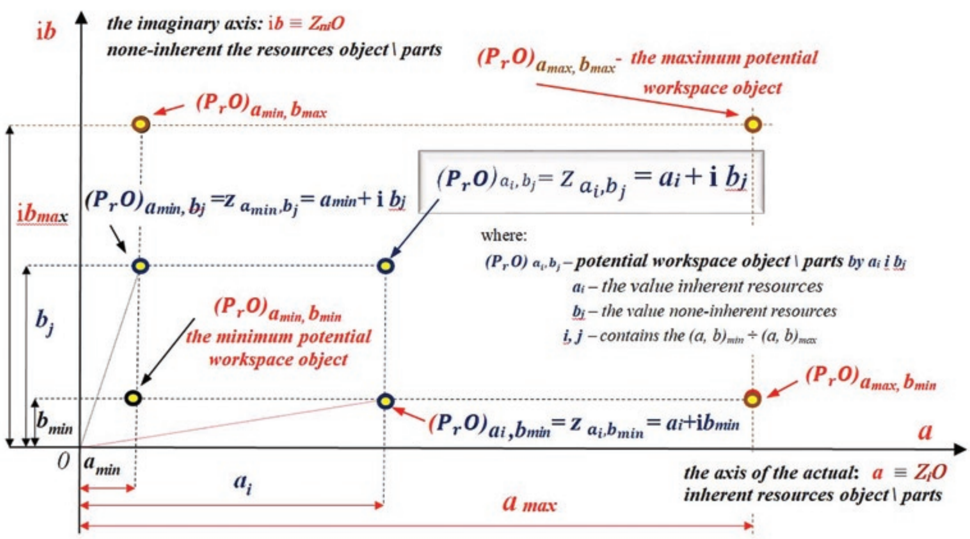


Fig. 3. Model potential workspace object (P_rO) on the I. quarter of the plane of complex numbers (P_rO) a_i,b_j – working potential object obtained on the basis of the adopted a_i,b_j ib $_{max}$ – the maximum value of the resource non-inherent object a $_{max}$ – the maximum value of the resource object inherent

However, the developed model binding P_rO resources Z_iO, $Z_{ni}O$ in the area of complex numbers contains one more, very useful information, in the form of the module number of P_rO.

Intuitively, assume that the value of the module with the number of P_rO can be regarded as axiomatically as the value of the productivity (P_o). After a number of considerations in the search for the inadequacy of such an approach, it was considered that but it has meaning and can be practically implemented. The model relationship P_o with P_rO shows in p. 2.2.

2.3. Mathematical model P_o the set of complex numbers

Although the value of a complex number P_rO does not meet the requirement of arithmetic, but the module P_rO so. This module axiomatically is assigned (in the proposed method) as the value P_o. In accordance with [27] expression $|z_1| > |z_2|$ (in our case $P_rO_1 | > |P_rO_2 |$) it is completely doable, because (8):

$$|Z_1|, |Z_2| \in \mathbb{R}; |P_rO_1| > |P_rO_2| \in \mathbb{R} \tag{8}$$

and the real numbers are the body ordered

Interpretation of geometric module P_rO on the complex plane, is the distance of a point of a complex number (representing the P_rO) from the origin. Hence module, or otherwise the absolute value of the number of $z \in \mathbb{C}$ save as (9):

$$|z| = |a_i + i b_i| = \sqrt{a_i^2 + b_i^2} \quad \text{thus} \quad P_o = |P_rO| = \sqrt{(Z_iO)^2 + (Z_{ni}O)^2}. \tag{9}$$

In Fig. 4 illustrates three cases of specific pairs of values Z_iO, $Z_{ni}O$ for which it was obtained $(P_rO)_{1,1}; (P_rO)_{2,min}; (P_rO)_{min,2}$ giving the same value P_o i.e. $P_{o1,1} = P_{o2,min} = P_{omin,2}$ what has been described the set of equations (10)

$$P_{omin,2} = \left|(P_rO)_{min,2}\right| = \sqrt{x_{min}^2 + y_2^2}; \; P_{o1,1} = \left|(P_rO)_{1,1}\right| = \sqrt{x_1^2 + y_1^2}$$
$$P_{o2,min} = \left|(P_rO)_{2,min}\right| = \sqrt{x_2^2 + y_{min}^2}; \; P_{o1,1} = P_{omin,2} = P_{o2,min}. \tag{10}$$

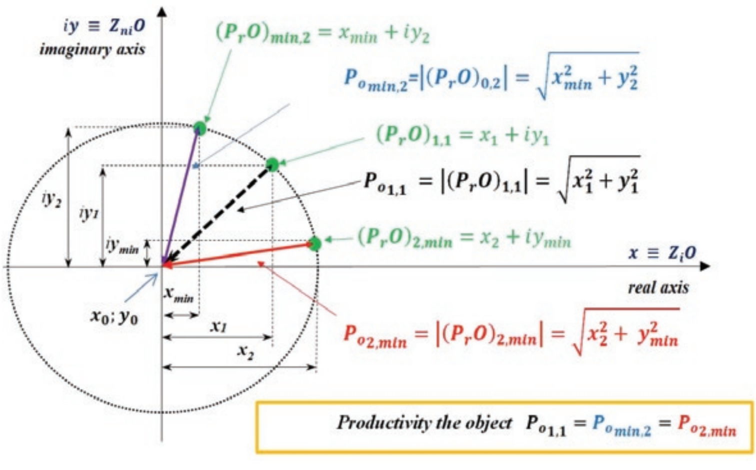


Fig. 4. The model relationship P_o with P_rO illustrated on I. quarter of the plane of complex numbers: $(P_rO)_{1,1}$ – working potential object obtained from the resources of $Z_iO=x_1$ and $Z_{ni}O=y_1$; $(P_rO)_{2,min}$ – potential working object obtained from the resources of the $Z_iO=x_2$ and $Z_{ni}O=y_{min}$; $(P_rO)_{min,2}$ – potential working object obtained from the resources of the $Z_iO=x_{min}$ i $Z_{ni}O=y_2$

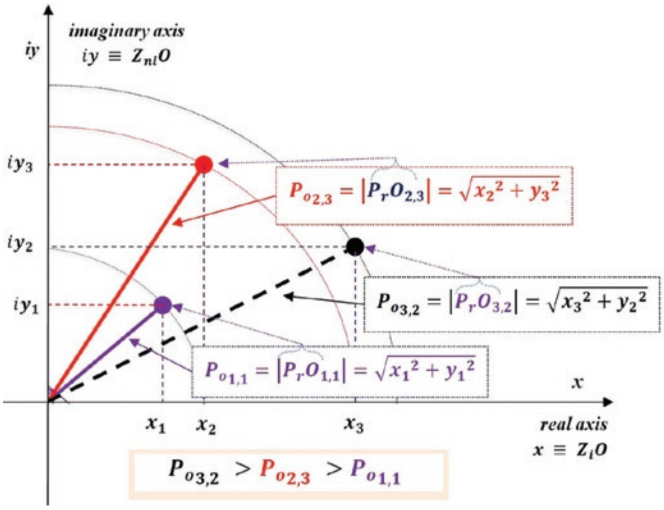


Fig. 5. Illustration of increase P_o by simultaneous improvements in exploitation procedures and service quality (increasing $Z_{ni}O$; $y_1 < y_2 < y_3$) and the improvement of the object or a more effective use of its resource (increasing the Z_iO; $x_1 < x_2 < x_3$)

While the Fig. 5 illustrates the three cases of the characteristic values of the Z_iO, $Z_{ni}O$ from which were obtained $(P_rO)_{1,1}; (P_rO)_{2,3}; (P_rO)_{3,2}$ giving different values P_o. Resource values Z_iO, $Z_{ni}O$ have been selected that obtained increasing the value productivity of the object (P_o) i.e. $P_{o3,2} > P_{o2,3} > P_{o1,1}$ what has been described the set of equations (11):

$$P_{o2,3} = |P_rO_{2,3}| = \sqrt{x_2^2 + y_3^2}; \; P_{o3,2} = |P_rO_{3,2}| = \sqrt{x_3^2 + y_2^2} \tag{11}$$
$$P_{o1,1} = |P_rO_{1,1}| = \sqrt{x_1^2 + y_1^2}; \; P_{o3,2} > P_{o2,3} > P_{o1,1}$$

By analyzing the information that contains the graph in Fig. 4 and Fig. 5, we conclude that we can adjust to changing in time resource values to productivity was fixed (Fig. 4). Simultaneously (Fig. 5) we have a simple illustration of that when, for the same category of objects, but different their perfections and various possibilities of them maintenance resources, their productivity (P_o) is different. Hence by obtaining or having the knowledge of existing or projected changes to the technical level service staff, their technical culture (described in the form of changes in the value of $Z_{ni}O$) and the possible updating of the standards use objects and changes environmental conditions (described in the form of changes in the value of Z_iO) we can preemptively correct our plans for the expected productivity. In conclusion, it must be find that the developed models (P_rO and P_o) should be very useful especially for predictive analytics. Using the conclusions of such analysis, you can make optimal decisions on construction schedules, allocation of resources to the given the production activity. We can also assess possession resources on the system exploitation and possession resources object in the forecast productivity necessary to undertake further production jobs or services.

2.4. Mathematical model of R_o in the collection of com plex numbers

If the value P_o the express using a numeric indicator, called here a R product technical object or production organization (R_o) with their production process (service), the model $P_{o1,1}$ based

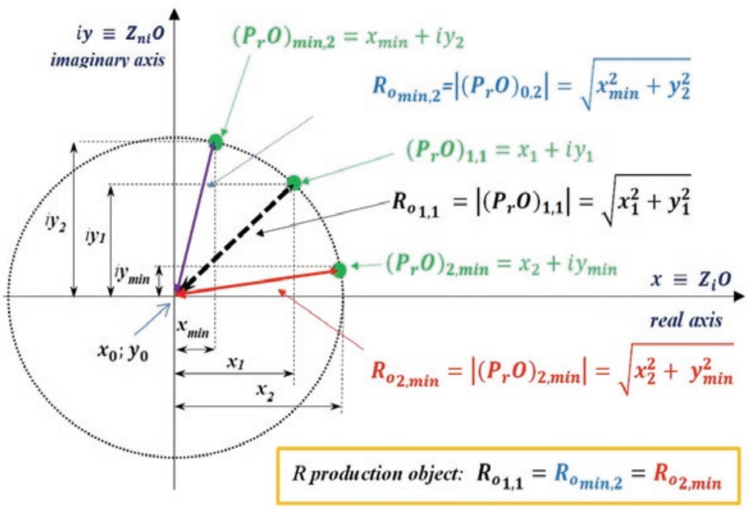


Fig. 6. The model calculated R production object (Ro 1,1) on I. quarter of the plane of complex numbers: (PrO)1,1 – potential working object from the resources of the ZiO=x1 and ZniO=y1; (PrO)2,min – potential working object from the resources of the ZiO=x2 and ZniO=y0min; (PrO)min,2 – potential working object from the resources of the ZiO=xmin and ZniO=y2

on dependencies (9, 10, 11) takes the form of (12). *R* object (R_{a_i,b_j}) is described (12) as a module complex number $(P_rO)_{a_i,b_j}$ (9) whose components are the contractual work resource object $(uZ_pO)_{a_i}$ with the value a_i and contractual service resource object $(uZ_oO)_{b_j}$ with the value b_j. The concept of a „contractual", has all the necessary match the types and the size of these types of resources in the group and detailed models to assign them the appropriate measures numbers. Consider the special case values $uZ_pO = x_1$ and $uZ_oO = y_1$ described a complex number $(P_rO)_{1,1}$ and shown in the Fig. 6. The value of the *R* object $R_{o1,1}$ is module of complex number $(P_rO)_{1,1}$ and is calculated in accordance with the equation (12):

$$R_{a_i,b_j} = \left.\sqrt{(a_i)^2 + (b_j)^2}\right|_{\substack{i=1 \\ j=1}} = R_{o1,1} = \left|(P_rO)_{1,1}\right| = \sqrt{(x_1)^2 + (y_1)^2} \quad (12)$$

where:

$R_{o\,1,1}$ – *R* productive(service) object *for* $(P_rO)_{1,1}$

$(P_rO)_{1,1}$ – working potential object for $uZ_pO = x_1$ and $uZ_0O = y_1$

x_1 – value uZ_pO , y_1 – value uZ_oO , i – is $i_{min} \div i_{max}$, j – is $j_{min} \div j_{max}$.

In Fig. 6 shows the two characteristic of the complex numbers i.e. $(P_rO)_{min,2}$ i $(P_rO)_{2,min}$ and with calculated them the value of R_o:

1) $R_{omin,2}$ for zero values $uZ_pO = x_{min}$ – described the expression (13):

$$(P_rO)_{min,2} = x_{min} + iy_2 \Rightarrow R_{omin,2} = \left|(P_rO)_{min,2}\right|, \quad (13)$$

2) $R_{o2,min}$ for zero values $uZ_oO = y_{min}$ – described the expression (14):

$$(P_rO)_{2,min} = x_2 + iy_{min} \Rightarrow R_{o2,min} = \left|(P_rO)_{2,min}\right|. \quad (14)$$

Marked on the Fig. 6 the two extreme cases can have the following interpretation:

a) when $R_o \cong R_{omin,2}$ (after $R_{min,2} \to x_{min}$) which means that it is a object with a small post-production excellence. Only for very good service or control can it achieve the accepted value of $R_o \cong R_{o2,min}$ (13). This type of resource allocation, we prefer for objects far less time than waiting time for work, a large resource maintenance for that object (e.g. for air cannons).

b) the second case, when the $R_o \cong R_{o2,min}$ (after $R_{o2,min} \to y_{min}$) means that the object is almost maintenance-free, i.e. is so perfect technically, that adopted the values R_o of control slightly and support requires (14). The type we prefer for use in a continuous manufacturing process, where we want to minimize the interruption to the service.

In the analysis of exploitation process illustrate the presented in Fig. 6 enables good visualization of existing joins between adopted the value R_o, production quality objects (Z_pO) and perfection of exploitation procedures (Z_oO). Also gives great opportunities to the theoretical estimation of quality (or numeric) changes P_rO and R_o due to changes in the uZ_oO and/or uZ_pO . Extreme cases R_o (using equations 13 and 14) describes the equations (15) and (16). On the basis of the appropriate selection of contractual value uZ_oO about assuming the y_{min} calculates the value of the R_o resulting from the capabilities and vulnerability of the maintenance object – $R_{o\,min,2}$ (the equation 15):

$$R_{omin,2} = \sqrt{x_{min}^2 + y_2^2} \quad (15)$$

On the basis of the appropriate selection of contractual value uZ_pO about assuming a_{min} calculates the value of the R_o resulting from the **inherent** of work by object $R_{o2,min}$ (the equation 16):

$$R_{o2,min} = \sqrt{x_2^2 + y_{min}^2} \quad (16)$$

Important is that all the objects that have the same value P_rO (or the object for different exploitation conditions with the same value P_rO) have the same value of the module and thus obtain the same value R_o (17).:

$$R_{o1,1} = R_{omin,2} = R_{o2,min} \quad (17)$$

This means that the selection of the value of the R_o object, to the required level, we can shape both by modifying its technical excellence, and by changing the environment to support it in a way that maintains the expected value of the R_o. Dependencies (12, 15, 16) can be applied in practical computer data collection system exploitation management system (machinery maintenance) [10, 29, 32, 33].

2.5. Illustration of the P_rO and R_o the object as a function of the Z_pO and Z_oO on the plane of complex numbers

To illustrate the potential of optimization resulting from the model P_rO and model evaluation it R_o shown in Fig. 7, Fig. 8, Fig. 9 the characteristic three cases how to obtain a specified value R_o by matching value of Z_oO and/or Z_pO :

1) case I. (Fig. 7) - the constant value uZ_oO and three different values of uZ_pO,
2) case II. (Fig. 8) - three different values uZ_oO and constant value uZ_pO,
3) case III. (Fig. 9) - three different values uZ_oO and three different values uZ_pO to ensure that the a fixed value R_o.

Case I. (Fig. 7) To increase the R_o was obtained by increasing the uZ_pO (e.g. improving the technical object and/or more effective use of its Z_pO or/and by changing standards of use [31]) while maintaining a constant value uZ_oO (stability environmental conditions, quality service and control object).

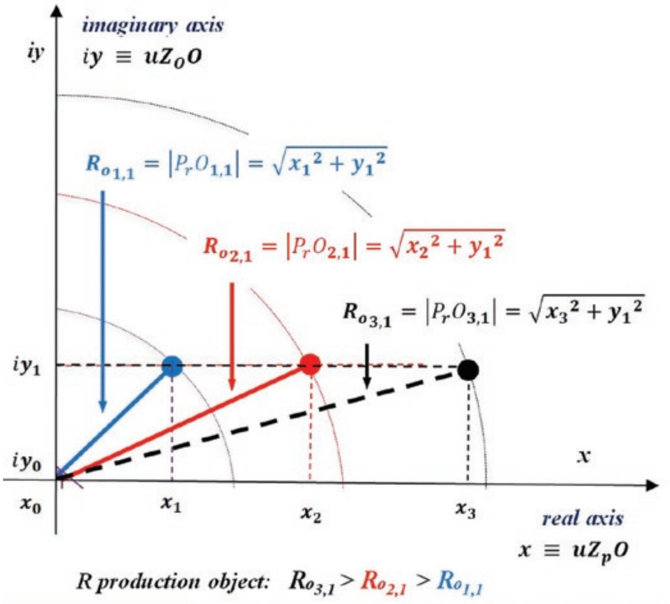


Fig. 7. (Case I.) Illustration values $Ro_{1,1}$, $Ro_{2,1}$, $Ro_{3,1}$ for three values of uZ_pO ($x_1 < x_2 < x_3$) with a constant value uZ_oO amount to y_1

Case II. (Fig. 8) Increase R_o was obtained by improvement exploitation procedures and improve the quality of service (uZ_oO takes the values $+iy_1 < +iy_2 < +iy_3$). By what the technical excellence of the object (uZ_pO) is fixed at is x_1. Points to describe the complex number P_rO flow on a simple in parallel to the imaginary axis.

We can see in Fig. 7, that exploitation procedures and quality of service were adopted here as constants ($uZ_oO = +iy_0$). Increasing the R_o, to assuming the constancy of the characteristics of non-inherent is possible by improving the technical excellence of the object or reduce workloads of the object (uZ_pO takes the values $x_1 < x_2 < x_3$). Points describing the complex number P_rO flow on a simple a parallel to the real axis, and R_o amount to the values $R_{o3,1} > R_{o2,1} > R_{o1,1}$.

Case III. (Fig. 9) Expected constant value R_o. Behavior of the established value of R_o illustrated by choosing three pairs of values uZ_oO and uZ_pO ($y_1, x_3; y_2, x_2; y_3, x_1$). We can see that the decrease in the value uZ_oO of the increase in the value of forces

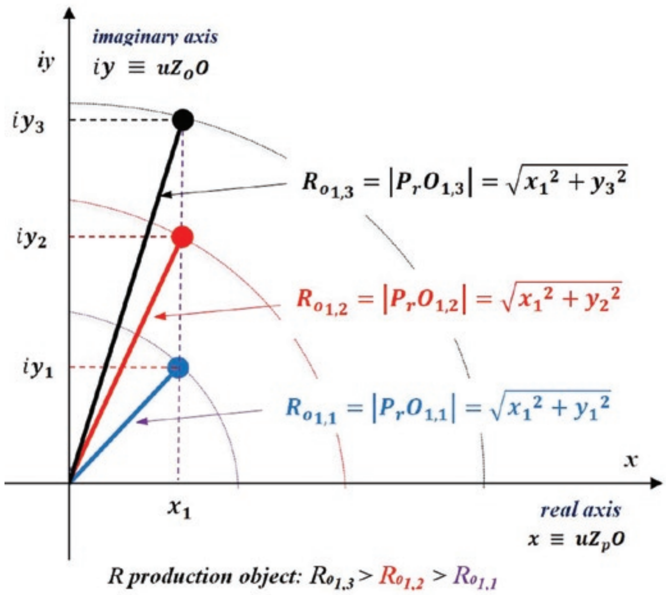


Fig. 8. (Case II.) Illustration resources $Ro_{1,1}$, $Ro_{1,2}$, $Ro_{1,3}$ for three values uZ_oO ($y_1 < y_2 < y_3$) at constant values of the Z_pO amount to x_1. The value of the R_o is growing, because $uZ_oO \rightarrow$ grows by uZ_pO = constants

uZ_pO and vice versa diminished value of uZ_pO forces increase in the value of uZ_oO.

In conclusion (Fig. 7, Fig. 8) it can be concluded that increasing the R_o is possible by increasing the uZ_pO (Fig. 7) and by increasing the uZ_oO (Fig. 8). Of course, it is possible in certain specific limits the possibility of technological and technical object and organizational setting and its exploitation process.

2.6. Summary of the picture value P_rO and P_o (R_o) as a function of changes Z_pO and Z_oO

The main purpose of illustration P_o (R_o) (Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9) was to show how changes to the Z_iO (Z_pO) and $Z_{ni}O$ (Z_oO) affect change the position of the P_rO on the

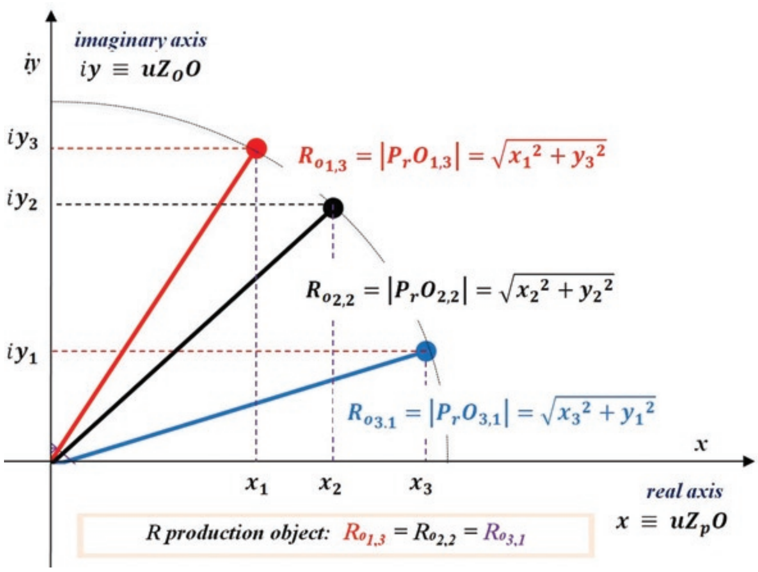


Fig. 9. (Case III.) Illustration of the increase R_o obtained by improving exploitation procedures and improving service quality (increasing uZ_oO; $y_1 < y_2 < y_3$) and improvement of the technical object or the more efficient use of the resource (increasing uZ_pO; $x_1 < x_2 < x_3$).

complex plane, and as shall be determined from the P_rO value $P_o\,(R_o)$. P_rO points equal distant from the origin shall determine the same value $P_o\,(R_o)$ for various combinations of the Z_pO and Z_oO. P_rO, and placed on the line parallel from the imaginary point to the stability of the Z_pO, and placed on a parallel line from the real axis on the stability of the Z_oO. This means that:

1) To preserve the value of $P_o\,(R_o)$ = constant, at changing *a* needs to be changed i*b* and vice versa, when you change the i*b* needs to be changed *a*.
2) To $P_o\,(R_o)$ increased, required is growth of at least one factor *a* or i*b* at non-decreasing the value of the second (but the most adequate evaluation of growth is when the sum of the geometric changes *a* and i*b* is positive).
3) If the value of the i*b* is decreasing at a constant value *a*, a force to be reckoned with reduction in the $P_o\,(R_o)$, which is the ability to:
   a) the increased number of failures,
   b) accelerate the achievement of the limit states,
   c) faster wear of the object.
4) If the exploitation research shows the deterioration of the i*b* is to maintain a constant value $P_o\,(R_o)$ should accordingly increase the value *a* the meaning:
   a) improve the quality/frequency service,
   b) apply the new operating procedures, etc.

2.7. Sample calculation of R_o for hypothetical value uZ_oO and uZ_pO

To illustrate how to use the developed models to determine the P_rO and $P_o\,(R_o)$ for technical objects shows the characteristic example of exploitation analysis (Table 1 and Chart 1). In Table 1. located for seven object, hypothetical value - uZ_pO = (5, 10, 15, 22.5, 30, 35, 40 *c.u.*) each of which is supported on the other system exploitation with numbers 1 to 7, which was allocated for the object values - uZ_oO = (40, 35, 30, 22.5, 15, 10, 5 *c.u.*). For the adopted data calculated P_rO and $P_o\,(R_o)$ for individual objects. Summary of data and calculation results are presented in Table 1 and Chart 1.

From the sample (Table 1 and Chart 1) shows that on the same simple arithmetic sum of units uZ_pO and uZ_oO of 45 *c.u.* value P_o (R_o) calculated from the different P_rO does not have a constant value. Interestingly, minimum of $P_o\,(R_o)$ received by using a balanced level of resources of the Z_pO and Z_oO. This can be explained by that for the average excellence service system and an average of excellence quality of the object, the risk of state failure is the biggest and the largest is the risk of inefficient use of those resources. For objects with low technical excellence (small values of the Z_pO), but in a very good exploitation system (large value Z_oO) is more likely to fully use the resource work (Z_pO). However, for objects with high reliability and durability (a large value of the Z_pO) in a small resource maintenances exploitation (low value Z_oO) we have minimized the impact of the low quality of service tasks. In addition, such systems typically support professionals and highly qualified service and uses them with dedicated diagnostic systems.

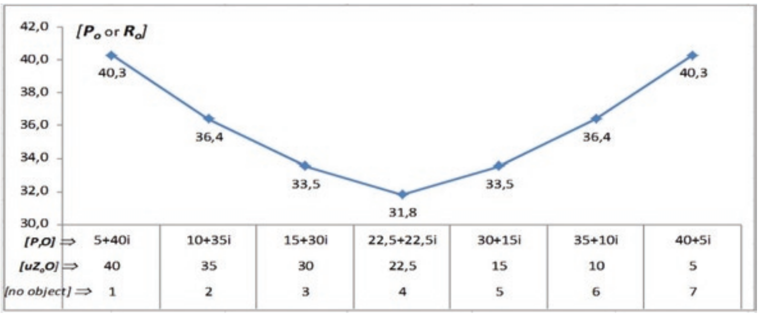
This is what it would seem natural that you should balance the chapter of excellence to object and its control system, it turned out to be by example calculated challenged.

One way to interpret the description of the contractual of the units (*c.u.*) value can be a description of the cost. We assume that the objects 1, 2, 3, 4, 5, 6, 7, pursue the same production. Using the proposed model we can make analysis of this, when it pays us to incur more costs, whether at the stage of the production (or buying) object, whether at the stage of his operating and maintenances of exploitation.

Table 1. Data sheet hypothetical values uZ_oO and uZ_pO and the results of calculations P_rO and R_o

Technical object	uZ_pO (Z_iO)	The type of exploitation system	uZ_oO ($Z_{ni}O$)	P_rO	$P_o\,(R_o)$
[No objects]	[c.u.]	[No systems]	[c.u.]	a complex number	[c.u.]
1	5	1	40	5+40i	**40,3**
2	10	2	35	10+35i	**36,4**
3	15	3	30	15+30i	**33,5**
4	22,5	4	22,5	22,5+22,5i	**31,8**
5	30	5	15	30+15i	**33,5**
6	35	6	10	35+10i	**36,4**
7	40	7	5	40+5i	**40,3**

Chart 1. Contract value $P_o\,(R_o)$ for the object as a function of simultaneous changes uZ_pO = (5; 10; 15; 22.5; 30; 35; 40 c.u.) and uZ_oO = (40; 35; 30; 22.5; 15; 10; 5 c.u.) so that was maintained constant cost value units (45 c.u.) in the number of complex P_rO.



For example, for the manufacture of the first object we had to spend 5 *c.u.* costs. Due to used savings in its manufacture (compared to 7 to produce which we have used 40 *c.u.*) necessary to apply more expensive procedures for its exploitation (in terms of use of his entire *R*) measured at 40 *c.u.* with costs. Contractual units (*c.u.*) can be for example in thousands of dollars. Because P_rO describes the costs is obtained with the chart R_o (Chart 1.) shows us how to choose the cost of production and exploitation for maximum effect R_o. Assuming we have limitations when increasing production quality (or buy a property for example, 1) to get a specific R_o (e.g. 40.3 *c.u.*) we have to incur the costs of the exploitation in terms of 40 *c.u.* costs. It is still only analysis quality but show us that to evaluate productivity and the R_o objects we lead simultaneous analysis of the cost of both production (or buying) objects, that are necessary to bear to the productivity $P_0\,(R_o)$. Of course, the model has entered restrictions to min\max production cost and max\min cost of exploitation for specific objects in specific in conditions of their exploitation.

3. Conclusion

Developed an innovative method of analysis, estimation and optimum amount of resources inherent Z_iO (Z_pO) and non-inherent $Z_{ni}O$ (Z_oO) in evaluation the productivity $P_o\,(R_o)$ object (or economic organization) with the use of the copyright the model potential workspace object (P_rO) in the space of complex numbers (in the body $\mathbb{C}$) brings new opportunities for theoretical and application in association areas technical sciences and economics. In the field of technical sciences can be used for optimal design object for fixed system exploitation or service resource selection to the perfection of the object and the intensity of its use. In the field of economics can be used to optimize productivity for available (or will plan) inherent and non-inherent manufacturing resource in order to of maximizing or optimize profits.

Developed the model P_rO gives you great possibilities of a figurative the conduct of analysis and estimates, using a single numeric indicator global impact of manufacturing resource selection i.e. quality of service/operator ($Z_{ni}O$, Z_oO) and/or quality changes production of objects (Z_iO, Z_pO). While the developed model use values P_rO to evaluation the value $P_o\,(R_o)$ allows easy transition from workspace potential analysis on the assessment of productivity object (or process or organization).

Therefore, the method includes two stages of analysis: possible technological and technical (engineering design object and engineering procedures exploitation) related to the selection of quality objects to exploitation systems (or system maintenance to the object or processes or production organizations), and resources-economic estimation of production capacity and the number of optimization products and profits.

Developed models can also be used (especially after the application of models to data collection and processing system) to monitor changes to the process exploitation and the technical state of object (or process or organization) on the changes of production capacity the test object in your environment .

In presented in the article as the models enable you to estimate P_o (R_o) on the large level of generality is dependent on how the selection of the contractual units (*c.u.*) for Z_iO (Z_pO) and $Z_{ni}O$ (Z_oO). Shown in article to illustrate the usefulness of the proposed models P_rO in estimate values P_o (R_o), for specific cases uZ_pO and uZ_oO, indicates on the application capabilities developed method with the developed models for it P_rO and P_o (R_o). Based on these examples, it can be concluded that the proposed model P_rO allows you to:

- very good to illustrate the impact of changes of the Z_pO and Z_oO to the value of the P_rO,
- simple estimate of the P_rO changes P_o (R_o) the technical object or production processes or production organizations,
- quick assist current production capacity of objects by changes in exploitation conditions and/or changes the condition of an object based on the values P_o (R_o) of objects.

Shown in the final part of the article (using the expert method) diagram for estimating values P_o (R_o) objects for the adopted values of the uZ_pO and uZ_oO gave interesting results. Minimum P_o (R_o) we get at sustainable level the value of the uZ_pO and uZ_oO objects. That is, balancing the financial expenses incurred on the growth of uZ_pO and uZ_oO does not give optimum values P_o (R_o) of the object. The results obtained show that the proposed models P_rO and P_o (R_o) for objects bring a new quality in the rapid assessment of the adequacy of the adopted resources Z_iO (Z_pO) and $Z_{ni}O$ (Z_oO). This knowledge should be particularly useful to those who must change the use of objects or bring object to a different exploitation system than recommended by the manufacturer or want to point out the possible directions of modernization of the exploitation of its objects.

Presented in the article sample interpretations to illustrate only the areas of application and are not ready-to-use models for specific objects. First of all it would be the above analysis linked with the cost of obtaining increased uZ_pO and the costs of obtaining increased uZ_oO. The results of such analysis may allow a choice between the improvement of the object and the improvement of the system's maintenance and help in determining the optimum boundaries of these changes.

As the developed models may be particularly useful when you automate the process of calculating the aggregated values of uZ_pO and uZ_oO, and with them the summed values P_o (R_o).

The same P_rO model can be useful also as an indicator of the efficiency of production, and in the monitoring of the level of stability of the process exploitation and maintenance of objects.

References

1. Bartkowiak T, Gessner A. Modeling performance of a production line and optimizing its efficiency by means of genetic algorithm. Paper presented at the ASME 2014 12th Biennial Conference on Engineering Systems Design and Analysis, ESDA 2014; 3, https://doi.org/10.1115/ESDA2014-20141.
2 Dhillon B S. Maintainability, Maintenance, and Reliability for Engineers. Boca Raton, London, New York: Taylor & Francis Group, 2006, https://doi.org/10.1201/9781420006780.
3. Downarowicz O. System eksploatacji. Zarządzanie zasobami techniki. Politechnika Gdańska: Wydawnictwo Instytutu Technologii Eksploatacji, 1997.
4. Dwiliński L. Podstawy eksploatacji obiektu technicznego. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej, 2006.
5. EUR-Lex. Title and reference. Official Journal of the European Union June 2014; L 173 Volume 57: 12.
6. Fuqing Y, Barabadi A, Jinmei L. Reliability modelling on two-dimensional life data using bivariate weibull distribution: with case study of truck in mines. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (4): 650–659, https://doi.org/10.17531/ein.2017.4.20.
7. Gabryelewicz I, Sadłowska-Wrzesińska J. Tendencje zmian i rola czynnika ludzkiego w systemach zarządzania przedsiębiorstwem. Częstochowa: Wydawnictwo Wydziału Inżynierii Produkcji i Technologii Materiałów Politechniki Częstochowskiej, 2014.
8. Gapp R, Fisher R, Kobayashi K. Implementing 5s within a Japanese context: An integrated management system. Management Decision 2008; 46(4), 565-579, https://doi.org/10.1108/00251740810865067.
9. Griffin R W. Podstawy zarządzania organizacjami. Warszawa: Wydawnictwo Naukowe PWN, 2017.
10. Herterich M, Uebernickel F, Brenner W. The Impact of Cyber-Physical Systems on Industrial Services in Manufacturing. Procedia CIRP 2015; 30: 323-328, https://doi.org/10.1016/j.procir.2015.02.110.
11. Idziaszek Z, Grzesik N. Object characteristics deterioration effect on task realizability – outline method of estimation and prognosis. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 433-440.
12. Idziaszek Z, Olearczuk E. Alokacja obsług i zasobów w zarządzaniu trwałością obiektów technicznych. Zagadnienia Eksploatacji Maszyn 2005; Z.1: 115-124.
13. Jasiulewicz - Kaczmarek M. Współczesne koncepcje utrzymania ruchu infrastruktury technologicznej przedsiębiorstwa. Wydawnictwo Instytutu Inżynierii Zarządzania Politechniki Poznańskiej 2005: 127–134.
14. Jasiulewicz-Kaczmarek M, Bartkowiak T. Improving the performance of a filling line based on simulation. Paper presented at the IOP Conference Series: Materials Science and Engineering 2016; 145 (4), https://doi.org/10.1088/1757-899X/145/4/042024.
15. Jasiulewicz-Kaczmarek M, Drożyner P. Preventive and Pro-Active Ergonomics Influence on Maintenance Excellence Level, [in.] M.M. Robertson (eds.) Ergonomics and Health Aspects, HCII 2011, LNCS 6779 Springer-Verlag Berlin Heidelberg, ISBN 978-3-642-21715-9 2011; 49-58.
16. Kałmucki W S. Prognozirowanije resursov detalej maszin i elementov konstrukcji. Kisziniev:1989.

17. Kosieradzka A. Zarządzanie produktywnością w przedsiębiorstwie. Warszawa: Wydawnictwo C.H. Beck, 2012.
18. Koźmiński A, Jemielniak D, Latusek-Jurczak D. Zasady zarządzania. Warszawa: Wolters Kluwer Polska, 2014.
19. Legutko S. Trendy rozwoju utrzymania ruchu urządzeń i maszyn. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2009; 2(42): 8-16.
20. Liu T, Cheng L, Pan Z, Sun Q. Cycle life prediction of Lithium-ion cells under complex temperature profiles. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (1): 25–31, https://doi.org/10.17531/ein.2016.1.4.
21. Martinez-Jurado P J, Moyano-Fuentes J, Jerez-Gomez P. Human resource management in Lean Production adoption and implementation processes: Success factors in the aeronautics industry. BRQ Business Research Quarterly 2014; 17 (1): 47-68, https://doi.org/10.1016/j.cede.2013.06.004.
22. Moubray J. et al. Maintenance management-a new paradigm. Maintenance1996; 11: 1.
23. Mroczko F, Stańkowska M. Informacja jako kluczowy zasób współczesnych organizacji, [w:] Borowiecki R., Czekaj J. (red.), Zarządzanie zasobami informacyjnymi w warunkach nowej gospodarki. Warszawa: Difin, 2010.
24. Nakajima S. Introduction to TPM. Portland: Productivity Press, 1988.
25. Piasecki S. Elementy teorii niezawodności i eksploatacji obiektów o elementach wielostanowych. Warszawa: Polska Akademia Nauk – Instytut Badan Systemowych, 1995.
26. Prokopenko J, North K. Productivity and Quality Management: A modular programme. International Labour Office,1996.
27. Ronald Leandro Elizondo et al. Beyond Productivity and Continuous Improvement: Fundamentals required for Lean Complex transformation. IFAC-PapersOnLine 2016; 49-12: 467–472.
28. Shewhart W A, Wilks S S. System Reliability theory. New Jersey: John Wiley & Sons, 2004.
29. Stewart I, Tall D. Podstawy matematyki. Warszawa: Prószyński Media Sp. z.o.o., 2017.
30. Taylor A, Taylor M, McSweeney A. Towards greater understanding of success and survival of lean systems. International Journal of Production Research 2013; 51 (22): 6607-6630, https://doi.org/10.1080/00207543.2013.825382.
31. Tomaszek H, Idziaszek Z. Zarys metody oceny trwałości luf działek lotniczych. Zagadnienia Eksploatacji Maszyn 2004; Z.1: 99-110.
32. Trzeszczyński J, Gawron P, Murzynowski W. Wytyczne przedłużania eksploatacji zmodernizowanych bloków 100 MW – 360 MW. Biuletyn Pro Novum 2016; nr 2: 792-799.
33. Wang H, Pham H. Reliability and Optimal Maintenance, London: Springer-Verlag, 2006.
34. Woropay'a M. Podstawy racjonalnej eksploatacji maszyn. Bydgoszcz: Wydawnictwo Akademii Techniczno-Rolniczej, 1996.
35. Zio E. Reliability engineering: Old problems and new challenges Reliability Engineering & System Safety 2009; 94:125-41, https://doi.org/10.1016/j.ress.2008.06.002.
36. Żurek J. Żywotność śmigłowców. Warszawa: Instytut Techniczny Wojsk Lotniczych, 2006.

Zdzisław IDZIASZEK
Mechatronics and Aviation Faculty
Jaroslaw Dabrowski Military University of Technology
ul. Gen. Witolda Urbanowicza 2, 00-908 Warszawa 46, Poland
E–mail: zdzislaw.idziaszek@wat.edu.pl

Article citation info:
HRYCIÓW Z, KRASOŃ W, WYSOCKI J. The experimental tests on the friction coefficient between the leaves of the multi-leaf spring considering a condition of the friction surfaces. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 682–688, http://dx.doi.org/10.17531/ein.2018.4.19.

Zdzisław HRYCIÓW
Wiesław KRASOŃ
Józef WYSOCKI

THE EXPERIMENTAL TESTS ON THE FRICTION COEFFICIENT BETWEEN THE LEAVES OF THE MULTI-LEAF SPRING CONSIDERING A CONDITION OF THE FRICTION SURFACES

BADANIA EKSPERYMENTALNE WSPÓŁCZYNNIKA TARCIA POMIĘDZY PIÓRAMI RESORU WIELOPIÓROWEGO Z UWZGLĘDNIENIEM STANU POWIERZCHNI CIERNYCH*

In this study are presented the results of the simulation tests of the friction pairs occurring between the spring leaves while considering a condition of the mating surfaces and an impact of the velocity of their mutual dislocation on the values of the friction coefficients. It has been proposed a methodology in respect of a determination of the coefficients of the static and kinetic friction. Two kinds of the specimens have been prepared for the tests, which have been cut out from a spring leaf of the prototype spring – they have created the model friction pairs. The condition of the specimen surface and their selected mechanical properties have been evaluated. During the experimental tests have been considered: four sliding velocities, four variants of the surface conditions and two values of the normal load. The tests of the friction pairs have been performed at the laboratory stand for measuring the friction force. The results of the tests have been presented in a form of the time courses of friction force, graphs and tabular summaries of the friction coefficients. It has been conducted a comparative analysis of the results in order to determine an influence of the test results on the values of the determined friction coefficients. The proposed research conditions are approximate to the typical operating conditions of the road vehicles.

Keywords: *multi-leaf spring, laboratory tests, friction pairs, surface roughness, friction forces, static and kinetic friction coefficients.*

W pracy przedstawiono wyniki badań symulacyjnych węzłów tarcia występujących pomiędzy piórami resoru, z uwzględnieniem stanu powierzchni współpracujących oraz wpływu prędkości ich wzajemnego przemieszczania, na wartości współczynników tarcia. Zaproponowano metodykę wyznaczania współczynników statycznych i kinetycznych tarcia. Do badań przygotowano dwa rodzaje próbek, które wycięto z pióra resoru prototypowego - tworzyły one modelowe pary cierne. Oceniono stan powierzchni próbek i wybrane właściwości mechaniczne. W badaniach eksperymentalnych uwzględniono: cztery prędkości poślizgu, cztery warianty stanu powierzchni oraz dwie wartości obciążenia normalnego. Badania par ciernych wykonano na stanowisku laboratoryjnym do pomiaru siły tarcia. Wyniki badań przedstawiono w postaci przebiegów czasowych siły tarcia, wykresów i zestawień tabelarycznych współczynników tarcia. Wykonano analizę porównawczą wyników w celu określenia wpływu warunków badań, na wartości wyznaczonych współczynników tarcia. Zaproponowane warunki badań są zbliżone do typowych warunków eksploatacyjnych pojazdów drogowych.

Słowa kluczowe: *resor wielopiórowy, badania laboratoryjne, pary cierne, chropowatość powierzchni, siły tarcia, statyczny i kinetyczny współczynnik tarcia.*

1. Introduction

The spring is an elastic element of the suspension, which is responsible for transferring forces and torques between the vehicle wheels and its frame, reduces the impacts developed by the reactions of the ground that produce the effects on the riding vehicle and has an essential impact on the movement stabilisation and the ride comfort. The researches evidence that the friction in the spring significantly affects a dynamic stiffness and vehicle vibrations. As a result of the relative movement of the spring leaves in the spring is developed the friction, which may contribute to instability of the vehicle movement and the unfavourable increase of the inelastic friction in the suspension. It is noticeable, in particular, on the roads that feature the good road surfaces. The dynamic forces, generated by the road irregularities, can be too small to overcome the friction forces between the spring leaves. Then it is observed a phenomenon of the "spring interlock", which is disabled from the suspension operation, and the developed forces are transmitted from the road irregularities directly to from a wheel to the vehicle body. It leads to a deterioration of the ride comfort. In the study [14] have been researched the dynamic properties of the stiffness of the multi-leaf spring in terms of designing the structure with a particular attention to the ride comfort. The spring intended for a light-duty truck has constituted an object of the studies. A dependency between the friction, frequency and vibrations amplitude on the dynamic stiffness of the spring has been determined on the basis of

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

the experimental tests and the numerous numerical tests while using the finite elements method (FEM). The results of the conducted test have enabled to optimise a structure of the vehicle suspension in terms of the proper selection of the elastic and damping characteristics. As a result, the ride comfort has been significantly improved. According to [14] it has been found that the dynamic stiffness of the leaf spring changes as the frequency, amplitude and energy dissipated due to the friction are changed. In the study it has been evidenced that a decrease of the friction between the spring leaves of the spring is the most effective way to improve the dynamic stiffness of the multi-leaf spring. An operating status of the spring leaf surfaces, their susceptibility to wear is a subject of the researches presented in the study [8]. In the article it has been studied an influence of the residual stresses on the durability of the structural elements, while forecasting its wear. The residual stresses usually occur as a result of the surface processing, such as shot blasting or rolling. An objective of the experimental tests has consisted in determining an impact of friction and wear characteristic resulting from an interaction of the residual stresses for the sliding surfaces, in dry condition and a measurement of the interphase friction. The specimens for the tests are made from the SUP9 material intended for manufacturing the leaf springs. The residual stresses have been performed on the spring leaf surface due to shot blasting. The profiles of the residual stresses have been measured on the surface and under the surface using a method of X-ray diffraction. The sliding tests have been carried out under a different contact pressure for a specified sliding velocity of 0.035 m/s in order to determine the friction and wear characteristics of the leaf surfaces. The laboratory tests of the leaf springs have been performed on the strength testing machine. It has been obtained the load - displacement curves in a form of a hysteresis loop, which have constituted a basis for a determination of a friction force and a friction coefficient between the spring leaves. The friction coefficients, wear volumes and wear velocities for two conditions of the shot blasted surfaces and specimens, which have not been a subject of shot blasting. In this way, an influence of the residual stress on the tribological characteristic have been evaluated.

In the article [2] one has been proposed to evaluate a total force generated by the leaf spring as a composition of an elastic force and a dry friction force. It has been assumed a spring model, in which an influence of a velocity change of the relative values for the cutting forces of the spring leaves has not been considered. A hysteresis in this model is strongly non-linear. It has been demonstrated that an analysed model can be linked by a discretisation and a change of the parameters with a district time-varying model, which was proposed in 1980 by a team headed by Fancher. For the purposes of mapping a non-linear behaviour of the leaf spring, in the study [2] one has attempted to apply a classical spring model with a linear elastic force and a linear viscous friction force. It has been discussed a process of cross-validation these two non-linear models while considering a member of the elastic and dumping forces, originating from the dry friction and viscous friction. The non-linear model of the spring has been also tested during the experiments in the real heavy goods vehicle.

In the literature [4, 5, 9, 11] can be found the results of a computer simulation with a usage of the finite elements method for the different versions of the spring: beginning from a simple three-leaf spring and ending with the multi-object models [4, 9, 10]. This diversification makes that they are used at the different stages of the vehicle construction (modernization) process for an analysis of the physical phenomena that take place in the chosen elements, parts and assemblies and of the entire vehicle. They represent a different level of complexity and accuracy thus a labour input when developing them and required computational time

Article [11], in which the authors focus on the “multibody” approach and the finite elements method (FEM), in order to find a compromise between a computational accuracy and a computational velocity deserves a special attention. A research area includes an intended use of the vehicle, an evaluation of its operational dynamics within a scope of its handling and the ride comfort. In this article, an interesting variant of the multibody numerical model, as the multi-object model built in the LMS Virtual. Lab Motion [11] has been proposed. Each spring leaf of the spring is discretised in a form of the rigid bodies connected with the linear elastic beams. A contact is modelled as a simplified interaction between the balls, which has this advantage that is computationally efficient and it is suitable for this type of the applications, in which the contact areas are known a priori (leaf ends [5]).

Due to the wide use of the multi-leaf springs in the suspensions of the dependent suspension of the heavy-duty vehicles and trailers, a problem of their operational wear becomes very significant for the technical support facilities of the enterprises.

One of many reasons of the intensive wear of the leaf springs is the servicing negligence. In fig.1 is shown a typical case of the spring wear due to the corrosion (areas marked with the ellipses). The external environmental effects (water, dust, etc.) on the elastic elements causes their intensive operational wear. The generated corrosion products fill the apertures creating a distinctive notch between the mating spring leaves. The conditions of their mating are a subject of the change, and, in particular, the spring leaf curvatures and occur the further clearances between the spring leaves (change of the friction conditions) [3]. As a consequence of the variable dynamic loads, interacting with the vehicle suspension, fatigue strength of the spring changes radically and is durability decreases.

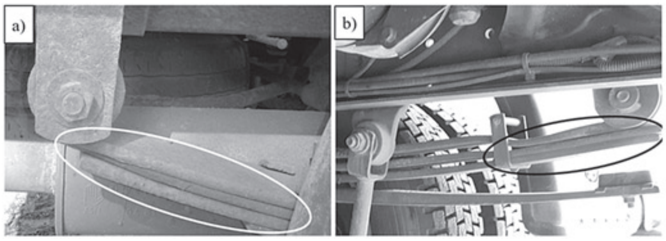


Fig. 1. The examples of the improper operational service of the elastic elements in the rear suspension of the vehicle: a) multi-leaf spring, b) double parabolic spring

The friction between the spring leaves has a great impact on the parameters of the vehicle suspension, and therefore on its operational characteristics. A possibility of considering the friction in the designing and diagnosing methods in respect of the spring is necessary to determine an operating condition of the suspension and the complete vehicle. The modern designing of the vehicle suspension requires a precise description of the friction phenomenon and an appropriate definition of the friction in a form of the different class of the models [4, 5, 9, 11]. For the purposes of the detailed defining of these models that map the friction phenomenon and changes of its interaction, it is necessary to know the static and kinematic friction coefficients. An answer to the question about an influence of the velocity of the mating leaf surfaces and an influence of the surface condition of the spring leaves on the values of the static and kinetic friction coefficients has not been found in the analyses papers, either. Due to the different friction cases, which occur between the spring leaves, an attempt to develop a complementary, consistent and universal model is a complex research problem.

In the study has been described a methodology of determining the friction coefficients in the research conditions similar to the typical operational conditions of the heavy-duty vehicles equipped with the metal multi-leaf springs. It has been conducted the comparative analysis of the simulation results and it has been determined an influence on the test conditions on the values of the determined friction coefficients.

2. Characteristic of the Test Object

2.1. Preparation of the Specimens and a Measurement of their Mechanical Properties

A basic objective of the experimental tests has consisted in the determination of the static and kinetic coefficients of the friction between the spring leaves for the typical for this element: structure condition of the surface layer and the chosen intermediate layers between the mating spring leaves. Therefore, the fragments of the spring leaf of the multi-leaf spring, intended for the rear suspension of the heavy-duty vehicle featuring the GVWR of 12 t have been used for the tests. In the programme of the laboratory tests have been foreseen two sets of the friction pairs comprising a specimen and a counter-specimen, which have been characterised by the different conditions of the friction surfaces. Four specimens have been cut out from the spring leaf:

- two specimens with the dimensions: 12 x 80 x 58 [mm];
- two counter specimens with the dimensions: 12 x 80 x 145 [mm].

Within the frames of the preparatory works in one set (friction pair) one has narrowed down only to its cleaning from the dust while remaining its surface layer in the existing condition (surface covered with a rust layer), but in the second friction pair, one has cleaned the surface layers, on which rust, pitting, cracks, etc. by grinding and smoothing. Then the friction pairs prepared like this have been used during the laboratory tests. The steel ingredients have been specified on the basis of the metallurgical certificate and it has been compared with PN-74/H-84032 standard (table 1).

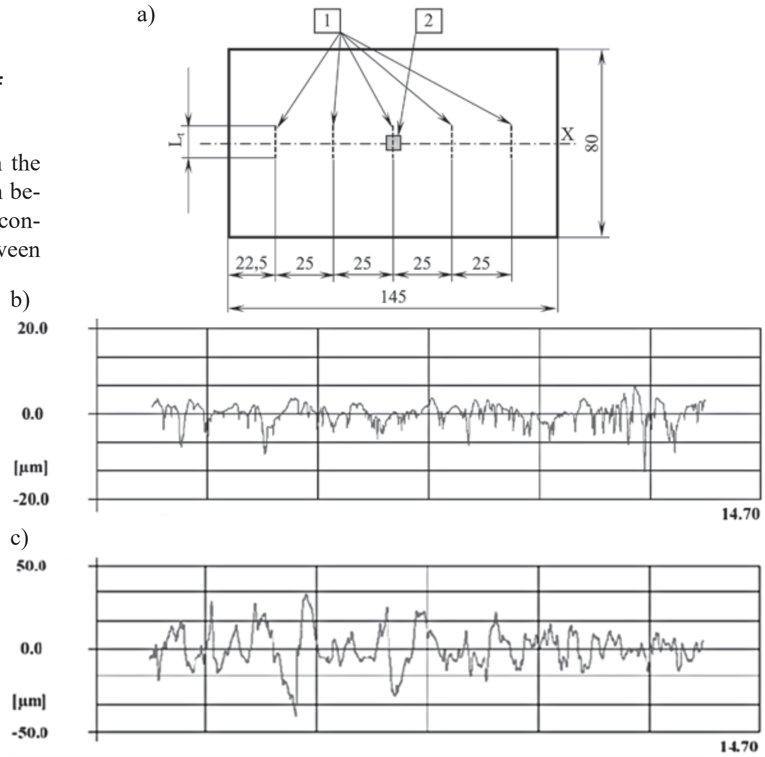


Fig. 2. Profilograms performed for the spring leaves: a) sites where the measurements have been carried out: 1 – location of the measurement sections, 2 – area, for which the 3D image of the micro-surface has been made, b) cleaned specimen, c) corroded specimen

Table 1. Chemical composition of steel 50HSA

Denotation		Proportion of elements in chemical composition [%]							
		C	Mn	Si	P	S	Cr	Ni	Cu
50HSA (by PN-74/H-84032)	min	0,45	0,3	0,8	-	-	0,9	-	-
	max	0,55	0,6	1,2	0,03	0,03	1,2	0,4	0,25
Test sample		0,46	0,38	0,88	0,013	0,01	0,93	0,19	0,15

While analysing the results contained in the table 1 it should be stated that the tested 50HSA spring steel meets, in terms of its chemical composition, the requirements comprised in the standard.

The mechanical properties of the steel have been evaluated using the specimens cut out from the spring leaf. The tests have been carried out pursuant to the PN-04310 standard on the strength testing machine, on which an elongation A_5, the percentage reduction of area Z and Young's modulus E of the steel have been determined. Furthermore, it has been determined hardness of the specimens with the Brinell hardness tester. The results of these tests are presented in the table 2.

Table 2. Selected mechanical properties of 50HSA steel

Denotation	R_m [MPa]	A_5 [%]	Z [%]	HB	E [MPa]
Value	1086	18,1	27,4	320	$2{,}03 \cdot 10^5$

A decrease in a tensile strength limit (R_m) by approx. 18% in respect of the value specified in the PN-74/H-84032 standard is worth noticing.

2.2. Evaluation of the Surface Condition of the Specimens

An evaluation of the surface condition of the friction pairs elements has been performed on the basis of a roughness measurement [6, 7, 13] at the laboratory stand equipped with a measurement instrument of the HOMMELL TESTER T1000 type. With regard to each specimen there have been performed the measurements have been carried out in the transverse direction in respect of a longitudinal axis x (fig. 2a) of the spring leaf (along its width) in the three (for the specimen) or five (for the counter-specimen) sites along its length.

A length of the measurement section has been assumed as L_t = 14.7 mm with a corresponding elementary section of L_c = 2.5 mm. A measuring head has moved with a velocity of 0.5 mm/s. The exemplary results of the measurements obtained for the specimens with two surface conditions are presented in a form of the profilograms (fig. 2b and c), performed in the regions of the geometrical centres of the specimen surfaces (intersection point of the longitudinal and horizontal axis of the specimen).

As a complement to the profilograms showed in fig. 2, table 3, there have been presented the basic parameters of the irregularities for the friction pair in the corroded and cleaned condition. They have been determined while using a smoothing filter compliant with the DIN 4777 standard.

The results presented in the table, constitute an arithmetic mean from five measurements. While analysing the results, contained in the table 3, it should be determined, that the maximum roughness heights (R_{max}), maximum ordinate values of the profile (R_z) and a total height of the profile (R_t) for the cleaned specimen are approx. one third (approx. 33%) of the values determined for the corroded specimen.

The selected results of the tests are shown in fig. 3. They constitute the 3D images of the tested micro-surfaces with a linear magnification 100x.

On these images, for the corroded specimen, a surface geometry is expressly shown. There are visible the sites where a layer of iron

Table 3. Selected parameters of surface roughness for the friction pair

Parameter [µm]	Specimen	
	corroded	cleaned
R_{max}	65,88	21,96
R_z	50,80	14,90
R_a	8,74	1,89
R_t	66,58	22,12
W_a	10,50	1,64

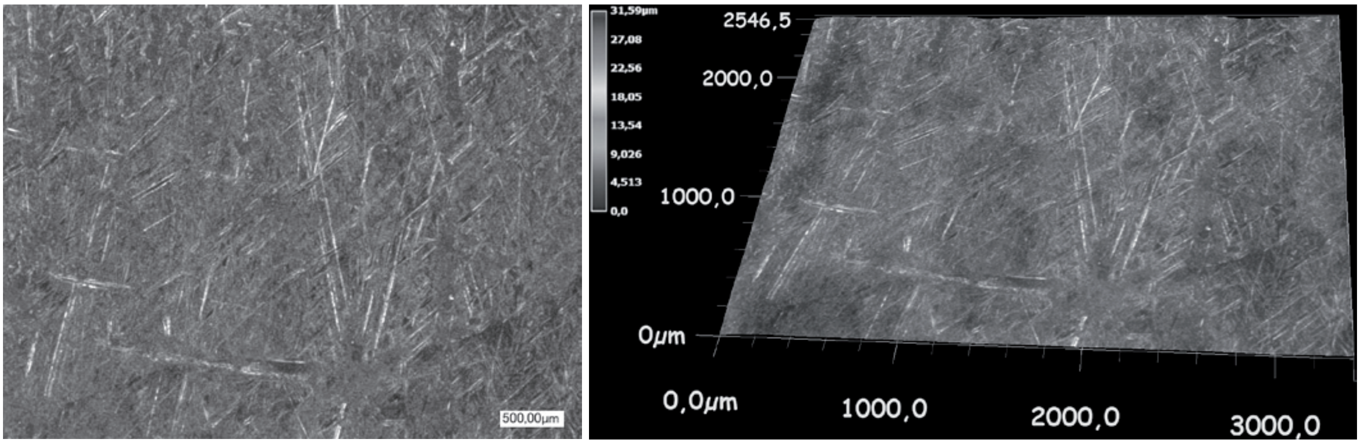
It is worth noticing a change in a waviness profile (W_a), which has decreased with regard to the cleaned specimen approx. 6.5 times in comparison with the corroded specimen (table 3). Moreover, the 3D images of the friction surfaces (fig. 2a, marking 2) have been made with a use of the optical microscope in the sites where roughness has been measured. For this purpose, the laboratory stand equipped with the measuring instrument of the KEYENCE VHX-1000 type has been used.

3. Experimental tests of the friction coefficient

3.1. Scope of the tests

The tests have been carried out at the laboratory stand for meas-

a)



b)

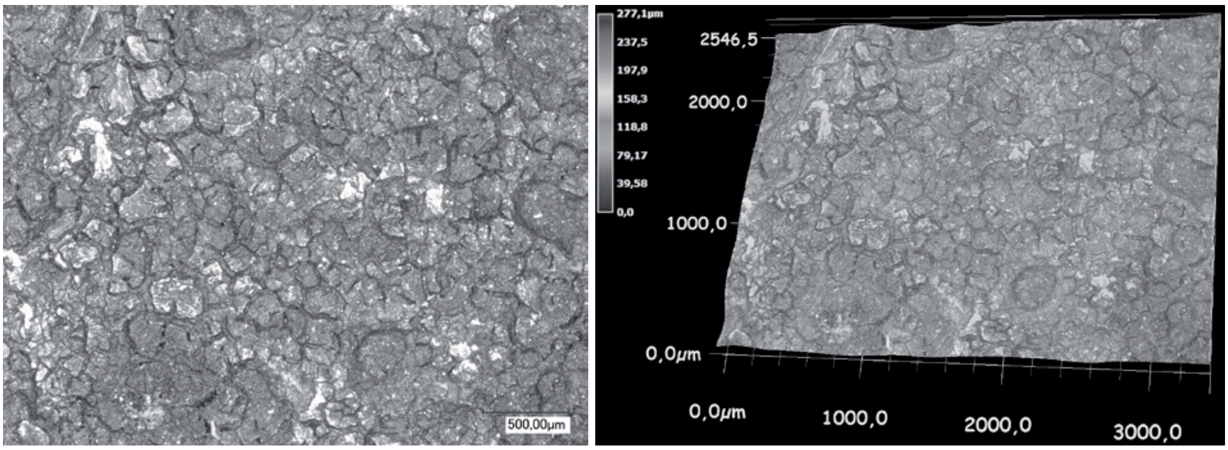


Fig. 3. The view of the specimen surface (linear magnification 100x) and the 3D images of the micro-surfaces: a) for the cleaned specimen – max. depth of the surface irregularities is of 37.6 µm, b) for the corroded specimen - max. depth of the surface irregularities is of 277.1 µm

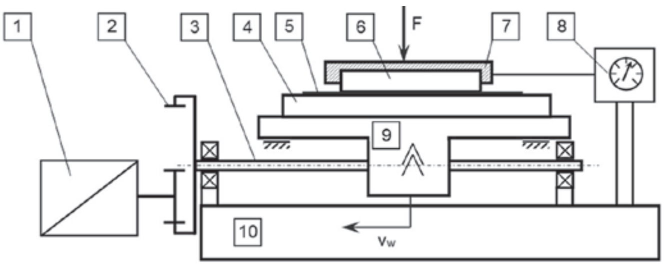


Fig. 4. The diagram of the measuring stand: 1 – motor, 2 – planetary reducer, 3 – screw propeller, 4 – specimen, 5 – intermediate layer (graphite grease or water), 6 – counter-specimen, 7 – handle, 8 – force indicator, 9 – sliding table, 10 – body

oxides is built-up [15], and the indentations and pitting between them. It will have an influence on obtaining the values of the friction coefficients [12]. The cracks visible on the surface of the cleaned surface are an effect of its former grinding.

uring the friction force (fig. 4), in the closed space, at the ambient temperature on the level of 22 ±0,10°C. The stand enables to obtain a relative movement of one element of the friction pair, assembled on the sliding table, in respect of the other one, maintained in the handle. The handle is connected through a force sensor, with the stand body. The sliding table is put into motion by a planetary reducer, powered by an electric motor. A change of the rotational velocity of the motor enables to change a sliding velocity of the friction pair elements. The constant sliding velocity is maintained in the course of a test. A change of the normal load (F) is realised by a change of a quantity of the weights featuring the known masses that exert pressure on the tested elements.

The experimental tests have been carried out for two types of the friction pairs. The first pair has been covered with a rust layer, but the second one has been cleaned prior to the test.

In respect of each pair, there have been changed both the load conditions, condition of the intermediate layer, and the sliding velocity. The tests have been performed for two loads, which are of

F = 58 and 107 N, respectively. The relative sliding velocity has been v_w = 0.0515; 0.111; 0.225 and 0,348 mm/s, respectively. These values result, first and foremost, from the possibilities offered by the test stand. Nevertheless, they are similar to the sliding velocity of the spring leaf ends at the typical operational conditions of the vehicles. In the case of the considered spring, at the maximum deflection (i.e. by 150 mm) a relative displacement of the longest spring leaves is of 3.49 mm, but of the shortest ones is of 0.866 mm. While deflecting with a frequency of 1 Hz, the maximum velocity values are of 10.9 and 2.72 mm/s, respectively. In the real operational conditions such large deflections (apart from the extreme situations) do not take place. If an average value of the spring deflection amplitude is assumed on the level of 20 mm, so with a frequency of 1 Hz a maximum velocity of the relative displacements of the spring leaves will be of 0.7 up to 2,9 mm/s, respectively for the shortest and the longest spring leaves.

At the first stage, the tests have been carried out for the dry surfaces. Them the friction surfaces of the corroded specimens have been moistened with water, and the cleaned specimens have been covered with a graphite grease layer. The surface conditions assumed for the tests result from these, which are found in the real operational conditions for the multi-leaf springs.

3.2. Results of the tests

In the course of the experimental tests, it has been recorded a value of the friction force between the specimen and the counter-specimen for each variant. A sampling period has been of Δt=0.02 s, and a quantity of the specimens n has been changed depending on the sliding velocity. For each variant, 3 up to 5 repetitions have been performed. In fig. 5 are presented the exemplary courses of the friction forces (three repetitions) obtained for the specimen covered with the rust for the dry surface condition at the determined sliding velocity. During the conducted trials, for most associations, it has been observed a great repeatability of the recorded results.

On the basis of the recorded courses, upon referring the friction force to the interacting normal load, a value of the static and kinetic friction coefficient has been determined. On the graphs (fig. 6) can be

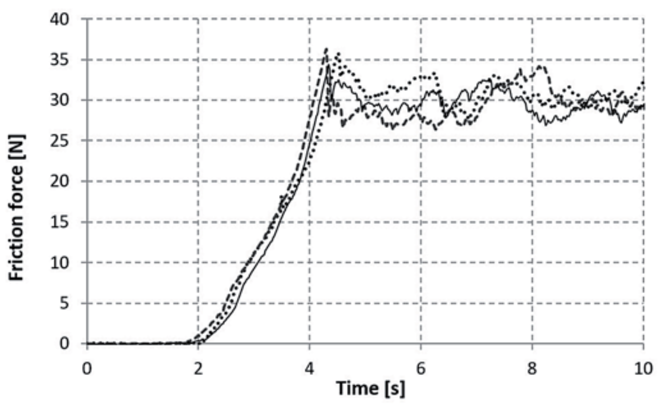


Fig. 5. The changes of a friction force as a function of time (three repetitions) at the determined sliding velocity obtained for the corroded specimen featuring the dry surface

observed three characteristic areas. In the first are, the friction force increases progressively, achieving its maximum value. In the second area, a transition from a condition of the sliding friction, and a value of the friction force decreases progressively. In the third area is noticed a stabilisation of the force value.

An issue of importance has constituted a determination of the borders of the particular areas. A border between I and II area corresponds to a moment, when it has been observed the maximum value of the coefficient. As a static coefficient has been assumed a maximum value observed prior to a beginning of the relative movement between the tested elements (i.e. on the border of the I and II area). The more complex issue has constituted a determination of the borders between II and III area (i.e. time $t_{III} = k \cdot \Delta t$, where k-times of sampling period (fig. 6)). It begins upon at the moment, upon which a coefficient of course variation has been lower than 10%. The coefficient of variation V has been calculated on the basis of a dependency (1):

$$V = \frac{\sigma}{\bar{\mu}} * 100\% \tag{1}$$

where: $\bar{\mu}$ – average value in the range from k to n described by the formula (2):

$$\bar{\mu} = \frac{1}{(n-k)} \sum_{i=k}^{n} \mu_i \tag{2}$$

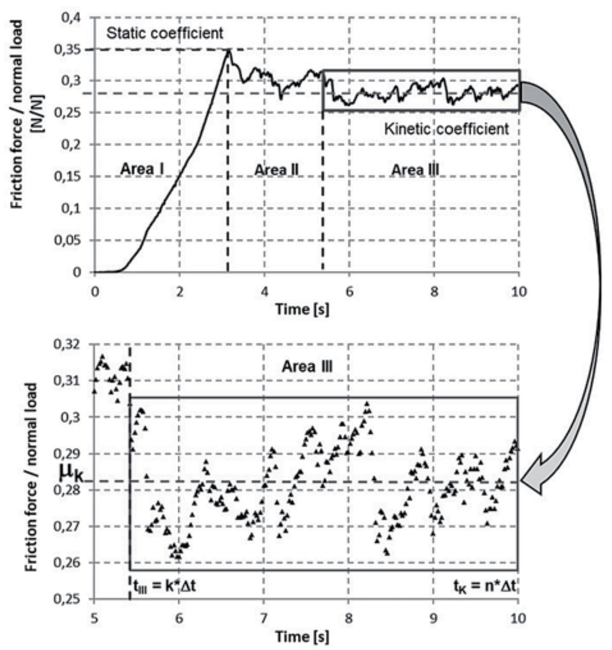
σ – standard deviation in the range from k to n, described by dependence (3):

$$\sigma = \sqrt{\frac{\sum_{i=k}^{n} (\mu_i - \bar{\mu})^2}{(n-k)}} \tag{3}$$

μ_i – another value of the coefficient in area III.

A kinetic coefficient has been determined as an average value related to a fragment, for which a stabilization of the friction value has been observed (area III) ($\mu_k = \bar{\mu}$).

Fig. 6. Principle of a determination of the static and kinetic friction coefficients



Then, an arithmetic mean from the obtained results has been calculated in respect of a few repetitions (from 3 up to 5), for each test variant. The obtained test results have been summarized in a form of the graphs of the static and kinetic friction coefficient as a function of the normal load for the different sliding velocities (fig. 7 and fig. 8). It has been used a uniform scale in order to facilitate a comparison of the results presented in all graphs. Moreover, in table 4 and 5, the results of the friction coefficients (μ_s and μ_k), average values of the coefficients

($\mu_{s\text{-}sr}$ and $\mu_{k\text{-}sr}$), determined for a measuring series and corresponding with them values of the standard deviations ($\sigma_{\mu s}$ and $\sigma_{\mu k}$) have been summarised. While analysing the results for the cleaned surfaces it should be stated that there cannot be noticed a significant influence of the sliding velocity on the obtained values of the static and kinetic friction coefficient. The differences do not exceed even a few percent. For the dry surfaces, it is observed a slight (approx. 8%) increase of a value of the static friction coefficient. Covering the surface with the graphite grease decreases a value of the static and kinetic friction coefficient by approx. 7% in the case of a load of 58 N and even by 20% in the case of a load of 107 N. Furthermore, it causes a decrease in an influence of the normal load on a value of the friction coefficients.

The values of the static and kinetic friction coefficient for the dry surfaces covered with the rust are approx. 230-270% higher than for the clean surfaces, without any rust.

Moreover, it is observed a larger influence of the sliding velocity on the values of the friction coefficients. The differences between the maximum and minimum values of the coefficients are approx. of 20%. For the larger load, a decrease of the value of the friction coefficients has been observed. It is mainly caused by the surface smoothing and removing the corroded layer during the subsequent measuring series. It has been noticed that a larger load promotes a disintegration of the rust particles. It has been built-up in the indentations developing the distinctive intermediate layer. Moistening the specimen surface with water has resulted in a decrease of the friction coefficient value from 17 to 20%. Additionally, it has been observed that it has caused a slight decrease of the value of the standard deviation to the average value ratio (for the dry surfaces covered with the rust it is approx. of 12-13%, and for the moistened surface – approx. of 10-11%).

Table. 4. Values of friction coefficients for vertical load F = 58 N

Surface condition	Velocity [mm/s]	Static friction coefficients			Kinetic friction coefficients		
		μ_s	$\mu_{s\text{-}sr}$	$\sigma_{\mu s}$	μ_k	$\mu_{k\text{-}sr}$	$\sigma_{\mu k}$
Clean, dry	0,0515	0,130	0,135	0,0073	0,134	0,135	0,0051
	0,111	0,131			0,134		
	0,225	0,141			0,138		
	0,348	0,139			0,135		
Clean, covered with graphite grease	0,0515	0,135	0,125	0,0114	0,136	0,121	0,0127
	0,111	0,123			0,118		
	0,225	0,120			0,113		
	0,348	0,123			0,117		
Covered with rust, dry	0,0515	0,419	0,369	0,0486	0,348	0,315	0,0300
	0,111	0,385			0,323		
	0,225	0,338			0,298		
	0,348	0,335			0,291		
Covered with rust and moistened with water	0,0515	0,350	0,311	0,0343	0,301	0,270	0,0232
	0,111	0,297			0,261		
	0,225	0,280			0,254		
	0,348	0,315			0,263		

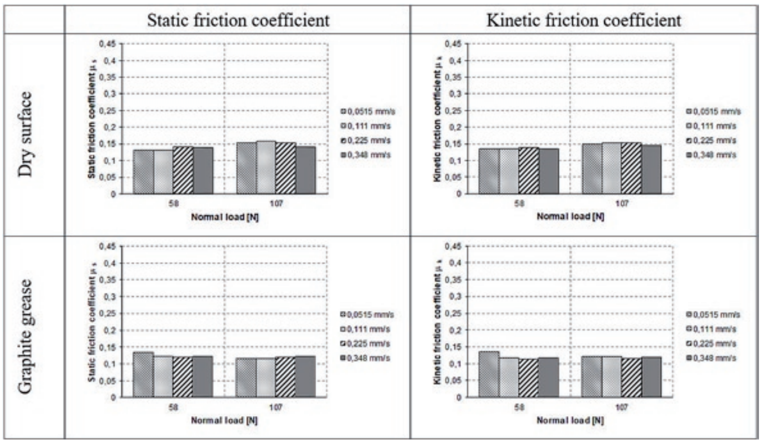


Fig. 7. Juxtaposition of static and kinetic friction coefficients for cleaned surfaces (Tab. 4 and 5)

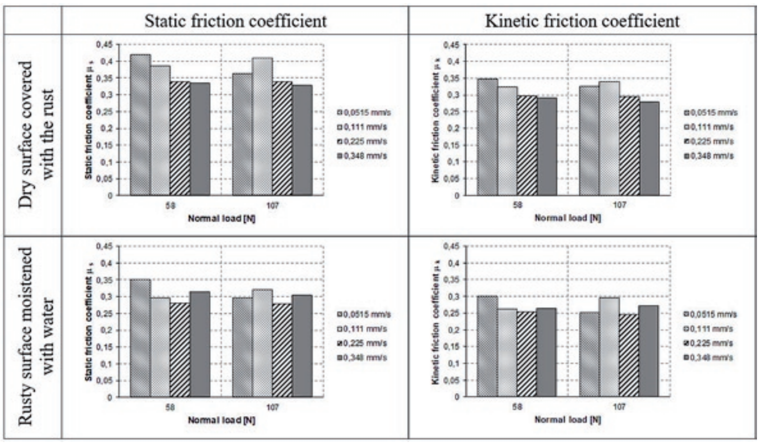


Fig. 8. Juxtaposition of static and kinetic friction coefficients for rusty surfaces (Tab. 4 and 5)

4. Conclusion

The selected results of the laboratory tests related to the friction coefficient values for the different surface condition of the spring leaves of the multi-leaf spring have been presented in the study. Due to a nature of its operating in the vehicle suspension, an existing friction significantly influences the spring static and dynamic characteristics of the spring (suspension). In the course of the vehicle operating takes place a removal of the grease layer from the mating surfaces of the spring leaves. A failure to provide a proper maintenance favours corrosion, which by changing the friction conditions results in a change of the spring elasticity and an increase of the dissipated energy. In the extreme cases it leads to a blockade of the spring leaves at the small amplitudes of the suspension deflections.

The spring leaf surfaces covered with the rust significantly change its operating conditions. It has an obvious impact on worsening the ride comfort, a change of the dynamic loads and a vibration frequency structure of the vehicle.

In the study, one has been considered four surface conditions: clean and dry, covered with the graphite grease (it is a desired condition), as well as covered with a layer of the rust and pitting, which has been moistened with water, at the subsequent stage. The carried-out tests of the surface roughness have visualised the significant differences related to all roughness parameters for the cleaned and corroded specimens. It has affected the results of the friction coefficient tests.

The obtained values of the static and kinetic friction coefficients for the cleaned specimens are similar to those presented in the literature [1].

While analysing the results for the cleaned surfaces it should be stated that there is no noticeable impact of the sliding velocity on the obtained values of the static and kinetic friction coefficients. These differences do not exceed even a

Table. 5. Values of friction coefficients for vertical load F = 107 N

Surface condition	Velocity [mm/s]	Static friction coefficients			Kinetic friction coefficients		
		μ_s	$\mu_{s\text{-}sr}$	$\sigma_{\mu s}$	μ_k	$\mu_{k\text{-}sr}$	$\sigma_{\mu k}$
Clean, dry	0,0515	0,154	0,151	0,0104	0,149	0,150	0,0080
	0,111	0,157			0,153		
	0,225	0,154			0,154		
	0,348	0,141			0,145		
Clean, covered with graphite grease	0,0515	0,116	0,119	0,0092	0,122	0,120	0,0092
	0,111	0,117			0,121		
	0,225	0,120			0,116		
	0,348	0,123			0,120		
Covered with rust, dry	0,0515	0,364	0,360	0,0438	0,326	0,310	0,0300
	0,111	0,409			0,339		
	0,225	0,339			0,296		
	0,348	0,328			0,278		
Covered with rust and moistened with water	0,0515	0,297	0,301	0,0315	0,251	0,266	0,0246
	0,111	0,321			0,296		
	0,225	0,279			0,246		
	0,348	0,305			0,271		

few percent. Introducing the graphite grease has resulted in a decrease of the friction coefficient values. It is particularly noticeable in the case of the larger load, for which the changes by approx. 20% have been determined.

The measurement of the static and kinetic friction coefficients has demonstrated the 230-270% increase of their values in comparison with the clean surfaces. In the course of the tests, one has observed smoothing of the surfaces – especially in the case of the larger load. Moistening the friction surfaces with water has caused developing a distinctive lubricant layer that has decreased the friction coefficients approx. by 15%.

References

1. Avallone E A, Baumeister T, Sadegh A. Marks' Standard Handbook for Mechanical Engineers. Eleventh Edition. New York: McGraw-Hill Education, 2007.
2. Clavel A, Sorine M, Zhang Q. Modeling and Identification of a Leaf Spring System. IFAC Advances in Automotive Control 2001; 34 (1): 307-312, https://doi.org/10.1016/S1474-6670(17)34415-4.
3. Drozd K. Causes of Leaf Springs damages verifiable analysis. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2003; 2(18): 24-27.
4. Krason W, Wysocki J. Experimental verification of numerical test results for a double multi-leaf spring. Journal of Kones 2009; 16 (3):185–193.
5. Krason W, Wysocki J. Investigation of friction in dual leaf spring. Friction and Wear 2017; 38 (3): 214–220, https://doi.org/10.3103/S1068366617030096.
6. Mitutoyo. Surface roughness measurement. Practical tips for laboratory and workshop. Bulletin No. 1984. https://www.mitutoyo.com/wp-content/uploads/2015/04/Surface_Roughness_Measurement.pdf, 26.05.2018.
7. Napadłek W, Burakowski T, Woźniak A, Chrzanowski W. Mikrostruktury napoiny typu Stellit Co-6 wytworzonej techniką laserową na stali X22CrNoV12-1. Wybrane Problemy w Mechatronice i Inżynierii Materiałowej 2016: 258-267.
8. Oh S D, Jung W W, Bae D H, Lee Y Z. Friction and Wear Characteristics for Automotive Leaf Spring Material due to the Influence of the Residual Stress, Key Engineering Materials 2005; 297-300: 1388-1394, https://doi.org/10.4028/www.scientific.net/KEM.297-300.1388.
9. Patil P P, Patil A V. Critical Review of Research Paper for Analysis of Multi Leaf Spring, International Journal of Latest in Engineering and Technology 2014; 3 (3): 171-174.
10. Qing L, Wei L. A contact finite element algorithm for the multileaf spring of vehicle suspension systems. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering 2004; 218 (3): 305-314, https://doi.org/10.1243/095440704322955821.
11. Toso A, Facchin U, Melzi S. Multibody and Finite Element models of a leaf-spring suspension for vehicle dynamics applications: numerical model, tests and correlation. ECCOMAS Thematic Conference on Multibody Dynamics 2015.
12. Wieczorek J, Sleziona J, Myalski J, Dolata-Grosz A, Dyzia M, Sluzalek G. Surface wear and structure after collaboration in dry sliding conditions in MMCS AK12 and ceramic particles AL2O3. Kompozyty (Composites) 2002; 2 (4): 212-216.
13. Wiśniewska M. Surface roughness measurements: emerging importance of the uncertainty evaluation, PhD Interdisciplinary Journal 2014; 1: 243-250.
14. Zhou Z, Guo W, Shen T, Wang F, Ju J, Wang H, Song E. Research and Application on Dynamic Stiffness of Leaf Spring. SAE-China, FISITA Proceedings of the FISITA 2012 World Automotive Congress. Lecture Notes in Electrical Engineering 2012; 198: 105-119, https://doi.org/10.1007/978-3-642-33795-6_10.
15. Żurowski W. Structural factors contributing to increased wear resistance of steel friction couples. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2012; 14 (1): 19-23.

Zdzisław HRYCIÓW
Wiesław KRASOŃ
Józef WYSOCKI
Department of Mechanical Engineering
Military University of Technology
gen. Witolda Urbanowicza 2, 00-908 Warsaw, Poland

E-mails: zdzislaw.hryciow@wat.edu.pl, wieslaw.krason@wat.edu.pl, jozef.wysocki@wat.edu.pl

Article citation info:
GONERA J, NAPIÓRKOWSKI J. Model for forecasting the geometry of the floor panel of a passenger car during its operation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (4): 689–695, http://dx.doi.org/10.17531/ein.2018.4.20.

Jarosław GONERA
Jerzy NAPIÓRKOWSKI

MODEL FOR FORECASTING THE GEOMETRY OF THE FLOOR PANEL OF A PASSENGER CAR DURING ITS OPERATION

MODEL PROGNOZOWANIA STANU GEOMETRII PŁYTY PODŁOGOWEJ SAMOCHODU OSOBOWEGO W TOKU EKSPLOATACJI*

A number of vehicle users pay attention to the impact of changes in the car body geometry during long-term use on the safety level. However, this issue has not been properly dealt with in research studies. The aim of this study was to identify changes in the floor panel, to develop a model to forecast the geometry during the car use and to identify the points which undergo the maximum displacement. The paper presents the effect of the car mileage on the floor panel condition, taking into account variable environmental factors. In the course of the study, the position of points fixing the front suspension, front bench and rear suspension was determined, as was the position of points situated on parts of the load bearing structure of the car body. The results were used to develop a model for forecasting changes of the floor panel geometry during car use. The probability of changes in the floor panel geometry was found to increase with the mileage. The probability of reaching the maximum permissible geometric changes (3 mm) in a floor panel is accurately described by the probabilistic model in the form of the Rayleigh distribution. Diverse models of the floor panel geometry changes were obtained depending on the environmental conditions and type of the base points under analysis.

Keywords: *passenger car, car body, floor panel, car body geometry, safety.*

Wielu użytkowników samochodów osobowych zwraca uwagę na istotność wpływu na poziom bezpieczeństwa zmian geometrii nadwozia pojazdów podczas ich wieloletniej eksploatacji. Jednak dotychczas zagadnienie to nie znalazło odpowiedniego odzwierciedlenia w literaturze. Celem pracy była identyfikacja zmian geometrii płyty podłogowej, opracowanie modelu prognozującego stan geometrii w toku eksploatacji i zidentyfikowanie punktów ulegającym największym przemieszczeniom. W pracy przedstawiono wpływ przebiegu pojazdu na stan geometrii płyty podłogowej z uwzględnieniem zróżnicowanych warunków środowiskowych. Podczas badań określano położenie punktów mocujących zawieszenie przednie, przednią ławę i zawieszenie tylne oraz położenie punktów znajdujących się na elementach struktury nośnej nadwozia. Na podstawie uzyskanych wyników opracowano model prognozowania zmian geometrii płyty podłogowej w toku eksploatacji. Stwierdzono, że prawdopodobieństwo zmian geometrii płyty podłogowej podczas eksploatacji rośnie w czasie, wraz ze wzrostem przebiegu. Prawdopodobieństwo osiągnięcia stanu dopuszczalnego (3 mm) zmian geometrycznych na płycie podłogowej dobrze opisuje model probabilistyczny w postaci rozkładu Rayleigha. Uzyskano zróżnicowane modele zmiany geometrii płyty podłogowej w zależności od warunków środowiskowych oraz rodzaju analizowanych punktów bazowych.

Słowa kluczowe: *samochód osobowy, nadwozie, płyta podłogowa, geometria nadwozia, bezpieczeństwo.*

1. Introduction

The issue of an assessment of the condition of a passenger car body geometry is usually considered in the context of repair work [17, 22]. To this end, procedures have been developed for car approval for traffic by car manufacturers, as well as relevant regulations [22]. The issue of the car body technical condition is linked inextricably with the safety of its use [3].

Safety system development in modern cars is oriented mainly towards reducing the risk of a car accident and minimising the injury if such an accident happens [6, 11]. The construction of the car body, aimed at limiting the accident effects by minimising the injuries of the car driver and passengers, is one of the most important elements of passive safety [5, 15, 18, 25]. Active car body-related safety elements mainly include the appropriate deployment of the fixing points for parts of the suspension and the steering system which directly affect the wheel geometry [4, 10, 20, 22, 23].

The car body geometry is understood to denote the appropriate, in line with the manufacturers' requirements, deployment of all the base points on the floor panel and on the upper parts of the car body relative to the three reference planes [1, 8]. There are base points on car bodies which are used for geometry measurements. These points usually include structural holes, which are used to fix subassemblies, and auxiliary holes, used especially for measurements [1, 16]. Manufacturers of passenger cars usually assume that the difference between the required and the actual position of base points should not be greater than 3 mm [12, 16, 17]. When base points in crumple zones are displaced during an accident or a road collision, some unexpected distortions of the body may occur, which absorb virtually no energy [7, 8, 24]. In modern cars, it is not possible to regulate many parameters of the suspension and steering systems. Therefore, for example, the camber angle, the kingpin inclination or the castor angle are not adjustable. A change of the car body geometry will therefore result in a change of these parameters, which may make it difficult to maintain the right kinematics of motion [22, 26].

To date, few authors have dealt with changes of the technical condition of accident-free car bodies during their use. The issue is also omitted in regular vehicle technical inspections. The main causes of changes in the technical condition of a car body include road ac-

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

cidents or collisions, weather conditions and prolonged use of a car on low quality roads [2, 12, 19, 27]. The issue of the car body wear during its use has been mentioned by the authors of [2, 3, 12, 19]. These are usually general remarks, stating that the manner of a car use affects the car body condition [14]. Nevertheless, the authors of [9] have analysed the effect of the passenger car mileage on changes in the floor panel geometry. According to their findings, the maximum permissible changes of the body geometry (3 mm) at some base points occur after 150,000 km [9]

2. A system approach to changes in the car body geometry

During car use, responses to the road pavement are transmitted through the suspension system [27]. When driving, a passenger car is subjected to continuous kinematic and forced excitations of a broad range of values [14]. During car use, its body is subjected to static and dynamic loads. The static loads result from the torque originating from forces from the pavement, transmitted through the suspension system. They are also a result of bending loads, being a consequence of the mass of the car, passengers and cargo. On the other hand, the dynamic loads are associated with speed and acceleration of a vehicle; they originate while taking bends, driving along a bumpy road, braking and gaining speed [13, 27]. During vehicle use, its body is exposed to such factors as corrosion and fatigue, etc., which contribute to progressive degradation [15].

Compared to those in Western Europe, roads in Poland are in a worse technical condition [21]. According to data received from GDDKiA, up to 38.3% of the trunk roads in Poland require repair work. The condition of regional, county and commune roads is even worse.

During car use, wear processes take place which include changes in the geometry of the floor panel and upper parts of the body (Fig. 1). The geometry of a car body is characterised by a set of characteristics C. A body during car use is subjected to a variety of excitations W, which bring about changes of geometry, which in these cases are responses to the wear process Z.

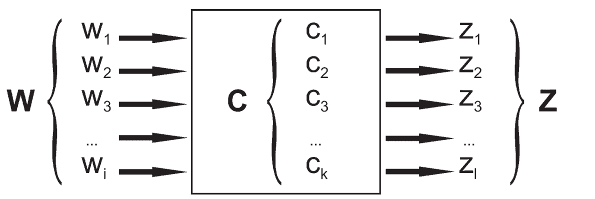


Fig. 1. A graphic illustration of the process of a car body wear during its use: C – a set of characteristics of an object, W – a set of excitations, Z – a set of responses of the object

C = {c_1, c_2, c_3, …, c_k} k = $\overline{1,K}$, where: c_k is a representation of the actual characteristics of the body geometry, k = 1, 2, 3, …K;

- c_1 – position of base points which characterise the active safety on the right side of the floor panel;
- c_2 – position of base points which characterise the active safety on the left side of the floor panel;
- c_3 – position of base points which characterise the passive safety on the right side of the floor panel;
- c_4 – position of base points which characterise the passive safety on the left side of the floor panel;
- c_k – k-th characteristic of the car body geometry.

W = {w_1, w_2, w_3, …, w_i} i = $\overline{1,I}$, where: w_i is a representation of the real excitations acting on the car body during its use, i = 1, 2, 3, …I;

- w_1 – total mileage of a vehicle;
- w_2 – environmental conditions in which a car is used, associated to the country in which it is used;
- w_3 – age of the car;
- w_4 – characteristics of the car use so far;
- w_5 – road incidents in which the car may have participated;
- w_6 – factors exceeding standard use;
- w_7 – environmental conditions of use;
- w_i – i-th excitations acting on the car.

Z = {z_1, z_2, z_3, …, z_l} l = $\overline{1,L}$, where: z_l is a representation of actual responses, l = 1, 2, 3, …L;

- z_1 – change of position of points which characterise the active safety on the right side of the floor panel;
- z_2 – change of position of points which characterise the active safety on the left side of the floor panel;
- z_3 – change of position of points which characterise the passive safety on the right side of the floor panel;
- z_4 – change of position of points which characterise the passive safety on the left side of the floor panel;
- z_l – l-th characteristic of changes of the car body geometry.

The literature analysis has shown that there have been no studies aimed at determining the effect on the changes of the car body geometry of: w1 – total mileage of a vehicle and w2 – environmental conditions of use. The quantitative effect of these parameters on changes in the body condition has not been identified so far. Therefore, the effect of these characteristics has been examined.

The aim of this study was to identify changes in the floor panel and to develop a model to forecast its geometry during car use. As an auxiliary objective, characteristic points of the body with the greatest displacement were identified.

3. Study methodology

A total of 120 passenger cars with diverse mileage from 10,000 to 360,000 km were included in the study. The vehicles were divided into two categories. The first category, marked PL, included cars used on domestic roads. The other, marked EU, included cars used on the roads of Western Europe. Each of the two categories included the same number of cars – 60. The cars had the same construction parameters (i.e. hatchback type body, spark-ignition engine and the front-wheel drive). None of them had been in an accident or a collision, in none of them had any events been identified which would go beyond normal use (e.g. exceeding the maximum allowed capacity).

The floor panel geometry was measured with an accuracy of 1 mm with a Gysmeter device manufactured by Gys (Fig. 2). The measurement range of the Gysmeter device was between 400 and 2650 mm. The measurement device was equipped with a dedicated set of measurement tips, fitted to the base points under analysis. Owing to the set, measurements could be conducted without dismounting parts of the suspension and steering system. The actual position of the characteristic base points was measured and compared to the position required by the vehicle manufacturers. The measurements were made relative to the reference points, situated at the back part of the car, in the rear of the passenger compartment. No changes in geometry, including deformations, were identified at the reference points in the cars under study. The data on the position of the base points on the floor panel were taken from the database in the Allvis Light programme. This provides the required distance (in mm) between individual base points.

A total of 12 characteristic base points were selected, which could be found in each of the passenger cars under study. Six of them were on the right and six on the left side of a car. Three of the points were associated with active safety and three with passive safety.

Fig. 2. Measurements of the floor panel geometry with a GYSMETER device manufactured by GYS

An analysis was conducted of the characteristics associated with (Fig. 3):

- geometry of the suspension and the steering systems (active safety) - the front suspension fixing points, the rear points fixing the front bench and the rear suspension fixing points;
- passive safety - points situated near the bulkhead, at the beginning of front longitudinals and at the end of the rear longitudinals.

During the measurements, each car was placed on a 2-column lift and fixed as recommended by the manufacturer. Dedicated magnetic tips on a permanent external pole and suitable measurement tips on a moveable pole were used with the Gysmeter device.

Changes in geometry for the given points of the floor panel Pzg were determined from the formula:

$$P_{zg} = |\mathbf{WOPB} - \mathbf{ROPB}| [\text{mm}], \quad (1)$$

where:

WOPB – required distance between the base points [mm];
ROPB – actual distance between the base points [mm].

The uncertainty for measurements of the actual distance between the base points (ROPB) was 1 mm. In consequence, uncertainty for P_{zg}, i.e. change of the geometry for the given points of the floor panel, was also 1 mm.

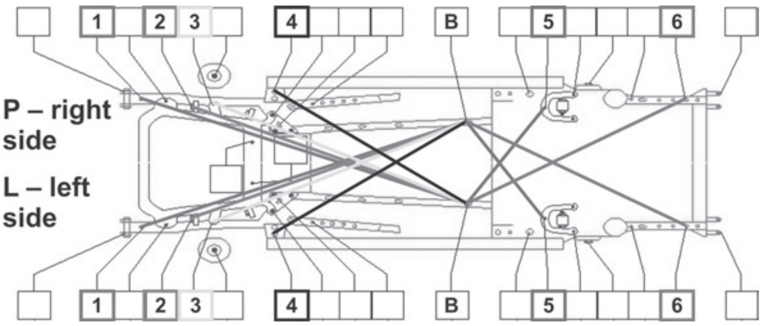


Fig. 3. Deployment of points at which measurements of the floor panel geometry was conducted, where: B – reference points, initial during the measurements; 1 – points situated at the beginning of the front longitudinals; 2 – front suspension fixing points; 3 – rear points fixing the front bench; 4 – points situated near the bulkhead; 5 – rear suspension fixing points; 6 – points situated at the end of the rear longitudinals

4. Analysis of the floor panel geometry changes

The size of the changes of the base points position (Table 1) was related to their deployment on the floor panel. The smallest changes in the floor panel geometry were observed near the bulkhead. Particularly large changes were observed at the base points of fixing the front suspension, rear suspension and at the front bench fixing point. The average displacement in the cars under study in these areas exceeded 6 mm. The point position changes exceeding 10 mm were also identified.

Much greater displacements of the base points were observed in the cars used in Poland compared to those used in Western Europe (Table 1). This applied to all the areas of the floor panel. Regardless of the place, the displacements of characteristic points of the floor panel in the cars used in Poland were greater than in the cars used in Western Europe. The displacements near the bulkhead (4P and 4L) were slightly greater in cars used in Poland. The differences between the cars used in Poland and those used in Western Europe were the smallest in this area (up to 0.3 mm). Similarly, relatively small differences between the two categories were observed at the points situated at the beginning of the front longitudinals (1P and 1 L) and at the end of the rear longitudinals (6P and 6L). The differences reached approx. 0.5 mm. Greater differences in the geometry changes between cars used in Poland and in Western Europe were observed at the base points connected with fixing the front suspension (2P and 2L), the front bench (3P and 3L) and the rear suspension (5P and 5L). The differences in these areas exceeded 1.5 mm. And the greatest base point displacements were observed at the rear suspension fixing points situated on the right side of the car body. Differences exceeding 3 mm were observed at these points in cars used in Poland and in Western Europe.

The analysis of variance has shown that changes in the floor panel geometry in cars used in Poland and in Western Europe differed significantly (Fig. 4).

Table 1. A list of mean changes of the geometry at the base points depending on where a car is used

The geometry changes at the base points			
Place of car use		**PL [mm]**	**EU [mm]**
1P	Front longitudinal	2,13	1,83
1L		2,13	1,67
2P	Fixing of front suspension	6,25	4,67
2L		5	3,5
3P	Fixing of front bench	4,88	3,33
3L		4	2,67
4P	Near the bulkhead	1,5	1,33
4L		1,63	1,33
5P	Fixing the rear suspension	6,5	3,33
5L		4,88	3
6P	Rear longitudinal	2,63	2
6L		2,13	1,67

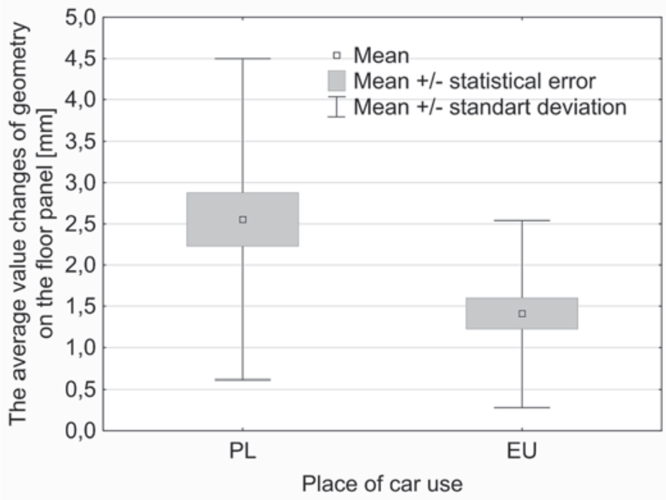
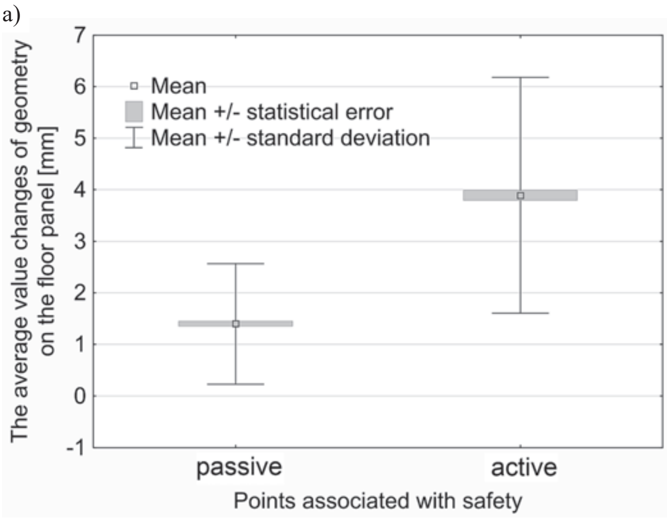


Fig. 4. A comparison between the base points displacement in cars depending on the country of use

a)



b)

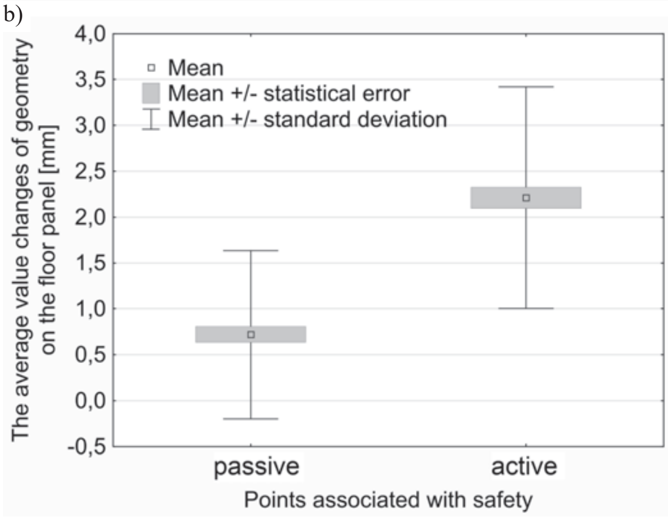
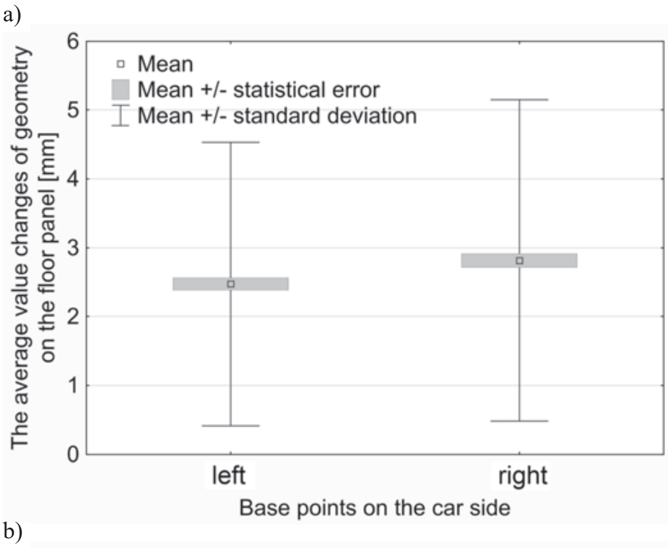


Fig. 5. A comparison between the base point displacements in cars used on roads: a. in Poland. b. in Western Europe. depending on the type of safety for which they are responsible

This study also employed an analysis of variance. whose aim was to determine the effect on the geometry changes in the type of safety for which the given base points are responsible (Fig. 5) and the side of the car where they are situated (Fig. 6).

a)



b)

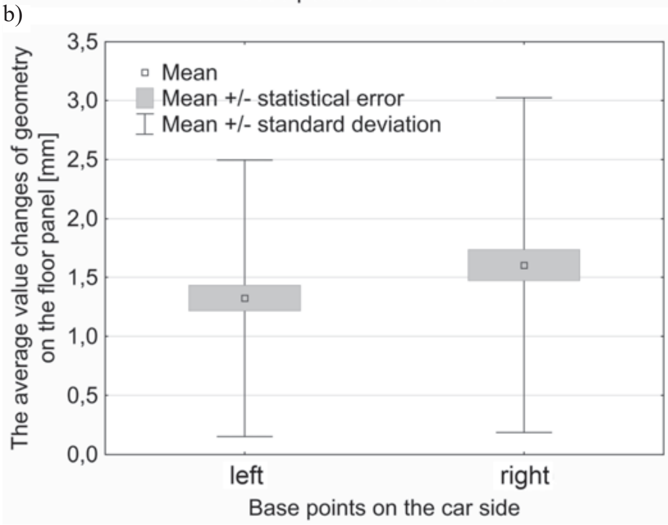


Fig. 6. A comparison between the base point displacements in cars used on roads: a. in Poland. b. in Western Europe. depending on the side of a car where the points were situated

Significant changes were found between the base points related to passive and active safety. Such differences were observed both in cars used in Poland and in those used in Western Europe. Greater geometry changes were observed at the base points related to active safety. Moreover. greater changes in the floor panel geometry - both in the cars used in Poland and in those used in Western Europe - were observed at the points situated on the right than on the left side of the car body.

The intensity of the floor panel geometry changes was constant. regardless of the car mileage (Fig. 7). However. an increase in the total geometry changes was observed with increasing mileage. The intensity of the geometry changes depended on the country in which a car was used. the type of safety for which the base points were responsible and the side on which they were situated. Tables 2 and 3 show the intensity of the floor panel geometry changes observed in this study.

Greater changes of the floor panel geometry changes in regard to the mileage were observed in the cars used in Poland. They were twice greater in virtually all cases. Moreover. several times greater changes were observed at points responsible for active than those responsible for passive safety of a car. The differences between the geometry changes on the right and left side of a car were not significant.

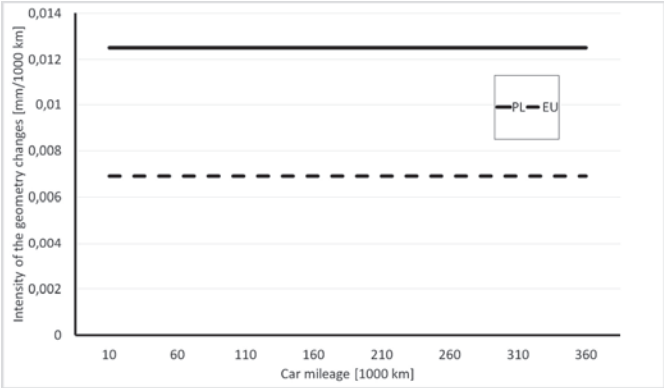


Fig. 7. Intensity of the floor panel geometry changes in cars used in Poland and in those used in Western Europe

Table 2. Intensity of the geometry changes in cars used in Poland

Cars used in Poland					
Type of safety	Passive		Active		At all the floor panel points under study
Side of a car	Right	Left	Right	Left	
Intensity of the geometry changes [mm/1000 km]	0,0060	0,0059	0,0214	0,0168	0,0125

Table 3. Intensity of the geometry changes in cars used in the EU

Cars used in Western Europe					
Type of safety	Passive		Active		At all the floor panel points under study
Side of a car	Right	Left	Right	Left	
Intensity of the geometry changes [mm/1000 km]	0,0026	0,0025	0,0128	0,0099	0,0069

5. A model of the floor panel geometry changes during car use

The following assumptions were adopted to develop a model of the floor panel geometry changes during the car use;

- the critical displacement size of the characteristic points is 3 mm;
- changes of the base point positions are linear for the analysed operational excitations.

Therefore. a model of the floor panel geometry changes in the deterministic approach will have the following form:

$$Z_{pp} = P_p \cdot I_{zg}, \text{ where} \quad (2)$$

Z_{pp} – change of the base points position;
P_p – car mileage;
I_{zg} – intensity of the floor panel geometry changes.

The probability of reaching the maximum allowable displacement of 3 mm grows linearly with increasing car mileage. After the geometry changes reach the critical value. a car must be withdrawn from use or transferred to a garage for repair. The model was verified based on the cumulative distribution function for the Weibull distribution:

$$F(X) = 1 - e^{-(x/\gamma)^k}, \text{ where:} \quad (3)$$

F(x) – probability of reaching the critical values (3 mm) of the geometry changes;
k > 1 – parameter of the distribution shape;
γ > 0 – parameter of the distribution scale.

The probability of changes in the floor panel geometry was found to increase with the mileage. which corresponds to the distribution shape parameter of 2. In consequence. the model adopted a specific form of the Weibull distribution (shape parameter k=2). called the Rayleigh distribution. It applied both to cars used in Poland and in those used in Western Europe. The probability of the floor panel geometry changes was described with the following relationship:

$$F(Z_{pp}) = 1 - e^{-(Z_{pp}/\gamma)^2}, \text{ where:} \quad (4)$$

Z_{pp} – change of the floor panel geometry at a given base point.

Based on the measurement results. the scale parameter was taken as two (γ=2). Therefore. the probability of floor panel changes during use had the following form:

$$F(Z_{pp}) = 1 - e^{-(Z_{pp}/2)^2} \quad (5)$$

The formula describing changes of the floor panel geometry at a given base point was transformed and supplemented with the statistics (w) determined in the Statistica software. The following relationship was obtained:

$$P_p = \frac{Z_{pp}}{w \cdot I_{zg}} \quad (6)$$

Tables 4 and 5 present the probability of changes of the floor panel geometry depending on the country in which the car was used. the side where the given base point is situated and the type of safety for which it is responsible.

Data presented in Tables 4 and 5 show that there are considerable differences in the mileage at which 3 mm displacements of the points appear. The differences depend on the condition of roads associated with the country in which a car is used. type of safety for which the given base points are responsible and the side of a car on which they are situated. For example. a geometry change of 3 mm will appear at the base points on the right side of a car. responsible for active safety. with a probability of p=0.05 at a mileage of 41 thousand km in cars used in Poland and at a mileage of 68 thousand km in cars used in Western Europe. On the other hand. the mileage for the left side will be 52 thousand km and 88 thousand km. respectively. For the base points associated with passive safety. a geometry change of 3 mm will appear at a mileage of 143 thousand km in cars used in Poland and at a mileage of 328 thousand km in cars used in Western Europe. It will also be 147 thousand km and 341 thousand km. respectively. on the left side of the car body.

6. Summary

With the mileage of a passenger car exceeding 120.000 km. changes of its floor panel geometry take place which have an impact on its safety. Changes in the floor panel geometry with the growing mileage are especially apparent at points important from the active safety perspective. These changes include displacements of points of fixing the front suspension. the front bench and the rear suspension. An average displacement of these points ranged from 6 mm to 10 mm.

Table 4. The probabilistic model of the floor panel geometry changes for points responsible for active safety on the right side in cars used in Poland

			Points associated with passive safety				Points associated with active safety			
			Left side		Right side		Left side		Right side	
p	w	Z_{gn} [mm]	I_{zg} [mm/1000km]	P [1000km]	I_{zg} [mm/1000km]	P [1000km]	I_{zg} [mm/1000km]	P [1000km]	I_{zg} [mm/1000km]	P [1000km]
0,95	0,45296			1123		1095		395		310
0,9	0,649186			784		764		276		216
0,8	0,944761			538		525		189		149
0,7	1,194445			426		415		150		117
0,6	1,429441			356		347		125		98
0,5	1,665109	3	0,005898	305	0,006049	298	0,016771	107	0,021379	84
0,4	1,914462			266		259		93		73
0,3	2,194514			232		226		82		64
0,2	2,537272			200		195		71		55
0,1	3,034854			168		163		59		46
0,05	3,461637			147		143		52		41
0,01	4,291932			119		116		42		33

Table 5. The probabilistic model of the floor panel geometry changes for points responsible for active safety on the right side in cars used in Western Europe

			Points associated with passive safety				Points associated with active safety			
			Left side		Right side		Left side		Right side	
p	w	Z_{gn} [mm]	I_{zg} [mm/1000km]	P [1000km]	I_{zg} [mm/1000km]	P [1000km]	I_{zg} [mm/1000km]	P [1000km]	I_{zg} [mm/1000km]	P [1000km]
0,95	0,45296			2603		2508		676		517
0,9	0,649186			1816		1750		472		361
0,8	0,944761			1248		1202		324		248
0,7	1,194445			987		951		256		196
0,6	1,429441			825		795		214		164
0,5	1,665109	3	0,002544	708	0,002641	682	0,009793	184	0,012812	141
0,4	1,914462			616		593		160		122
0,3	2,194514			537		518		140		107
0,2	2,537272			465		448		121		92
0,1	3,034854			389		374		101		77
0,05	3,461637			341		328		88		68
0,01	4,291932			275		265		71		55

Displacements of points associated with passive safety were smaller. Moreover. smaller geometry changes were observed in the cars used in Western Europe. The intensity of the changes depended on the position of the points under analysis on the floor panel. For example. displacements of points on the right were as much as 33% bigger than those on the left.

The probability of reaching the maximum permissible (3 mm) geometric changes on a floor panel is accurately described by the probabilistic model in the form of the Rayleigh distribution. Its mathematical form was developed for cars used in various environmental conditions. A 3 mm displacement of a base point associated with active safety on the right side will be reached with the probability of 0.9 at a mileage of 216.000 km in cars used in Poland and at a mileage of 361.000 km in cars used in Western Europe.

The study results presented in this paper are of utilitarian importance. There is a need for introducing compulsory floor panel measurements in cars with mileages exceeding 120.000 km. The legal requirements in this regard will contribute to an improvement in road traffic safety.

References

1. Aguilar J J, Sanz M, Guillomía D, Lope M, Bueno I. Analysis, characterization and accuracy improvement of optical coordinate measurement systems for car body assembly quality control. International Journal Advanced Manufacturing Technology 2006; 30: 1174–1190,https://doi.org/10.1007/s00170-005-0143-5.
2. Ahrens G, Dellmann T, Gies S, Hecht M, Hefazi H, Henke R, Pischinger S, Schaufele R, Tegel O. Applications in Mechanical Engineering. Transport Systems, Würzburg: Springer Science Media, 2009.
3. Anderson R. Numerical and experimental evaluation of spring back in a front side member. Division of Production and Materials Engineering. Lund 2007.
4. Bera T K, Bhattacharya K, Samantaray A K. Evaluation of antilock braking system with an integrated model of full vehicle system dynamics. Simulation Modelling Practice and Theory 2011; 19: 2131–2150, https://doi.org/10.1016/j.simpat.2011.07.002.
5. Boruta G, Piętak A. Mechatronika samochodu. Układy bezpieczeństwa czynnego i biernego. Olsztyn: Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego,2012.
6. Chen H, Gong X, Hu Y F, Liu Q F, Gao B Z, Guo H Y. Automotive Control: the State of the Art and Perspective. Acta Automatica Sinica 2013: 39.
7. Duffy J E. I-Car Professional Automotive Collision Repair. New York: Delmar a division of Thomson Learning 2001.
8. Duffy J. E. Auto Body Repair Technology. Boston: Delmar Cengage Learning 2009.
9. Gonera J, Napiórkowski J. Effect of the mileage of a passenger car on changes in its body geometry. Seoul: Conference 2nd Annual International Symposium on Material Science and Engineering (ISMSE), 2018, https://doi.org/10.1063/1.5030315.
10. Guoye W, Juanli Z, Yanli F, Yanru Z. Study on ESP Control Principle of Light Off-road Vehicle Based on Brake / Drive Integrated Control. Physics Procedia 2012; 25: 834–841, https://doi.org/10.1016/j.phpro.2012.03.165.
11. Habibovic A, Davidsson J. Causation mechanisms in car-to-vulnerable road user crashes: Implications for active safety systems. Accident Analysis and Prevention 2012; 49: 493– 500, https://doi.org/10.1016/j.aap.2012.03.022.
12. Heißing B, Ersoy M. Introduction and Fundamentals, Chassis Handbook: Fundamentals, Driving Dynamics, Components, Mechatronics, Perspectives. Wiesbaden GmbH: Vieweg + Teubner Verlag – Springer Fachmedien 2011.
13. Jackowski J, Łęgiewicz J, Wieczorek M. Samochody osobowe i pochodne. Warszawa: WKiŁ 2011.
14. Konieczny Ł. Wykorzystanie metod drganiowych w ocenie stanu technicznego mechanicznych i hydropneumatycznych zawieszeń pojazdów samochodowych. Gliwice: Wyd. Politechniki Śląskiej 2015.
15. Liss M, Żółtowski B. Analiza modalna jako metoda diagnozowania stanu nadwozi samochodów osobowych. Postępy w inżynierii mechanicznej Developments in mechanical engineering 2014; 3: 21–28.
16. Livesey W A, Robinson A. The repair of vehicle bodies. Oxford 2006.
17. Michalski R, Gonera J. Kompleksowa ocena stanu technicznego nadwozia samochodu. Studia i materiały polskiego stowarzyszenia zarządzania wiedzą; Bydgoszcz: Polskie Stowarzyszenie Zarządzania Wiedzą 2011; 47: 169–183.
18. Milanés V, González C, Naranjo J E, Onieva E, Pedro T D. Electro-hydraulic braking system for autonomous vehicles. International Journal of Automotive Technology 2010; 11: 89‒95, https://doi.org/10.1007/s12239-010-0012-6.
19. Mizuno D, Suzuki S, Fujita S, Hara N. Corrosion monitoring and materials selection for automotive environments by using Atmospheric Corrosion Monitor (ACM) sensor. Corrosion Science 2014; 83: 217–225, https://doi.org/10.1016/j.corsci.2014.02.020.
20. Ozdalyan B. Development of a slip control anti-lock braking system model. International Journal of Automotive Technology 2008; 9: 71–80, https://doi.org/10.1007/s12239-008-0009-6.
21. Przybyłowski A. Stan infrastruktury transportu drogowego w Polsce z uwzględnieniem aspektów bezpieczeństwa. Research papers of Wrocław University of Economics 2014; 367: 261–271, https://doi.org/10.15611/pn.2014.367.29.
22. Raatz B. Nowoczesne technologie pomiarów i napraw karoserii powypadkowych. Ząbrowo: Oficyna Wydawnicza Troton 2005.
23. Reński A. Bezpieczeństwo czynne samochodu. Zawieszenia oraz układy hamulcowe i kierownicze. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej 2011.
24. Stack M. Trade of vehicle body repair – Vehicle Measuring. Dublin: SOLAS 2014.
25. Straky H, Kochem M, Schmitt J, Isermann R. Influences of braking system faults on vehicle dynamics. Control Engineering Practice 2003; 11: 337–343, https://doi.org/10.1016/S0967-0661(02)00301-5.
26. Tobota A. Naprawy powypadkowe nadwozi, a bezpieczeństwo. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej 1997.
27. Wallentowitz H. Virtuelle Fahrzeugentwicklung–Netzwerkeals Voraussetzungenzur Problemlosung. Berlin: Erfolg in Netzwerken; Springer – Verlag Berlin Heidelberg 2002.

Jarosław GONERA
Jerzy NAPIÓRKOWSKI
Department of Vehicle and Machinery Construction and Operation
Faculty of Technical Sciences
University of Warmia and Mazury in Olsztyn
Oczapowskiego str. 11, 10-716 Olsztyn, Poland

E-mails: jaroslaw.gonera@uwm.edu.pl, jerzy.napiorkowski@uwm.edu.pl

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.

No text or photograph published in „Maintenance and Reliability" can be reproduced without the Editor's written consent.

Wydawca:
Polskie Naukowo Techniczne
Towarzystwo Eksploatacyjne
Warszawa

Publisher:
Polish Maintenance Society
Warsaw

Członek:
Europejskiej Federacji
Narodowych Towarzystw
Eksploatacyjnych

Member of:
European Federation of National
Maintenance Societies

Patronat naukowy:
Polska Akademia Nauk
Oddział Lublin

Scientific Supervision:
Polish Academy of Sciences
Branch in Lublin