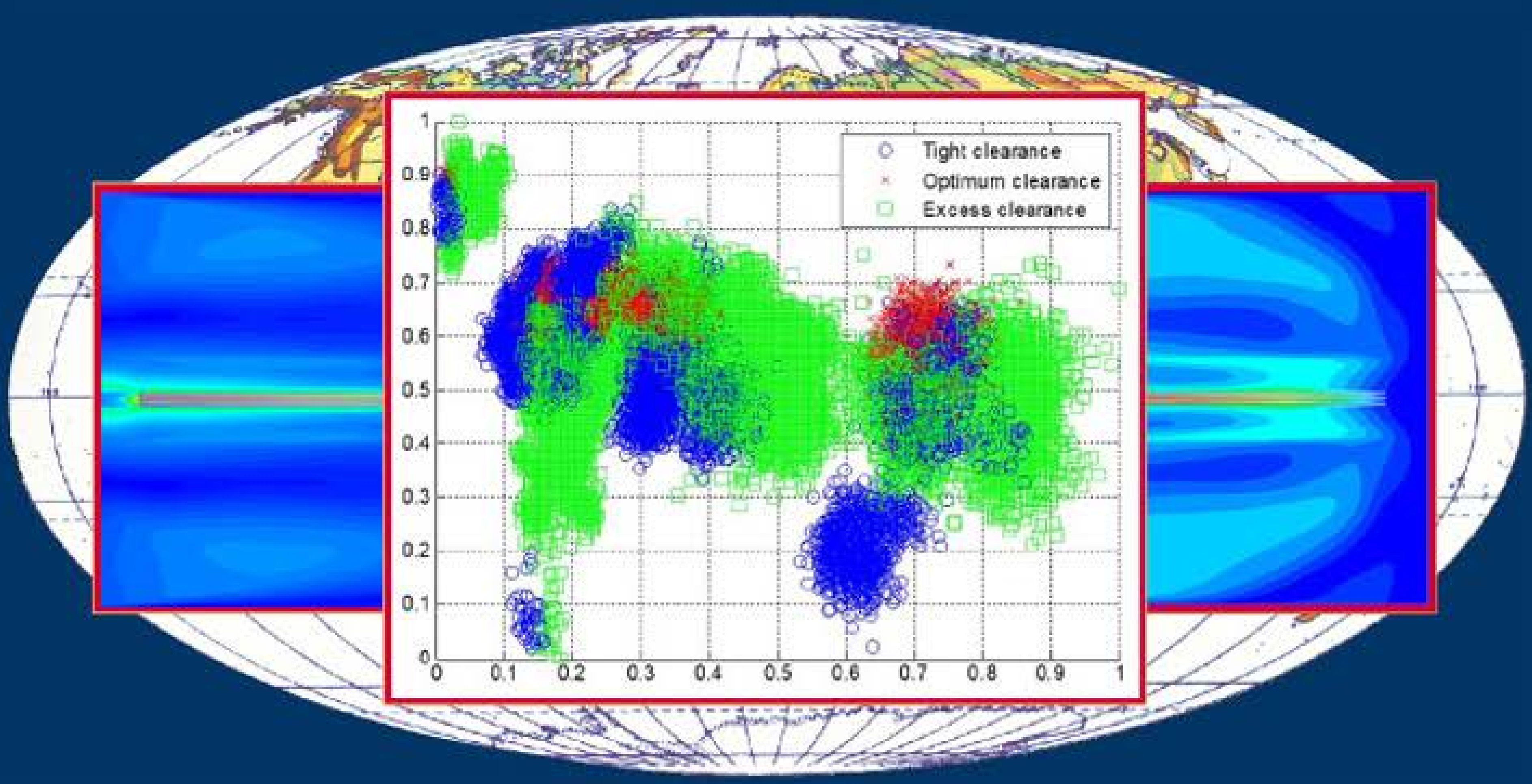
Vol. 22. No 2, 2020

ISSN 1507-2711
Cena: 25 zł

EKSPLOATACJA I NIEZAWODNOŚĆ
MAINTENANCE AND RELIABILITY



Polskie Naukowo Techniczne Towarzystwo Eksploatacyjne
Warszawa

Polish Maintenance Society
Warsaw

VALIS D, FORBELSKÁ M, VINTR Z. **Forecasting study of mains reliability based on sparse field data and perspective state space models**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 179–191, http://dx.doi.org/10.17531/ein.2020.2.1.

The elements of critical infrastructure have to meet demanding dependability, safety and security requirements. The article deals with the prognosis of water mains reliability while using sparse irregular filed data. The data are sparse because the only thing we know is the number of mains failures during a given month. Since it is possible to transform the data into a typical reliability measure (rate of failure occurrence – ROCOF), we can examine the course of this measure development in time. In order to model and predict the ROCOF development, we suggest novel single and multiple error state space models. The results can be used for i) optimizing mains operation and maintenance, ii) estimating life cycle cost, and iii) planning crisis management.

VALIS D, FORBELSKÁ M, VINTR Z. **Prognozowanie niezawodności elementów sieci wodociągowej na podstawie rzadkich danych terenowych i modeli przestrzeni stanów**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 179–191, http://dx.doi.org/10.17531/ein.2020.2.1.

Elementy infrastruktury krytycznej muszą spełniać wysokie wymagania w zakresie niezawodności, bezpieczeństwa i ochrony. Artykuł dotyczy prognozowania niezawodności sieci wodociągowej przy wykorzystaniu nieregularnie rejestrowanych rzadkich danych. Wykorzystane w pracy dane są rzadkie, ponieważ dostarczają jedynie informacji na temat liczby uszkodzeń wodociągu w danym miesiącu. Przekształcenie tych danych w typową miarę niezawodności (wskaźnik występowania uszkodzeń – ROCOF), pozwala zbadać przebieg rozwoju tej miary w czasie. Rozwój ROCOF można modelować i przewidywać za pomocą zaproponowanych w pracy innowacyjnych modeli przestrzeni stanów uwzględniających pojedynczy błąd lub wiele błędów. Uzyskane wyniki można wykorzystać do i) optymalizacji pracy i eksploatacji sieci wodociągowej, ii) szacowania kosztów cyklu życia, oraz iii) planowania zarządzania kryzysowego.

KOZŁOWSKI E, BORUCKA A, ŚWIDERSKI A. **Application of the logistic regression for determining transition probability matrix of operating states in the transport systems**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 192–200, http://dx.doi.org/10.17531/ein.2020.2.2.

Transport companies can be regarded as a technical, organizational, economic and legal transport system. Maintaining the quality and continuity of the implementation of transport requisitions requires a high level of readiness of vehicles and staff (especially drivers). Managing and controlling the tasks being implemented is supported by mathematical models enabling to assess and determine the strategy regarding the actions undertaken. The support for managing processes relies mainly on the analysis of sequences of the subsequent activities (states). In many cases, this sequence of activities is modelled using stochastic processes that satisfy Markov property. Their classic application is only possible if the conditional probability distributions of future states are determined solely by the current operational state. The identification of such a stochastic process relies mainly on determining the probability matrix of interstate transitions. Unfortunately, in many cases the analyzed series of activities do not satisfy Markov property. In addition, the occurrence of the next state is affected by the length of time the system remains in the specified operating state. The article presents the method of constructing the matrix of probabilities of transitions between operational states. The values of this matrix depend on the time the object remains in the given state. The aim of the article was to present an alternative method of estimating the parameters of this matrix in a situation where the studied series does not satisfy Markov property. The logistic regression was used for this purpose.

KOZŁOWSKI E, BORUCKA A, ŚWIDERSKI A. **Zastosowanie regresji logistycznej do wyznaczania macierzy prawdopodobieństw przejść stanów eksploatacyjnych w systemach transportowych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 192–200, http://dx.doi.org/10.17531/ein.2020.2.2.

Przedsiębiorstwa transportowe mogą być traktowane jako wyodrębniony pod względem technicznym, organizacyjnym, ekonomicznym i prawnym system transportowy. Zachowanie jakości i ciągłości realizacji zleceń przewozowych wymaga wysokiego poziomu gotowości pojazdów oraz personelu (szczególnie kierowców). Kontrolowanie i sterowanie realizowanymi zadaniami wspierane jest modelami matematycznymi, umożliwiającymi ocenę i określenie strategii dotyczącej podejmowanych działań. Wsparcie procesów zarządzania polega głównie na analizie sekwencji kolejnych, realizowanych czynności (stanów). W wielu przypadkach taki ciąg czynności jest modelowany za pomocą procesów stochastycznych, spełniających własność Markowa. Ich klasyczne zastosowanie możliwe jest tylko w przypadku, gdy warunkowe rozkłady prawdopodobieństwa przyszłych stanów są określone wyłącznie przez bieżący stan eksploatacyjny. Identyfikacja takiego procesu stochastycznego polega głównie na wyznaczeniu macierzy prawdopodobieństw przejść międzystanowych. Niestety w wielu przypadkach analizowane ciągi czynności nie spełniają własności Markowa. Dodatkowo, na wystąpienie kolejnego stanu wpływa długość interwału czasowego pozostania systemu w określonym stanie eksploatacyjnym. W artykule przedstawiono metodę konstrukcji macierzy prawdopodobieństw przejść pomiędzy stanami eksploatacyjnymi. Wartości tej macierzy zależą od czasu przebywania obiektu w danym stanie. Celem artykułu było zaprezentowanie alternatywnej metody estymacji parametrów tej macierzy w sytuacji, gdy badany szereg nie spełnia własności Markowa. Wykorzystano w tym celu regresję logistyczną.

MACIÁN V, TORMOS B, BASTIDAS S, PÉREZ T. **Improved fleet operation and maintenance through the use of low viscosity engine oils: fuel economy and oil performance**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 201–211, http://dx.doi.org/10.17531/ein.2020.2.3.

For heavy-duty vehicles and road transportation, fuel consumption and associated CO_2 emissions have been of great concern, which has led to the development and implementation of technologies to reduce their impact on the environment. Low viscosity engine oils have arisen as one proven cost-effective solution to increase the engine efficiency; however, for the heavy-duty vehicle segment, engine protection against wear is a priority for end-users, and therefore there is some reluctance to the use of that new oil formulations. In this study, eight lubricant oils, representative of the HTHS viscosity reduction that heavy-duty oils have been undergoing and new API CK-4 and FA-4 categories, were evaluated for fuel economy, oil performance and engine wear, in a long-term test involving a fleet of 49 heavy-duty vehicles of four different engine technologies, some of them with diesel fuel and others with compressed natural gas. Results of fuel economy were positive for most of the buses' models. Regarding oil performance and wear, most of the formulations were found to be suitable for extended oil drain intervals (ODI); and although no alarming results were found, overall performance of the formulations of the fourth stage could lead to significant wear if the oil drain interval is extended. In this study, it should be noted that some of the information has been presented by the authors in other publications, here they are presented with the purpose of complementing the new results and summarize the entire test.

MACIÁN V, TORMOS B, BASTIDAS S, PÉREZ T. **Poprawa efektywności eksploatacji floty dzięki zastosowaniu olejów silnikowych o niskiej lepkości: oszczędność paliwa i wydajność oleju**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 201–211, http://dx.doi.org/10.17531/ein.2020.2.3.

W przypadku pojazdów o dużej ładowności, i transportu drogowego w ogóle, ważny problem stanowi zużycie paliwa i związana z nim emisja CO2, które wymagają opracowywania i wdrażania technologii zmniejszających ich wpływ na środowisko. Jednym ze sprawdzonych i finansowo korzystnych rozwiązań w tym zakresie są oleje silnikowe o niskiej lepkości, które zwiększają wydajność silnika. Jednak w segmencie pojazdów ciężkich, priorytetem dla użytkowników końcowych jest ochrona silnika przed zużyciem, co pociąga za sobą niechęć do stosowania tych nowych preparatów olejowych. W pracy, przedstawiono badania ośmiu olejów smarowych o obniżonej lepkości wysokotemperaturowej HTHS reprezentatywnych dla produkowanych obecnie kategorii olejów do pojazdów ciężkich, z uwzględnieniem nowych kategorii oleju API CK-4 i FA-4. Oleje oceniano pod kątem oszczędności paliwa, wydajności oleju i zużycia silnika w badaniu długoterminowym obejmującym flotę 49 autobusów o silnikach opartych na różnych technologiach, z których część była zasilana olejem napędowym a część sprzężonym gazem ziemnym. Wyniki dotyczące oszczędności zużycia paliwa były pozytywne dla większości modeli badanych autobusów. Jeśli chodzi o wydajność oleju i zużycie silnika, większość preparatów okazała się być przystosowana do dłuższych okresów wymiany oleju; chociaż nie zaobserwowano niepokojących wyników, to jednak ogólna wydajność preparatów w czwartym etapie testu, mogłaby prowadzić do znacznego zużycia silnika przy wydłużeniu okresu wymiany oleju. Część przedstawionych danych publikowaliśmy już w innych pracach. Niniejszy artykuł stanowi uzupełnienie poprzednich wyników oraz podsumowanie całego badania.

CHUDZIK A, WARDA B. **Fatigue life prediction of a radial cylindrical roller bearing subjected to a combined load using FEM**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 212–220, http://dx.doi.org/10.17531/ein.2020.2.4.

The article presents the results of studies on the impact of a combined load of a radial cylindrical roller bearing for its predicted fatigue life. The distributions of maximum equivalent subsurface stresses and their depths, necessary during calculations of fati-

CHUDZIK A, WARDA B. **Prognozowanie trwałości zmęczeniowej promieniowego łożyska walcowego poddanego złożonemu obciążeniu z wykorzystaniem MES**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 212–220, http://dx.doi.org/10.17531/ein.2020.2.4.

W artykule zaprezentowano wyniki badań wpływu złożonego obciążenia promieniowego łożyska walcowego na jego prognozowaną trwałość zmęczeniową. Rozkłady maksymalnych zastępczych naprężeń podpowierzchniowych oraz głębokości ich występowania, gue life, were determined using the finite element method, using the basic package of the ANSYS program. The calculations took into account the geometrical parameters of the bearing, including radial clearance and the shape of the rolling elements generators. The calculation results showed that the axial load of the radial cylindrical roller bearing and the tilt of the rollers associated with its operation reduces fatigue life. The obtained results were compared with the results of calculations according to the SKF catalogue method, obtaining satisfactory compliance.

niezbędne podczas obliczeń trwałości zmęczeniowej, określono za pomocą metody elementów skończonych, z wykorzystaniem pakietu podstawowego programu ANSYS. W obliczeniach uwzględniono geometryczne parametry łożyska, w tym luz promieniowy i kształt tworzących elementów tocznych. Wyniki obliczeń wykazały, że obciążenie osiowe promieniowego łożyska walcowego i przechylenie wałeczków towarzyszące jego działaniu powoduje zmniejszenie trwałości zmęczeniowej. Otrzymane wyniki porównano z wynikami obliczeń według katalogowej metody firmy SKF, otrzymując zadowalającą zgodność.

LIU Y, WANG y, FAN Z, HOUZ, ZHANG S, CHEN X. **Lifetime prediction method for MEMS gyroscope based on accelerated degradation test and acceleration factor model**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 221–231, http://dx.doi.org/10.17531/ein.2020.2.5.

The reliability analysis of MEMS gyroscope under long-term operating condition has become an urgent requirement with the enlargement of its application scope and the requirement of good durability. In this study we propose a lifetime prediction method for MEMS gyroscope based on accelerated degradation tests (ADTs) and acceleration factor model. Firstly, the degradation characteristic (bias instability) is extracted based on Allan variance. The effect of temperature stress on the degradation rate of bias instability is analyzed, and it shows that the degradation rate of bias instability would increase with the increase of the temperature. Secondly, the ADTs of MEMS gyroscope are designed and conducted, the degradation model of MEMS gyroscope is established based on the output voltage of MEMS gyroscope and Allan variance. Finally, the acceleration factor model of MEMS gyroscope under temperature stress is derived, and the lifetime of the MEMS gyroscope is predicted based on two group tests data under high stress level. The results show that the lifetime calculated by the acceleration factor model and mean lifetime under high stress levels is close to the mean lifetime calculated by the linear equation at normal temperature stress.

LIU Y, WANG y, FAN Z, HOUZ, ZHANG S, CHEN X. **Metoda prognozowania czasu pracy żyroskopu MEMS na podstawie testu przyspieszonej degradacji i modelu współczynnika przyspieszenia**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 221–231, http://dx.doi.org/10.17531/ein.2020.2.5.

Analiza niezawodności żyroskopu MEMS w warunkach długotrwałej pracy stała się pilną koniecznością wraz z rozszerzeniem zakresu jego zastosowania i wprowadzeniem wymogu dobrej trwałości. W niniejszym artykule, zaproponowano metodę prognozowania czasu pracy żyroskopu MEMS w oparciu o testy przyspieszonej degradacji i model współczynnika przyspieszenia. W pierwszej kolejności, wyznaczono charakterystykę degradacji (niestabilność wskazań) na podstawie wariancji Allana. Analizowano wpływ naprężenia cieplnego na szybkość degradacji w zakresie niestabilności wskazań. Analiza wykazała, że szybkość degradacji wzrastała wraz ze wzrostem temperatury. Następnie, opracowano i przeprowadzono testy przyspieszonej degradacji żyroskopu MEMS, a model jego degradacji ustalono na podstawie napięcia wyjściowego żyroskopu i wariancji Allana. Na koniec, wyprowadzono model współczynnika przyspieszenia dla żyroskopu MEMS w warunkach naprężenia cieplnego, a żywotność żyroskopu prognozowano na podstawie danych z dwóch testów grupowych przeprowadzonych w warunkach wysokiego naprężenia. Wyniki pokazują, że czas pracy obliczony na podstawie modelu współczynnika przyspieszenia i średni czas pracy przy wysokich poziomach naprężeń są zbliżone do średniego czasu pracy obliczonego na podstawie równania liniowego przy normalnym naprężeniu cieplnym.

CZAPP S, SZULTKA S, RATKOWSKI F, TOMASZEWSKI A. **Risk of power cables insulation failure due to the thermal effect of solar radiation**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 232–240, http://dx.doi.org/10.17531/ein.2020.2.6.

Low-voltage, as well as high-voltage power cable lines, are usually buried in the ground. The ampacity of the power cables in the ground mainly depends on the thermal resistivity of the soil, which may vary in a wide range. A common practice in power cable systems performance is to supply them from a pole of an overhead line. If so, a section of the line is located in free air and can be directly exposed to solar radiation. In some cases, the ampacity of power cables placed in free air is lower than in the ground. Differences in ampacities can be very high if thermal resistivity of the soil is very low, and simultaneously solar irradiation of cables in air occurs. This paper presents the risk of power cables overheating and in consequence the risk of their failure, when part of the underground power cable line is placed in free air. Temperature distribution of cables in the air (with and without solar radiation) for various load currents is presented. Thermal endurance of power cables insulation, operating with the overheating, is estimated.

CZAPP S, SZULTKA S, RATKOWSKI F, TOMASZEWSKI A. **Ryzyko uszkodzenia cieplnego izolacji kabli elektroenergetycznych z powodu oddziaływania promieniowania słonecznego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 232–240, http://dx.doi.org/10.17531/ein.2020.2.6.

Linie kablowe zarówno niskiego, jak i wysokiego napięcia zwykle buduje się jako podziemne. Obciążalność kabli układanych w ziemi w znacznym stopniu zależy od rezystywności cieplej gruntu, a może się ona zmieniać w bardzo szerokim zakresie. Obecnie powszechną praktyką jest zasilanie linii kablowych z linii napowietrznych, co sprawia, że pewien odcinek linii kablowej znajduje się w powietrzu i może być poddany bezpośredniemu oddziaływaniu promieniowania słonecznego. W pewnych przypadkach obciążalność prądowa długotrwała kabli w powietrzu jest niższa niż w ziemi – różnice w tej obciążalności mogą być bardzo duże, jeżeli grunt ma niską rezystywność cieplną, a na odcinek linii w powietrzu oddziałuje promieniowanie słoneczne. W artykule przedstawiono problem przegrzania kabli elektroenergetycznych, gdy przyjęta obciążalność linii kablowej wynika z warunków dla ułożenia w ziemi, a na pewnym odcinku linia jest umieszczona w powietrzu. Przedstawiono rozkłady temperatury kabli w powietrzu (z uwzględnieniem i bez uwzględnienia promieniowania słonecznego) dla różnych prądów obciążenia kabli. Oszacowano trwałość termiczną izolacji kabli, mających przez znaczny przedział czasu temperaturę wyższą niż dopuszczalna długotrwale.

YU G, DU Y, YAN L, REN F. **Stress-strength interference-based importance for series systems considering common cause failure**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 241–252, http://dx.doi.org/10.17531/ein.2020.2.7.

Series systems, whose structures are simple, are widely discovered in practical engineering, but the interdependency between the components is complex, such as common cause failure. With the consideration of the components' strength, this paper focuses on ranking the importance measure of components considering the common cause failure based on the stress-strength interference (SSI) model. The weakest component can be identified by integrating the SSI model with the importance measure when the strength mean and variance of the component under the load stress is known. Firstly, the analytic methods are proposed to calculate the SSI-based importance of components in the series systems. Then, the monotonicity of SSI-based importance is analyzed by changing the strength mean or strength variance of one component. The results show that the SSI-based importance of components, whose parameters are changed, will reduce monotonically with the increase of strength mean or increase monotonically with the increase of strength variance. Finally, a component replacement method is developed based on the rules that both the importance of replaced component and the importance ranks should be unchanged after the replacement. SSI-based importance can help engineers to make maintenance decisions, and the component replacement method can increase the diversity of spare parts by finding the equivalent components.

YU G, DU Y, YAN L, REN F. **Ocena opartej na modelu obciążeniowo-wytrzymałościowym ważności elementów systemu szeregowego z uwzględnieniem uszkodzeń wywołanych wspólną przyczyną**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 241–252, http://dx.doi.org/10.17531/ein.2020.2.7.

Systemy szeregowe, które są szeroko stosowane w praktyce inżynieryjnej, charakteryzują się prostą strukturą, jednak współzależności między ich elementami są złożone, czego przykładem są uszkodzenia wywołane wspólną przyczyną. Rozważając wytrzymałości składowych systemu, opracowano metodę szeregowania miar ważności składowych z uwzględnieniem uszkodzeń wywołanych wspólną przyczyną. Metoda ta pozwala zidentyfikować najsłabsze ogniwo systemu. Miarę istotności zintegrowano z modelem obciążeniowo-wytrzymałościowym (SSI), biorąc pod uwagę średnią i wariancję wytrzymałości elementu pod obciążeniem. W pierwszym kroku opracowano metody analityczne pozwalające na obliczanie opartej na SSI ważności elementów w systemach szeregowych. Następnie analizowano monotoniczność opartej na SSI ważności zmieniając średnią lub wariancję wytrzymałości jednego z elementów. Wyniki pokazują, że mierzona w oparciu o SSI ważność elementów, których parametry są zmieniane, maleje monotonicznie wraz ze wzrostem średniej wytrzymałości lub rośnie monotonicznie wraz ze wzrostem wariancji wytrzymałości. Na podstawie przeprowadzonych badań, opracowano metodę wymiany części, opartą na zasadzie polegającej na tym, że zarówno ważność zastąpionego elementu, jak i rangi ważności powinny pozostać niezmienione po wymianie. Możliwość określania ważności opartej na modelu SSI może pomóc inżynierom w podejmowaniu decyzji dotyczących konserwacji, zaś proponowana metoda wymiany elementów systemu pozwala zwiększyć różnorodność części zamiennych poprzez znalezienie równoważnych elementów.

CHE H, ZENG S, GUO J. **A reliability model for load-sharing k-out-of-n systems subject to soft and hard failures with dependent workload and shock effects**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 253–264, http://dx.doi.org/10.17531/ein.2020.2.8.

A component in a k-out-of-n system may experience soft and hard failures resulting from exposure to natural degradation and random shocks. Due to load-sharing characteristics, once a component fails, the surviving components share an increased workload, which increases their own degradation rates. Moreover, under the larger workload, random shocks may cause larger abrupt degradation increments and larger shock sizes. Therefore, the system experiences the dependent workload and shock effects (DWSEs). Such dependence will cause the load-sharing system to fail more easily, though it is often not considered in existing methods. In this paper, to evaluate the system reliability more accurately, we develop a novel reliability model for load-sharing k-out-of-n systems with DWSEs. In the model, the joint probability density function of shock effects to soft and hard failures is developed to describe the DWSEs on a component. To derive an analytical expression of system reliability with load-sharing characteristics and DWSEs, conditional probability density function is used to model the random component failure times. A load-sharing Micro-Electro-Mechanical System (MEMS) is then utilized to illustrate the effectiveness of the reliability model

CHE H, ZENG S, GUO J. **Model niezawodności dla systemów typu k-z-n z podziałem obciążenia podlegających uszkodzeniom parametrycznym i katastroficznym, w których zachodzi zależność między obciążeniem pracą a skutkami obciążeń losowych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 253–264, http://dx.doi.org/10.17531/ein.2020.2.8.

Element systemu k-z-n może ulegać uszkodzeniom parametrycznym i katastroficznym wynikającym z ekspozycji na naturalne procesy degradacji i obciążenia losowe. Ze względu na równomierny podział obciążenia między wszystkie elementy systemu, gdy jeden element ulega awarii, obciążenie pracą przypadające na pozostałe komponenty zwiększa się, podnosząc tempo degradacji każdego z nich. Ponadto, przy większym obciążeniu pracą, obciążenia losowe mogą powodować większe nagłe przyrosty degradacji i zwiększać rozmiary obciążeń. Mówi się wtedy o istnieniu zależności między obciążeniem pracą a skutkami obciążeń losowych (dependent workload and schock effects (DWSE). Taka zależność powoduje, że system z podziałem obciążeń łatwiej ulega uszkodzeniom. Fakt ten jest często pomijany w obecnie stosowanych metodach oceny niezawodności. W niniejszym artykule przedstawiamy nowatorski model oceny niezawodności systemów k-z-n z podziałem obciążenia i zależnością DWSE, który pozwala dokładniej ocenić niezawodność takich systemów. W modelu, opracowano wspólną funkcję gęstości prawdopodobieństwa skutków obciążeń losowych dla uszkodzeń parametrycznych i katastroficznych, która pozwala opisać zależność DWSE dla elementu systemu. Aby wyprowadzić analityczne wyrażenie niezawodności systemu z podziałem obciążenia i DWSE, do modelowania czasów losowych uszkodzeń elementów systemu wykorzystano funkcję warunkowej gęstości prawdopodobieństwa. Skuteczność modelu niezawodności zilustrowano na przykładzie układu mikroelektromechanicznego z podziałem obciążenia (MEMS).

NIKOŃCZUK P, ROSOCHACKI W. **The concept of reliability measure of recuperator in spray booth**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 265–271, http://dx.doi.org/10.17531/ein.2020.2.9.

Overspray sediments deposited on the recuperator fins gradually reduce the cross-section of the recuperator channels. The result of this process is the increase in airflow resistance and thermal resistance during heat transfer. Both phenomena have a negative impact on the reliability of the device. This paper presents the concept of recuperator reliability measures. For this purpose, the essential requirement of reliability (indestructibility) was formulated and damage was defined by identifying it with the loss of air flow reserve and reserve of heat transfer efficiency. On this basis ability features of the heat recovery unit were assessed. Limits of features and critical time of recuperator loss of ability were also assessed.

NIKOŃCZUK P, ROSOCHACKI W. **Koncepcja miary niezawodności rekuperatora kabiny lakierniczej**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 265–271, http://dx.doi.org/10.17531/ein.2020.2.9.

Odkładające się na lamelach rekuperatora osady lakiernicze powodują stopniowe zmniejszanie przekroju poprzecznego kanałów rekuperatora. Skutkiem tego procesu są wzrosty oporów przepływu powietrza oraz oporu termicznego przy wymianie ciepła. Oba zjawiska wpływają negatywnie na niezawodność urządzenia. W artykule przedstawiono koncepcję miary niezawodności rekuperatora. W tym celu sformułowano podstawowe wymaganie niezawodnościowe (nieuszkadzalność) oraz zdefiniowano uszkodzenia utożsamiając je z utratą zapasu strumienia powietrza oraz zapasu efektywności wymiany ciepła. Na tym tle określono cechy zdatności urządzenia, granice ich obszarów oraz krytyczny czas utraty zdatności rekuperatora.

LIU C, KRAMER A, NEUMANN S. **Reliability assessment of repairable phased-mission system by Monte Carlo simulation based on modular Sequence-Enforcing Fault Tree model**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 272–281 http://dx.doi.org/10.17531/ein.2020.2.10.

Phased-mission system (PMS) is the system subject to multiple, consecutive and non-overlapping tasks. Much more complicated problems will be confronted when the PMS is repairable since the repairable system could perform the multi-phases mission with more diversity requirements. Besides, various maintenance strategies will directly influence the reliability analysis procedure. Most researches investigate those repairable PMSs that carry out the multi-phases mission with deterministic phase durations, and the mission fails once the system switches from up to down. In this case, one common maintenance strategy is that failed components are repairable as long as the system keeps in up state. However, many practical systems (e.g., construction machinery, agricultural machinery) may be involved in such multi-phases mission, which has uncertain phase durations but limited by a maximum mission time, within which failed components can be unconditional repaired, and the system can be restored from down state. Comparing with the former type of repairable PMS, the latter will also concern phase durations dependence, and both the system and components included have the state bidirectional transition. This paper makes new contributions to the reliability assessment of repairable PMSs by proposing a novel SEFT-MC method. Two types of repairable PMS mentioned above are considered. In our method, a specific sequence-enforcing fault tree (SEFT) is proposed to correctly depict failure logical relationships between the system and components included. In order to transfer the graphical fault tree (no matter its size and complexity) into a modular reliability model used in Monte Carlo (MC) simulation, an improved linear algebra representation (I-LAR) approach is introduced. Finally, a numerical example including two cases corresponding to the two types of repairable PMS is presented to validate the proposed method.

LIU C, KRAMER A, NEUMANN S. **Ocena niezawodności naprawialnego systemu z misjami okresowymi za pomocą symulacji Monte Carlo w oparciu o modułowy model drzewa niezdatności z bramkami SEQ**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 272–281 http://dx.doi.org/10.17531/ein.2020.2.10.

System z misjami okresowymi (phased-mission system, PMS) to system, który wykonuje wiele następujących po sobie i nienakładających się na siebie zadań. W przypadku naprawialnych systemów PMS, analiza niezawodności jest o wiele bardziej skomplikowana, ponieważ system naprawialny może wykonywać misje wielofazowe o bardziej różnorodnych wymaganiach. Poza tym systemy takie wymagają zastosowania różnych strategii utrzymania ruchu, co ma bezpośredni wpływ na procedurę analizy niezawodności. Większość badaczy bada naprawialne systemy PMS, które wykonują misje wielofazowe, w których czas trwania fazy jest wielkością deterministyczną, a misja kończy się niepowodzeniem, gdy system przechodzi ze stanu zdatności do stanu niezdatności W takich przypadkach najczęściej przyjmuje się, że uszkodzone elementy można naprawić o ile system pozostaje w stanie zdatności. Jednak wiele systemów stosowanych w praktyce (t.j. maszyny budowlane czy maszyny rolnicze) może wykonywać misje wielofazowe, w których czas trwania fazy jest wielkością niepewną, ograniczoną jedynie przez maksymalny czas trwania misji, w którym to czasie uszkodzone komponenty mogą być bezwarunkowo naprawiane, dzięki czemu system może zostać przywrócony do stanu zdatności. W porównaniu z pierwszym rodzajem naprawialnego PMS, w drugim, czasy trwania faz są zależne od siebie. Ponadto, w systemie tego typu, zarówno poszczególne elementy, jak i cały system mogą przechodzić ze stanu zdatności do stanu niezdatności i odwrotnie. Niniejsza praca wnosi nowy wkład w ocenę niezawodności naprawialnych systemów PMS, proponując nowatorską metodę, która polega na wykorzystaniu dynamicznego drzewa niezdatności do przeprowadzenia symulacji Monte Carlo (SEFT-MC). Rozważane są dwa wymienione powyżej typy naprawialnego PMS. W naszej metodzie zaproponowano drzewo niezdatności z bramkami SEQ (SEFT), które pozwala poprawnie zobrazować logiczne zależności między systemem a jego komponentami w zakresie uszkodzeń. Do przeniesienia graficznego drzewa niezdatności (bez względu na jego rozmiar i złożoność) do modułowego modelu niezawodności wykorzystywanego w symulacji Monte Carlo, zastosowano udoskonaloną metodę reprezentacji algebry liniowej (I-LAR). Poprawność proponowanej metody wykazano na przykładzie numerycznym obejmującym dwa przypadki odpowiadające dwóm omawianym typom naprawialnego PMS.

CUI X, LI T, WANG S, SHI J, MA Z. **Reliability modeling based on power transfer efficiency and its application to aircraft actuation system**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 282–296, http://dx.doi.org/10.17531/ein.2020.2.11.

The power transfer systems (PTS) has special reliability properties, including multiple states and fault dependence. Consequently, traditional binary-state reliability modeling methods cannot accurately evaluate the reliability of PTS. In order to resolve the contradiction between terminal energy demand and power transfer capability of PTS, this paper proposes a novel multi-state reliability model based on power transfer efficiency (PTE) for reliability evaluation of PTS. The multi-state model caused by performance degradation based on PTE is considered in this paper. In addition, the failure correlation in virtue of the system structure and energy allocation mechanism is analyzed in the proposed model, and the corresponding reliability evaluation result is obtained under different terminal energy requirements. The approach is verified on the example of a dual hydraulic actuation system (DHAS), in which the stochastic model based on the generalized stochastic Petri nets (GSPNs) is established and combined with the power transfer capability via universal generating function (UGF). Though changing flow rate to face the degradation rate of hydraulic pump, the reliability assessment of DHAS based on the proposed reliability model is effective and accurate.

CUI X, LI T, WANG S, SHI J, MA Z. **Model niezawodności oparty na wydajności przesyłu energii i jego zastosowanie do oceny lotniczego układu hydrauliki siłowej**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 282–296, http://dx.doi.org/10.17531/ein.2020.2.11.

Układy przesyłu energii (power transfer systems, PTS) charakteryzują się szczególnymi właściwościami niezawodnościowymi, w tym wielostanowością i zależnością między błędami. W związku z tym, tradycyjne metody modelowania niezawodności, które sprawdzają się w przypadku systemów dwustanowych, nie pozwalają na dokładną ocenę niezawodności PTS. W przedstawionej pracy zaproponowano nowatorski model niezawodności systemu wielostanowego, który do oceny niezawodności PTS wykorzystuje dane o wydajności przesyłu energii (PTE). Model ten wiąże niezawodność zarówno z zapotrzebowaniem na energię końcową jak i zdolnością przesyłową PTS. Rozważano model wielostanowy opisujący proces degradacji komponentów systemu w oparciu o PTE. W proponowanym modelu analizowano korelacje między uszkodzeniami w świetle struktury systemu i mechanizmu alokacji energii, a niezawodność oceniano dla różnych stopni zapotrzebowania na energię końcową. Podejście to zweryfikowano na przykładzie podwójnego układu hydrauliki siłowej (DHAS), dla którego ustalono model stochastyczny oparty na uogólnionych stochastycznych sieciach Petriego (GSPN), który łączono ze zdolnością przesyłową za pomocą uniwersalnej funkcji tworzącej (UGF). Badania pompy hydraulicznej prowadzone dla różnych prędkości przepływu i różnych szybkości degradacji wykazały, iż ocena niezawodności DHAS na podstawie proponowanego modelu cechuje się skutecznością i trafnością.

WYMYSŁOWSKI A, JANKOWSKI K. **Strength analysis of solder joints used in microelectronics packaging**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 297–305, http://dx.doi.org/10.17531/ein.2020.2.12.

The aim of the research was the problem of damage accumulation for solder alloys used in microelectronics packaging due to creep and fatigue as a result of a combined profile of loading conditions. The selected failure modes affect the lifetime of contemporary electronic equipment. So far the research activities are focused on a single failure mode and the problem of their interaction is often omitted. Taking into account the failure modes interaction would allow more precise lifetime prediction of the contemporary electronic equipment and/or would allow for reduction of time required for reliability tests. Within the taken research framework the reliability analysis of solder joints was conducted for the Sn63Pb37 solder alloy using the Hot Bump Pull method. The results of the presented research contain: reliability tests, statistical analysis and the problem of a damage accumulation due to a combined profile of loading conditions.

WYMYSŁOWSKI A, JANKOWSKI K. **Badania wytrzymałości połączeń lutowanych stosowanych w montażu w mikroelektronice**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 297–305, http://dx.doi.org/10.17531/ein.2020.2.12.

Celem badań był problem kumulacji uszkodzeń dla stopów lutowniczych stosowanych w montażu w mikroelektronice w wyniku zmęczenia i pełzania na skutek złożonego profilu obciążeń. Wybrane rodzaje uszkodzeń przyczyniają się do ograniczenia czasu życia współczesnych urządzeń elektronicznych. Aktualnie prowadzi się badania z wykorzystaniem jednego rodzaju uszkodzeń i często pomijany jest problem ich wzajemnej interakcji. Uwzględnienie problemu wzajemnej interakcji pozwoliłoby na bardziej precyzyjne prognozowanie bezawaryjnego czasu pracy współczesnych urządzeń elektronicznych i/lub przyspieszenie testów niezawodnościowych. W ramach zrealizowanych badań przeprowadzono analizę wytrzymałości połączeń lutowanych dla stopu lutowniczego Sn63Pb37 z wykorzystaniem metody Hot Bump Pull. Wyniki przedstawionych badań obejmują: analizę wytrzymałości, analizę statystyczną oraz problem kumulacji uszkodzeń w wyniku złożonego profilu obciążeń.

PAULAUSKAS V, FILINA-DAWIDOWICZ L, PAULAUSKAS D. **Ships speed limitations for reliable maintenance of the quay walls of navigation channels in ports**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 306–315, http://dx.doi.org/10.17531/ein.2020.2.13.

There is a number of ports where approach or inside navigation channels are located close to the quay walls. In difficult hydro-meteorological conditions appropriate speed of ship is needed to keep proper ship's steering while passing through channel. The ships that pass near the quay walls with high speed create high interaction forces on moored ships and negatively interact on their mooring equipment and quay walls. Ports should ensure relevant maintenance and reliability of quay walls and ships' mooring equipment. That is why investigation of the ships interaction forces during ship passing near the ships moored to quay walls is very important to find limitations of the passing ship speed depending on passing ship's parameters, distances and environmental conditions to provide reliable maintenance of navigation channel. In the article the conditions of dynamic forces caused by passing ships are investigated, including possible external forces influencing on moored ships, mooring equipment and quay walls. The methodology to assess the forces exerted on ship moored to quay wall by ship passing close to them is created. On the basis of the case study analysis results, the recommendations for the limitations to ships passing near the ships moored to quay walls were proposed that will allow providing relevant maintenance and reliability of navigation channel and its infrastructure.

PAULAUSKAS V, FILINA-DAWIDOWICZ L, PAULAUSKAS D. **Ograniczenia prędkości statków do niezawodnego utrzymania ścian nabrzeży zlokalizowanych przy kanałach nawigacyjnych w portach**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 306–315, http://dx.doi.org/10.17531/ein.2020.2.13.

Istnieje wiele portów, w których tory podejściowe lub wewnętrzne kanały nawigacyjne znajdują się obok ścian nabrzeży. W trudnych warunkach hydrometeorologicznych należy utrzymywać odpowiednią prędkość statku, aby zapewnić jego prawidłowe sterowanie podczas przemieszczania się przez kanał. Statki, które przepływają z dużą prędkością w pobliżu ścian nabrzeży, wywierają duże siły na zacumowane jednostki i negatywnie oddziałują na urządzenia cumownicze statków i ściany nabrzeży. Porty powinny zapewniać należyte utrzymanie oraz niezawodność ścian nabrzeży i urządzeń cumowniczych statków cumowanych przy nabrzeżach. W związku z tym ważne jest zbadanie sił interakcji podczas przemieszczania się statku w pobliżu jednostek przycumowanych przy nabrzeżu, aby znaleźć ograniczenia prędkości przepływającego statku w zależności od jego parametrów, odległości i warunków środowiskowych, co pozwoli zapewnić niezawodną eksploatację kanału nawigacyjnego. W artykule zbadane są uwarunkowania sił dynamicznych wywieranych przez przepływające statki, w tym możliwe siły zewnętrzne, które wpływają na zacumowane jednostki, urządzenia cumownicze i ściany nabrzeży. Opracowano metodologię oszacowania sił wywieranych na statek przycumowany przy nabrzeżu przez przepływającą obok jednostkę. Na podstawie wyników analizy studium przypadku zaproponowano zalecenia dotyczące ograniczeń w odniesieniu do statków przepływających w pobliżu statków przycumowanych przy nabrzeżu, które pozwolą zapewnić należytą eksploatację i niezawodność kanału nawigacyjnego i jego infrastruktury.

PUCHALSKI A, KOMORSKA I, ŚLĘZAK M, NIEWCZAS A. **Synthesis of naturalistic vehicle driving cycles using the Markov Chain Monte Carlo method**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 316–322, http://dx.doi.org/10.17531/ein.2020.2.14.

Simulation methods commonly used throughout the design and verification process of various types of motor vehicles require development of naturalistic driving cycles. Optimization of parameters, testing and gradual increase in the degree of autonomy of vehicles is not possible based on standard driving cycles. Ensuring representativeness of synthesized time series based on collected databases requires algorithms using techniques based on stochastic and statistical models. A synthesis technique combining the MCMC method and multifractal analysis has been proposed and

PUCHALSKI A, KOMORSKA I, ŚLĘZAK M, NIEWCZAS A. **Synteza eksploatacyjnych cykli jezdnych samochodów przy wykorzystaniu metody Monte Carlo z zastosowaniem łańcuchów Markowa**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 316–322, http://dx.doi.org/10.17531/ein.2020.2.14.

Metody symulacyjne powszechnie stosowane w całym procesie projektowania i weryfikacji różnych typów pojazdów mechanicznych wymagają opracowania eksploatacyjnych cykli jezdnych. Optymalizacja parametrów, testowanie i stopniowe zwiększanie stopnia autonomiczności pojazdów nie jest możliwe na bazie standardowych cykli jezdnych. Zapewnienie reprezentatywności syntezowanych szeregów czasowych na podstawie zgromadzonych baz danych wymaga algorytmów wykorzystujących techniki bazujące na verified. The method allows simple determination of the speed profile compared to classic frequency analysis.

modelach stochastycznych i statystycznych. Zaproponowano i zweryfikowano technikę syntezy łączącą metodę Monte Carlo wykorzystującą łańcuch Markowa (MCMC) oraz analizę multifraktalną. Metoda umożliwia proste wyznaczenie profilu prędkości jazdy w porównaniu do klasycznej analizy częstotliwościowej.

KOSUCKI A, STAWIŃSKI Ł, MALENTA P, ZACZYŃSKI J, SKOWROŃSKA J. **Energy consumption and energy efficiency improvement of overhead crane's mechanisms**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 323–330, http://dx.doi.org/10.17531/ein.2020.2.15.

The article presents the numerical investigation of the overhead crane's energy consumption. The analysis is based on the hybrid model of the crane consisting of numerical model of drive mechanisms as bridge, trolley, hoist and also experimentally measured power consumption of each control unit. The numerical model was verified experimentally on the real crane. The investigation focuses on analyzing the energy consumption of the overhead crane in relation both to the travelled distance and also for the lifting and lowering heights of a suspended payload. Particular attention was paid on the cases straightly related to the hoist, as a main factor of improvements in the energetic efficiency of the overhead crane. Energy consumption was investigated for a variety of magnitudes of transported mass.

KOSUCKI A, STAWIŃSKI Ł, MALENTA P, ZACZYŃSKI J, SKOWROŃSKA J. **Energochłonność i poprawa efektywności energetycznej mechanizmów suwnicy pomostowej**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 323–330, http://dx.doi.org/10.17531/ein.2020.2.15.

W artykule przedstawiono badania symulacyjne energochłonności pracy suwnicy pomostowej. Podstawą badań jest model hybrydowy bazujący na numerycznych modelach mechanizmów mostu, wózka i wciągarki oraz na eksperymentalnie zmierzonym zapotrzebowaniu mocy dla układu sterowania. Model numeryczny został zweryfikowany na rzeczywistej suwnicy. W pracy przedstawiono analizę energochłonności mechanizmów jazdy w zależności od pokonanej drogi, jak również mechanizmu podnoszenia w zależności od wysokości podnoszenia i opuszczania ładunku. Zwrócono szczególną uwagę na mechanizm wciągarki, jako na główny czynnik poprawy efektywności energetycznej. Energochłonność została zbadana dla różnych mas transportowanego ładunku.

TABASZEWSKI M, SZYMAŃSKI GM. **Engine valve clearance diagnostics based on vibration signals and machine learning methods**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 331–339, http://dx.doi.org/10.17531/ein.2020.2.16.

A dynamic advancement of the design of combustion engines generates a necessity of introduction of strategies of operation based on the information related to their technical condition. The paper analyzes problems related to vibration based diagnostics of valve clearance of a piston combustion engine, significant in terms of its efficiency and durability. Methods of classification have been proposed for the assessment of the valve clearance. Experiments have been performed and described that aimed at providing information necessary to develop and validate the proposed methods. In the performed investigations, the vibration signals were obtained from a triaxial accelerometer located in the engine cylinder head. A parameterization of the obtained vibration signal has been carried out for the engine operating under different engine loads, rotation speeds and valve clearance settings. The parameterization pertained to the specific features of the vibration signals, the derivative of the vibration signal as a function of time as well as the envelope of this derivative. In the first approach, the authors developed a classifier in the form of a set of binary trees that additionally allowed distinguishing the features significant in terms of the identification of adopted classes. For comparison, the authors also developed classifiers in the form of a neural network as well as a k-nearest neighbors algorithm using the Euclidean metric. Based on the performed investigations and analyses a method of valve clearance assessment has been proposed.

TABASZEWSKI M, SZYMAŃSKI GM. **Diagnostyka luzu zaworów silnika spalinowego z wykorzystaniem sygnału drganiowego i metod uczenia maszynowego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 331–339, http://dx.doi.org/10.17531/ein.2020.2.16.

Dynamiczny rozwój konstrukcji silników spalinowych generuje potrzebę wprowadzenia strategii eksploatacji jednostek napędowych, opartej na znajomości ich stanu technicznego. W artykule poddano analizie zagadnienia, związane z drganiową diagnostyką luzu zaworów tłokowego silnika spalinowego, istotnego ze względu na efektywność pracy silnika i jego trwałość. Zaproponowano wykorzystanie metod klasyfikacji do oceny poprawności luzu zaworowego. Przeprowadzono i opisano eksperymenty, które miały na celu dostarczenie informacji koniecznych do zbudowania i zweryfikowania zaproponowanych metod. W przeprowadzonych badaniach pozyskano sygnały drganiowe z trójosiowego czujnika przyspieszeń drgań zlokalizowanego na głowicy silnika. Dokonano parametryzacji uzyskanych przebiegów czasowych sygnału drganiowego dla silnika pracującego pod różnym obciążeniem, z różnymi prędkościami obrotowymi oraz z różnymi luzami zaworowymi. Parametryzacja dotyczyła zarówno cech sygnału przyspieszeń drgań, pochodnej przyspieszeń drgań względem czasu jak i obwiedni tej pochodnej. W pierwszym podejściu zbudowano klasyfikator w postaci zbioru drzew binarnych, który przy okazji pozwolił na wyodrębnienie istotnych, ze względu na przyjęte klasy, cech. Dla porównania zbudowano także klasyfikatory w postaci sieci neuronowej jak i algorytmu k – najbliższych sąsiadów z metryką euklidesową. Na podstawie przeprowadzonych badań i analiz zaproponowano metodę oceny luzu zaworowego.

SUN B, LI Y, WANG Z, LI Z, XIA Q, REN Y, FENG Q, YANG D, QIAN C. **Physics-of-failure and computer-aided simulation fusion approach with a software system for electronics reliability analysis**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 340–351, http://dx.doi.org/10.17531/ein.2020.2.17.

Electronics, such as those used in the communication, aerospace and energy domains, often have high reliability requirements. To reduce the development and testing cost of electronics, reliability analysis needs to be incorporated into the design stage. Compared with traditional approaches, the physics of failure (PoF) methodology can better address cost reduction in the design stage. However, there are many difficulties in practical engineering applications, such as processing large amounts of engineering information simultaneously. Therefore, a flexible approach and a software system for assisting designers in developing a reliability analysis based on the PoF method in electronic product design processing are proposed. This approach integrates the PoF method and computer-aided simulation methods, such as CAD, FEM and CFD.The software system integrates functional modules such as product modeling, load-stress analysis and reliability analysis, which can help designers analyze the reliability of electronic products in actual engineering design. This system includes software and hardware that validate the simulation models. Finally, a case study is proposed in which the software system is used to analyze the filter module reliability of an industrial communication system. The results of the analysis indicate that the system can effectively promote reliability and can ensure the accuracy of analysis with high computing efficiency.

SUN B, LI Y, WANG Z, LI Z, XIA Q, REN Y, FENG Q, YANG D, QIAN C. **Metoda i oprogramowanie do analizy niezawodności urządzeń elektronicznych oparte na połączeniu metodologii fizyki uszkodzeń i symulacji komputerowej**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 340–351, http://dx.doi.org/10.17531/ein.2020.2.17.

Urządzenia elektroniczne, na przykład te używane w łączności, lotnictwie i energetyce, często muszą spełniać wysokie wymagania dotyczące niezawodności. Aby zmniejszyć koszty rozwoju i testowania tego typu urządzeń, należy opracować metodę analizy niezawodności, którą można wykorzystywać już na etapie projektowania. Metodologia fizyki uszkodzeń (PoF) pozwala, lepiej niż tradycyjne podejścia, rozwiązywać problemy związane z niezawodnością już na etapie powstawania projektu. Jednak jej zastosowanie w praktyce inżynierskiej nastręcza wielu trudności, związanych, między innymi, z koniecznością jednoczesnego przetwarzania dużych ilości informacji inżynieryjnych. W związku z tym, w przedstawionej pracy zaproponowano elastyczne podejście oraz system oprogramowania, które mogą być wykorzystywane przez projektantów do opracowania analizy niezawodności produktu elektronicznego w opaciu o PoF na etapie projektowania. Podejście to stanowi połączenie metody PoF i metod symulacji komputerowej, takich jak CAD, FEM i CFD. System oprogramowania zawiera moduły funkcjonalne, takie jak modelowanie produktu, analiza obciążeń, analiza niezawodności i inne, które mogą wspomagać projektantów w analizie niezawodności projektowanych przez nich produktów elektronicznych. Na system ten, oprócz oprogramowania składa się także sprzęt komputerowy, który służy do walidacji modeli symulacyjnych. W artykule przedstawiono studium przypadku, w którym zaproponowany system oprogramowania wykorzystano do analizy niezawodności modułu filtra wykorzystywanego w systemie łączności przemysłowej. Wyniki analizy pokazują, że opracowane oprogramowanie skutecznie poprawia niezawodność urządzeń jak też zapewnia dokładność analizy przy jednoczesnej wysokiej wydajności obliczeniowej.

ROMANIUK M, HRYNIEWICZ O. **Estimation of maintenance costs of a pipeline for a U-shaped hazard rate function in the imprecise setting**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 352–362, http://dx.doi.org/10.17531/ein.2020.2.18.

In this paper, we discuss imprecise settings for an evaluation of the maintenance costs of a water distribution system (WDS). Moments of failures of pipes are modelled using a newly proposed three-piece convex hazard rate function (HRF) for which number of previous failures is taken into account, too. Both fuzzy sets and shadowed sets are used to model the impreciseness of important parameters of this HRF and the costs of maintenance services. Contrary to more classical and widely-used approaches to cost analysis (i.e. a constant yield or nominal value of money), a strictly stochastic process (i.e. the one-factor Vasicek model) of an interest rate is assumed in the analysis of maintenance costs. This approach models future behaviour of the interest rate (i.e. the future value of money) in a more realistic way. Respective algorithms together with exemplary results of numerical simulations for two setups, which are related to fuzzy and shadowed sets, are also provided.

KORTA J, SKRUCH P. **The concept of integrating vapor chamber into a housing of electronic devices for increased thermal reliability**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 363–369, http://dx.doi.org/10.17531/ein.2020.2.19.

Systematic increase in computational power and continuous miniaturization of automotive electronic controllers pose a challenge to maintaining allowable temperature of semiconductor components, preventing premature wear-out or, in extreme cases, unacceptable shutdown of these devices. For these reasons, efficient and durable cooling systems are gaining importance in modern car technology design, showing critical influence on reliability of vehicle electronics. Vapor chambers (flat heat pipes) which could support heat management of automotive electronic controllers in the nearest future are passive devices, which transport heat through evaporation-condensation process of a working liquid. At present, vapor chambers are not commercially used in cooling systems of automotive controllers, being a subject of research and development endeavors aimed at understanding their influence on thermomechanical reliability of semiconductor devices used in cars. This paper presents a concept of an electronic controller aluminum housing integrated with a vapor chamber. The conceptual design was numerically validated in elevated temperature, typical for automotive ambient conditions. The paper discusses influence of the vapor chamber-based cooling system on the controller's thermal performance, as well as on its reliability, expressed as the expected lifetime of the device.

POURHASSAN MR, RAISSI S, HAFEZALKOTOB A. **A simulation approach on reliability assessment of complex system subject to stochastic degradation and random shock**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 370–379, http://dx.doi.org/10.17531/ein.2020.2.20.

Many systems are affected by different random factors and stochastic processes, significantly complicating their reliability analysis. In general, the performance of complicated systems may gradually, suddenly, or continuously be downgraded over times from perfect functioning to breakdown states or may be affected by unexpected shocks. In the literature, analytic reliability assessment examined for especial cases is restricted to applying the Exponential, Gamma, compound Poisson, and Wiener degradation processes. Consideration of the effect of non-fatal soft shock makes such assessment more challenging which has remained a research gap for general degraded stochastic processes. Through the current article, for preventing complexity of analytic calculations, we have focused on applying a simulating approach for generalization. The proposed model embeds both the stochastic degradation process as well randomly occurred shocks for two states, multi-state, and continuous degradation. Here, the user can arbitrarily set the time to failure distribution, stochastic degradation, and time to occurrence shock density function as well its severity. In order to present the validity and applicability, two case studies in a sugar plant alongside an example derived from the literature are examined. In the first case study, the simulation overestimated the system reliability by less than 5%. Also, the comparison revealed no significant difference between the analytic and the simulation result in an example taken from an article. Finally, the reliability of a complicated crystallizer system embedding both degradation and soft shock occurrence was examined in a three-component standby system.

ROMANIUK M, HRYNIEWICZ O. **Estymacja kosztów eksploatacyjnych rurociągu dla U-kształtnej funkcji intensywności uszkodzeń przy nieprecyzyjnym podejściu**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 352–362, http://dx.doi.org/10.17531/ein.2020.2.18.

W niniejszym artykule omawiamy nieprecyzyjne podejścia do problemu obliczenia kosztów eksploatacji systemu dystrybucji wody (WDS). Czasy uszkodzeń rur modelowane są z wykorzystaniem nowo zaproponowanej trzyczęściowej wypukłej funkcji intensywności uszkodzeń (hazard rate function, HRF) dla której brana jest pod uwagę również liczba wcześniejszych uszkodzeń. Do modelowania nieprecyzyjności istotnych parametrów tej HRF oraz kosztów działań serwisowych są wykorzystywane zarówno zbiory rozmyte jak i zbiory cieniowane. W przeciwieństwie do bardziej klasycznych i szeroko wykorzystywanych podejść do analizy kosztów eksploatacji (tzn. stałej stopy procentowej lub wartości nominalnej pieniądza), założono ściśle stochastyczny proces (tzn. jednoczynnikowy model Vasicka) dla stopy procentowej. Podejście to modeluje przyszłe zachowanie stopy procentowej (czyli przyszłej wartości pieniądza) w bardziej realistyczny sposób. Zaprezentowano również odpowiednie algorytmy wraz z przykładowymi wynikami symulacji numerycznych dla dwóch zestawów parametrów, związanych ze zbiorami rozmytymi i cieniowanymi.

KORTA J, SKRUCH P. **Koncepcja integracji komory parowej z obudową układów elektronicznych w celu zwiększenia ich niezawodności**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 363–369, http://dx.doi.org/10.17531/ein.2020.2.19.

Systematycznie wzrastająca moc obliczeniowa oraz postępująca miniaturyzacja urządzeń elektronicznych stosowanych w pojazdach samochodowych powodują trudności w utrzymaniu temperatury pracy elementów półprzewodnikowych w dozwolonym zakresie, przyczyniając się do ich przedwczesnego zużycia, a w skrajnych przypadkach, uniemożliwiając nawet ich normalną pracę. Wydajne i trwałe układy chłodzące stają się więc nieodzownym komponentem współczesnych podzespołów samochodowych, o krytycznym znaczeniu dla ich niezawodności. Urządzeniami mogącymi w niedalekiej przyszłości wspomagać działanie układów chłodzenia systemów elektronicznych wykorzystywanych w motoryzacji są komory parowe (płaskie rurki cieplne), w których transport energii termicznej zachodzi poprzez przemianę fazową i samoistne przemieszczanie się czynnika roboczego. Współcześnie, tego rodzaju urządzenia nie są komercyjnie stosowane w układach chłodzenia sterowników samochodowych, pozostając przedmiotem prac badawczo-rozwojowych związanych z ich wpływem na szeroko pojętą niezawodność termomechaniczną urządzeń elektronicznych. W niniejszym artykule opisano koncepcję zintegrowania komory parowej z aluminiową obudową kontrolera elektronicznego pracującego w warunkach podwyższonej temperatury otoczenia, odpowiadającej warunkom użytkowania komponentów samochodowych. Ponadto, ocenie poddano wpływ zastosowania tego urządzenia na temperaturę pracy chłodzonego elementu półprzewodnikowego i jego niezawodność, wyrażoną jako przewidywany czas jego bezawaryjnego funkcjonowania.

POURHASSAN MR, RAISSI S, HAFEZALKOTOB A. **Metoda symulacyjna oceny niezawodności złożonego systemu podlegającego procesom degradacji stochastycznej i narażonego na obciążenia losowe**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 370–379, http://dx.doi.org/10.17531/ein.2020.2.20.

Prawidłowe działanie wielu systemów zależy od różnych czynników losowych i procesów stochastycznych, co znacznie komplikuje analizę niezawodności tych układów. Parametry pracy skomplikowanych systemów mogą ulegać stopniowemu, nagłemu lub stałemu obniżeniu ze stanu doskonałego funkcjonowania do stanu awaryjnego. Wpływ na nie mogą też mieć niespodziewane obciążenia. W literaturze przedmiotu, analityczną ocenę niezawodności stosuje się do badania przypadków szczególnych i ogranicza do badania degradacji w oparciu o proces wykładniczy, proces gamma, złożony proces Poissona i proces Wienera. Ocena niezawodności z uwzględnieniem wpływu obciążeń miękkich, nieprowadzących do całkowitej awarii, stanowi większe wyzwanie tworząc lukę w badaniach nad ogólnymi stochastycznymi procesami degradacji. Aby uniknąć złożonych obliczeń analitycznych, w niniejszej pracy skupiliśmy się na zastosowaniu podejścia symulacyjnego w celu uzyskania generalizacji. Proponowany model obejmuje zarówno stochastyczny proces degradacji, jak i losowo występujące obciążenia i uwzględnia przypadki degradacji systemów dwustanowych, wielostanowych oraz degradacji ciągłej. Posługując się tym modelem, użytkownik może dowolnie ustawiać rozkład czasu do uszkodzenia, degradację stochastyczną, czas do wystąpienia obciążenia, funkcję gęstości prawdopodobieństwa wystąpienia obciążenia, a także jego nasilenie. Trafność oraz możliwości zastosowania przedstawionego modelu zilustrowano na podstawie dwóch studiów przypadków dotyczących cukrowni oraz przykładu zaczerpniętego z literatury. W pierwszym studium przypadku, poziom niezawodności systemu obliczony na podstawie symulacji różnił się o mniej niż 5% od wyniku otrzymanego na drodze analitycznej. Porównanie nie ujawniło również żadnej istotnej różnicy między wynikiem analitycznym a symulacyjnym w przykładzie pochodzącym z literatury. Artykuł wieńczy analiza niezawodności złożonego układu krystalizatora, obejmująca zarówno degradację, jak i występowanie miękkich obciążeń w trójelementowym systemie krystalizatora z rezerwą.

SCIENCE AND TECHNOLOGY

Article citation info:
VALIS D, FORBELSKÁ M, VINTR Z. Forecasting study of mains reliability based on sparse field data and perspective state space models. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 179–191, http://dx.doi.org/10.17531/ein.2020.2.1.

David VALIS
Marie FORBELSKÁ
Zdeněk VINTR

FORECASTING STUDY OF MAINS RELIABILITY BASED ON SPARSE FIELD DATA AND PERSPECTIVE STATE SPACE MODELS

PROGNOZOWANIE NIEZAWODNOŚCI ELEMENTÓW SIECI WODOCIĄGOWEJ NA PODSTAWIE RZADKICH DANYCH TERENOWYCH I MODELI PRZESTRZENI STANÓW

The elements of critical infrastructure have to meet demanding dependability, safety and security requirements. The article deals with the prognosis of water mains reliability while using sparse irregular filed data. The data are sparse because the only thing we know is the number of mains failures during a given month. Since it is possible to transform the data into a typical reliability measure (rate of failure occurrence – ROCOF), we can examine the course of this measure development in time. In order to model and predict the ROCOF development, we suggest novel single and multiple error state space models. The results can be used for i) optimizing mains operation and maintenance, ii) estimating life cycle cost, and iii) planning crisis management.

Keywords: *mains, critical infrastructure, reliability prognosis, sparse data, state space models.*

Elementy infrastruktury krytycznej muszą spełniać wysokie wymagania w zakresie niezawodności, bezpieczeństwa i ochrony. Artykuł dotyczy prognozowania niezawodności sieci wodociągowej przy wykorzystaniu nieregularnie rejestrowanych rzadkich danych. Wykorzystane w pracy dane są rzadkie, ponieważ dostarczają jedynie informacji na temat liczby uszkodzeń wodociągu w danym miesiącu. Przekształcenie tych danych w typową miarę niezawodności (wskaźnik występowania uszkodzeń – ROCOF), pozwala zbadać przebieg rozwoju tej miary w czasie. Rozwój ROCOF można modelować i przewidywać za pomocą zaproponowanych w pracy innowacyjnych modeli przestrzeni stanów uwzględniających pojedynczy błąd lub wiele błędów. Uzyskane wyniki można wykorzystać do i) optymalizacji pracy i eksploatacji sieci wodociągowej, ii) szacowania kosztów cyklu życia, oraz iii) planowania zarządzania kryzysowego.

Słowa kluczowe: *sieć wodociągowa, infrastruktura krytyczna, prognoza niezawodności, rzadkie dane, modele przestrzeni stanów.*

1. Introduction

Water mains are an important part of the country critical infrastructure. The level of their dependability, safety and security is very much required and it must be high. Because of different structural designs, materials used, running failures and repairs, the technical condition of mains varies. Moreover, the system is subject to extreme stress owing to weather conditions. Monitoring the condition of this device is therefore very difficult. Both direct and indirect diagnostics is very problematic and does not provide satisfactory results. The access to the field data of this device is also difficult, and is not that common. Despite all these limitations mentioned, we are able to examine and predict the water mains reliability. For this purpose we use the sparse data set, and in order to model the data, we apply novel single and multiple error state space models.

1.1. State of the art

This part is devoted to the literature research focusing mainly on technical applications containing i) irregular sparse data, ii) single and multiple error state space models, iii) ways of predicting reliability.

Current publications may contain the results which deal with irregular sparse data. As examples, we introduce a few selected works which we found interesting, for different reasons explained later. Antholzer et al [2] propose a solution for accurate image reconstruction in the form of algorithms. They investigate this issue for the sparse data problem in photoacoustic tomography (PAT). They develop a direct and highly efficient reconstruction algorithm based on deep learning. In their approach, image reconstruction is performed with a deep convolutional neural network (CNN), whose weights are adjusted prior to the actual image reconstruction based on a set of training data. The proposed reconstruction approach can be interpreted as a network that uses the PAT filtered back projection algorithm for the first layer, followed by the U-net architecture for the remaining layers. This work is inspirational but data set analyzed is not very large. Guo et al [17] bear in mind that fault diagnosis bearings in wind turbine and the drivetrain is very important to reduce the maintenance cost of the wind turbine and improve economic efficiency. However, the traditional diagnosis methods have difficulty in extracting the impulsive components from the vibration signal of the wind turbine because of heavy background noise and harmonic interference. In their paper, they propose a novel method based on data-driven multiscale dictionary construction. Firstly, they achieve the useful atom through training the K-means singular value decomposition (K-SVD) model with a standard signal. Secondly, they deform the chosen atom into different shapes and construct the final dictionary. Thirdly, the constructed dictionary is used to sparsely represent the vibration signal, and orthogonal matching pursuit (OMP) is performed to extract the impulsive component. The proposed method is robust to harmonic interference and heavy background noise. Moreover, the effectiveness of the proposed method is validated by numerical simulation and two experimental cases including the bearing fault of the wind turbine generator in the field test. This approach is interesting but the analyzed diagnostic signal is vibration, which includes some noise and is deformed by imperfection in terms of signal transfer. Wang and Lu [39] understand that incomplete modal data is mostly measured for experimental and engineering structures. However, its application into structural damage identification suffers from the drawback that the amount of the incomplete modal data is often insufficient, rendering the identification very sensitive to the measurement noise. Aiming to overcome this drawback, this paper proposes a new damage identification approach that combines the incomplete modal data with the sparse regularization. The realization of the proposed damage identification approach is mainly threefold: (a) The first is the establishment of a new goal function which is decoupled with respect to the damage parameters. To this end, the decomposition of the stiffness matrix so that the damage parameters are contained in a diagonal matrix must be introduced. (b) The second is the application of the alternating minimization approach to get the solution of the new goal function. (c) The third is the development of a novel and simple threshold setting method to properly determine the sparse regularization parameter. The feature of the proposed damage identification approach lies in that the sensitivity analysis is not involved and the exact orders of the modal data are not demanded. Numerical and experimental examples are conducted to verify the proposed damage identification approach. This outcome is also interesting, however we miss practical implementation of the proposed method. Experimental data are even more sparse. Sousa and Wang [35] present an application on bridges monitoring system. It uses on-structure sensors that are able to acquire signals sensitive to traffic load events, which can be used as an indirect indicator of the load magnitude. In this paper, sparse representation algorithms have been innovatively applied to the bridge weight in motion monitoring system data compression. A comparative study is performed based on measurements collected from a real bridge, by exploring different methods including Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), and two dictionary learning methods, i.e. Compressive Sensing (CS) and K-means Singular Value Decomposition (K-SVD). This outcome is interesting but again, vibration assessment causes some noise in the final system state determination. Inspirational examples on specific diagnostic data forms analysis can be found in [9, 13, 14, 25, 28, 32].

In the next part of our literature review we focus on applying state-space time series models related to technical systems and reliability. The state-space models have been used by numerous applications based on the data analysis. In the text below there are examples of the outputs we found particularly inspiring. Dos Santos et al [10] propose new reliability models whose likelihood consists of decomposition of data information in stages or times, thus leading to latent state parameters. Alternative versions of some well-known models such as piecewise exponential, proportional hazards, and software reliability models are shown to be included in our unifying framework. In general, latent parameters of many reliability models are high dimensional, and their inference requires approximating methods such as Markov chain Monte Carlo (MCMC) or Laplace. Latent states in their models are related across stages through a non-Gaussian state-space framework. This feature makes the models mathematically tractable and allows for the exact computation of the marginal likelihood function, despite the non-Gaussianity of the state. The proposed non-Gaussian evolution models circumvent the need for approximations, which are required in similar likelihood-based approaches. This model however, has limited practical applicability. Li et al [29] present method of reliability prediction based on state space model. Firstly, signals about machine working conditions are collected based on-line monitoring technology. Secondly, wavelet packet energy parameters are determined based on the monitored signals. Frequency band energy is regarded as characteristic parameter. Then, the degradation characteristics of signal to noise ratio is improved by moving average filtering processing. In the end, SSM is established to predict degradation characteristics of probability density distribution, and the degree of reliability is determined. Milling cutter is used to demonstrate the rationality and effectiveness of this method. Simpler version of state-space model is applied here. Distefano et al [8] present outcomes of their article where the main goal is to demonstrate how state-space based techniques can satisfy demands on reliability and availability assessment. For this purpose some examples of specific dynamic reliability behaviors, such as common cause failure and load sharing, are considered applying state-space based techniques to study the corresponding reliability models. Different repair policies in availability contexts are also explored. Both Markovian and non-Markovian models are studied via phase type expansion and renewal theory in order to adequately represent and evaluate the considered dynamic reliability aspects in case of generally distributed lifetimes and times to repair. Although there are a great number of state-space models applications used not only in a technical area, it is difficult to find the specific application working with a single or multiple error state-space model. The authors offering interesting results in [11, 19].

Last but not least, we focus on the results dealing with reliability forecasting/prognosis. Gobbato et al [15] present approach based on fatigue assessment of structural components. The first part of the paper provides an overview and extension of a comprehensive reliability-based fatigue damage prognosis methodology – previously developed by the authors – for recursively predicting and updating the remaining fatigue life of critical structural components and/or sub-components in aerospace structures. In the second part of the paper, a set of experimental fatigue test data, available in the literature, is used to provide a numerical verification and an experimental validation of the proposed framework at the reliability component level (i.e., single damage mechanism evolving at a single damage location). Gobbato et al [16] present the theoretical basis of a novel and comprehensive probabilistic methodology for predicting the remaining service life of adhesively bonded joints within the structural components of composite aircraft, with emphasis on a composite wing structure. Non-destructive evaluation techniques and recursive Bayesian inference are used to (i) assess the current state of damage of the system and (ii) update the joint probability distribution function (PDF) of the damage extents at various locations. A probabilistic model for future aerodynamic loads and a damage evolution model for the adhesive are then used to stochastically propagate damage through the joints and predict the joint PDF of the damage extents at future times. This information is subsequently used to probabilistically assess the reduced (due to damage) global aeroelastic performance of the wing by computing the PDFs of its flutter velocity and the velocities associated with the limit cycle oscillations of interest. Both of these methods are interest-

Table 1. Extracted sample of field data example – number of failures F_t during a respective month and ROCOF

Year	2000		2001		2002		2003	
Month	F_t	ROCOF	F_t	ROCOF	F_t	ROCOF	F_t	ROCOF
January	6	0.19354839	10	0.32258065	6	0.19354839	13	0.41935484
February	12	0.4137931	11	0.39285714	11	0.39285714	7	0.25
March	27	0.87096774	14	0.4516129	14	0.4516129	8	0.25806452
April	23	0.76666667	10	0.33333333	4	0.13333333	13	0.43333333
May	18	0.58064516	12	0.38709677	4	0.12903226	4	0.12903226
June	8	0.26666667	9	0.3	3	0.1	2	0.06666667
July	12	0.38709677	16	0.51612903	7	0.22580645	4	0.12903226
August	8	0.25806452	12	0.38709677	10	0.32258065	8	0.25806452
September	12	0.4	10	0.33333333	7	0.23333333	11	0.36666667
October	17	0.5483871	10	0.32258065	6	0.19354839	12	0.38709677
November	24	0.8	5	0.16666667	6	0.2	7	0.23333333
December	17	0.5483871	7	0.22580645	14	0.4516129	12	0.38709677
Total in year	184		126		92		101	

ing however small testing data set for approach validity confirmation has been presented. Dindarloo [7] presents alternatives to traditional reliability assessment approaches. Both the autoregressive integrated moving average (ARIMA or the Box-Jenkins technique) and artificial neural networks (ANNs) are viable alternatives to the traditional reliability analysis methods (e.g., Weibull analysis, Poisson processes, non-homogeneous Poisson processes, and Markov methods). Time series analysis of the times between failures via ARIMA or ANNs does not have the limitations of the traditional methods such as requirements/assumptions of a priori postulation and/or statistically independent and identically distributed observations for TBFs. The reliability of an load-haul-dump unit was investigated by analysis of time between failures. Seasonal autoregressive integrated moving average (SARIMA) was employed for both modeling and forecasting the failures. The results were compared with a genetic algorithm-based (ANNs) model. In this approach we can see interesting practical application. Kontrec et al [26] propose an approach that supports decision making process in planning and controlling of spare parts in aircraft maintenance systems. Reliability characteristics of aircraft consumable parts were analyzed in order to substantiate this approach. Moreover, the proposed reliability model was used to evaluate characteristics of subassemblies and/or assemblies these parts belong to. Finally, an innovative approach for determining the total amount of parts required in inventory and underage costs, based on observing the total unit time as a stochastic process, is presented herein. This application of stochastic process is also partially inspirational. Other encouraging results which deal with assessing systems reliability and include real field data but not only on water mains, are those introduced in [27,30,36-38].

1.2. Motivation

The main motivation of this article is to show that despite having only restricted information in the data, it is still possible to carry out certain estimates and prognoses of reliability measures. What is more, the data are very austere, and from the mathematics point of view, also very sparse. The only thing available is the number of failures which occurred during single months, the data information value is therefore very low. The assumptions about the future system behaviour are then very uncertain because of numerous influences, e.g. seasonal influences, water mains material, etc. That is the reason why we would like to introduce a suitable mathematical approach based on a few newly proposed single and multiple error state-space models. Using these models, we have the ambition not only to estimate the course of ROCOF, but also predict the system behaviour in the future. This information can significantly contribute to i) the optimization of this part of a critical infrastructure, ii) the planning of water mains maintenance, iii) the support of crisis management and emergency planning, iv) the rationalization of life cycle costs (LCC).

2. Field data analyzed

The analysed data are dichotomous and quantitative. They are the real field data collected in a broad area from the operators of the mains distribution system. In the area covered by the monitoring there are more than 5 million inhabitants supplied by water. The data cover the period longer than 17 years. Table 1 shows the example of such a data segment. We introduce the example of a complete form of the recorded data when having only the number of failures during single months. Although it would be possible to work with quarters which could filter out some data mistakes, we prefer to work with the number of failures during a month. The work with this variable would be possible but not entirely favourable. This value is therefore always transferred into the rate of failures occurrence (ROCOF), since it is necessary to consider and filter out the different number of days during a month. However, we presume that it will be possible to trace certain seasonal influences affecting the data.

The total course of ROCOF during a observed time elapse is shown in Fig. 1. The course is accompanied by trend smoothing non-parametric curves: lowess and cubic spline.

Next, in Fig. 2. we applied box plots to show a preliminary analysis of a time-series based on robust estimates of yearly and monthly ROCOF levels (thick lines in the boxes illustrate a median value) and on robust estimates of ROCOF variability during single years or months (see the box containing 75 % of the values of a given period). The left graph (panel) shows that the failure frequency decreases over

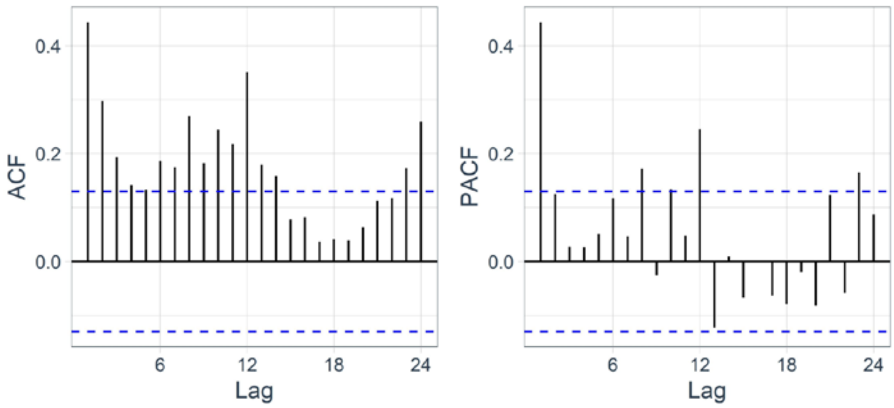


Fig. 1. *ROCOF – monthly of observed series: accompanied with lowess trend line and its 95% confidence intervals (left) and with cubic spline and its 95% confidence intervals (right)*

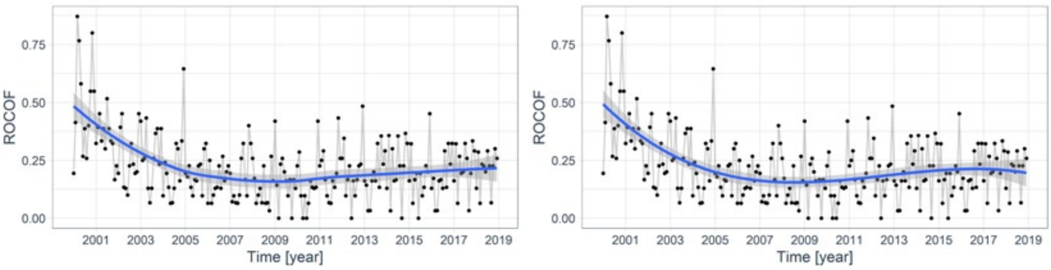


Fig. 2. *Visualisation of robust estimates of ROCOF levels and of ROCOF variabilities for each year (left panel) and months (right panel)*

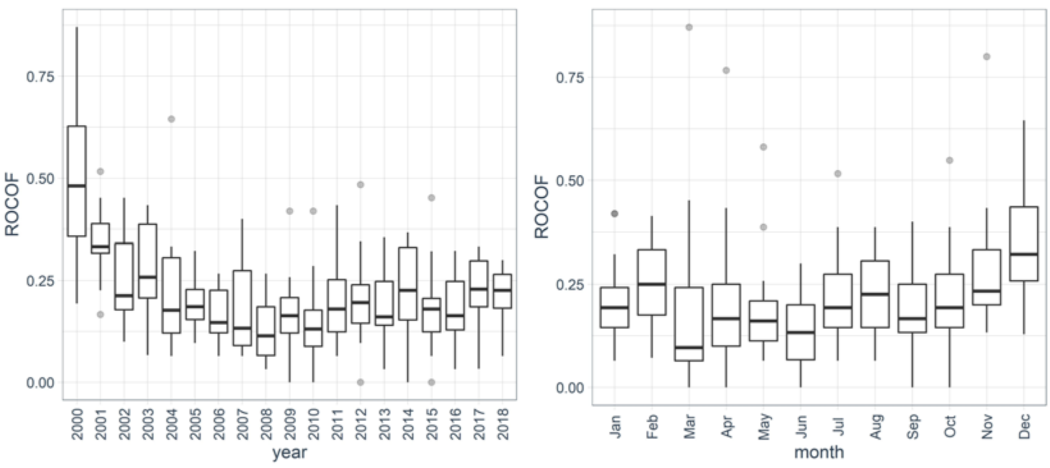


Fig. 3. *Course of autocorrelation function during respective quarters (left) and partial autocorrelation function (right)*

time – during years, which indicates that the system condition 'gets improved', perhaps due to gradual modernisation of network lines. We can also see that the ROCOF variability fluctuates during single years. The right graph (panel) shows the distribution of ROCOF in respective months and its variability. That is very useful since we work with and study the seasonality.

Since we have the idea of applying the following mathematical tools, we would like to verify the dependence or independence in the data – an autocorrelation function (ACF) and a partial correlation (PACF). Autocorrelation and partial autocorrelation plots are very often used in time series analysis and forecasting. These are plots that graphically summarize the strength of a relationship with an observation in a time series with observations at prior time steps, called lags. The blue dashed lines represent an approximate confidence interval for what is produced by white noise, by default a 95% interval. The ACF identifies the obvious seasonal variation (with high positive autocorrelations at lags 12, 24, …) and shows the slow decay typical for a non-stationary series. The results shown in Fig. 3 demonstrate that the data are independent.

No standard state-space models fit the data distributed in this manner. Therefore, we suggest special forms which capture, describe and fit the behaviour in the data a lot better. They are also able to depict the course of ROCOF in a better and more accurate way.

3. Modelling methodology

When modelling the system behaviour from the reliability point of view, we will follow the approach illustrated in Fig. 4. The first four steps have already been described above, and the remaining steps will be described in view of the applied theory.

Using the recorded data and the calculated ROCOF, we try to find the proper state-space model.

3.1. General theory of state space models

At first, we will describe briefly and generally the basic state-space models theory, later we will give a more detailed description of our newly proposed approaches and modifications of specially designed state-space models to be used for the examined specific technical case.

The state space representation of a linear time series model is given by:

$$y_t = w_t' x_{t-1} + \varepsilon_t \tag{1}$$

$$x_t = F_t x_{t-1} + \eta_t \tag{2}$$

for $t = 1, \dots, n$, where y_t is the observed time series value. w_t (with w_t' its transpose) is assumed to be known vector, F_t a known matrix and x_t is the (possibly unobservable) state vector. Furthermore, ε_t are serially uncorrelated disturbances assumed to have zero mean and variance σ_ε^2, while η_t is a vector of serially uncorrelated disturbances with zero mean and covariance matrix V_t. Usually, ε_t and η_t are assumed to be (multivariate) normally distributed and uncorrelated with each other at all time periods (i.e. $E(\varepsilon_t\ \eta_s) = 0$ for all $t, s = 1, \dots, n$). The model is thus characterized by w_t, F_t, σ_ε^2 and V_t. Equation (1) is known as the observation or measurement equation and (2) as the transition or state equation. If w_t, F_t and V_t do not change over time, the model is said to be time-invariant. Next we assume, the underlying state space models are time-invariant.

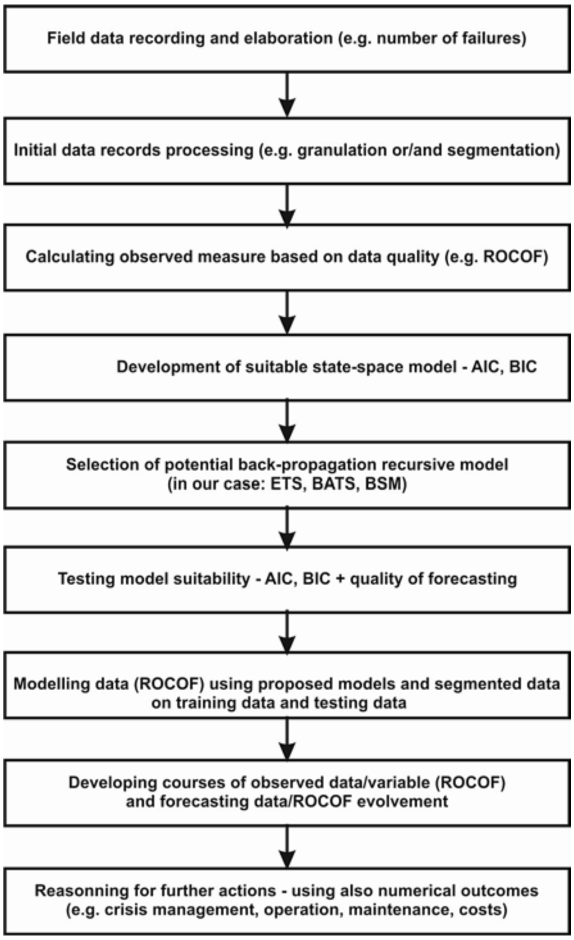


Fig. 4. Methodology of elaboration the recorded data on mains failure

One popular special case of state space models is the class of structural models. For these models, the vector x_t denotes a vector of states corresponding to the trend, cycle and seasonal components (see [18], chapter 2).

A special form of the state space model is the innovations model by Anderson and Moore (1979) [1]:

$$y_t = w_t' x_{t-1} + \varepsilon_t \tag{3}$$

$$x_t = F_t x_{t-1} + g\eta_t \tag{4}$$

where g is a fixed vector. The difference with the general state space model (1)-(2) is that the disturbance terms in the observation and transition equations are now perfectly correlated, implying that there is effectively only a single source of randomness. Since there is only one disturbance term in the model, it is referred to as a single source of error (SSOE) model. Models with more than one disturbance term are known as multiple sources of error (MSOE) models. This model has many common models as special cases, such as multiple regression, exponential smoothing, and ARIMA models. Any ARIMA model can be converted to this form. This model was first related to exponential smoothing in [33].

Since state space models contain both an unobservable state vector and unknown parameters, estimation for state space models has two aspects: i) estimating the unknown parameters, ii) estimating the unobservable state variables.

The first can be done by, for example, maximum likelihood estimation. The second is done by the Kalman filter (see e.g., [18]). The Kalman filter is a recursive procedure for computing the optimal estimator of the state vector at time t, based on the information available at time t (i.e. the observations up to and including y_t). Every time period, when new observations become available, this information is used to update the estimates. For state space models, optimal predictions of future observations and optimal estimates of unobserved components can be made using the Kalman filter. If the model is Gaussian (ε_t,η_t and the initial state are normally distributed), the Kalman filter is optimal in the sense that it yields minimum mean square error estimators (MMSE's).

For both the time-invariant MSOE and SSOE model, the Kalman filter converges to a steady state under some conditions.

In the next part we give a more detailed description of the particular proposed models used for examining our specific technical case.

3.1.1. Single error state space models for exponential smoothing

Exponential smoothing state space methods constitute a broad family of approaches to univariate time series forecasting that have been around for many decades and only in the twenty-first century placed into a systematic framework. The definitive book on the subject is [20]. In general, innovation ETS models are defined according to three model structure parameters: (E) error type, (T) trend type, and (S) seasonality type [20,24]. Each of the parameters can be an N (none), A (additive), or M (multiplicative) state. The trend component also can be d (damped) type.

ETS(A,Ad,A) (seasonal exponential smoothing with damped trends)

Consider the single error state space model given by:

observed series: $y_t = l_t + s_{t-m} + \varepsilon_t \quad \varepsilon_t \sim {}^\circ NID\left(0,\sigma_\varepsilon^2\right)$

latent level: $l_t = l_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t \quad 0 < \phi < 1$ – damping parameter

latent trend/drift: $b_t = \phi b_{t-1} + \beta\varepsilon_t$

latent seasonal: $s_t = s_{t-m} + \gamma\varepsilon_t \quad 0 \le \alpha, \beta, \gamma \le 1$ - smoothing parameters

A deterministic representation of the seasonal components can be obtained by setting the smoothing parameters equal to zero.

BATS model (exponential smoothing state space model with Box-Cox transformation, ARMA errors, trend and seasonal components)

De Livera et al. [6] propose modifications to the linear innovation models in order to include a wide variety of seasonal patterns and solve the problem of correlated errors. To avoid falling into nonlinearity problems, these authors restricted the models to those homoskedastic and the Box-Cox transformation [3] are used when there is some type of specific nonlinearity. The model including the transformation of Box and Cox, ARMA errors (see [4,5,40]) and seasonal patterns can be expressed as follows:

Box-Cox power transformation: $y_t^{(\omega)} = \begin{cases} \dfrac{y_t^{(\omega)} - 1}{\omega} & \omega \ne 0 \\ \ln y_t & \omega = 0 \end{cases}$

transformed observation: $y_t^{(\omega)} = l_t + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-m}^{(i)} + d_t$

latent level: $l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$

latent trend/drift: $b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t$

latent seasonal: $s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t$

$ARMA(p, q)$ process: $d_t = \sum_{i=1}^{p}\varphi_i d_{t-1} + \sum_{i=1}^{q}\theta_i \varepsilon_{t-1} + \varepsilon_t \ \varepsilon_t \sim NID\left(0,\sigma_\varepsilon^2\right)$

where:

ω	Box-Cox transformation parameter
ϕ	damping parameter of trend (see [12, 34]),
$m_1,\cdots,m_M$	seasonal periods,
b	long-run trend parameter,
b_t	short-run trend parameter in time t,
$\alpha,\beta,\gamma_1^{(i)},\gamma_2^{(i)}$	smoothing parameters.

These models are called BATS with arguments (ω,ϕ, p, q, m_1, …, m_M). A deterministic representation of the seasonal components can be obtained by setting the smoothing parameters equal to zero.

3.1.2. Multiple error state space model

Among state space models with multiple errors, the following model was chosen as suitable

observed series: $y_t = l_t + s_{1,t} + \varepsilon_t \qquad \varepsilon_t \sim NID\left(0,\sigma_\varepsilon^2\right)$

latent level: $l_{t-1} + b_{t-1} + \eta_t \qquad \eta_t \sim NID\left(0,\sigma_\eta^2\right)$

latent trend/drift: $b_t = b_{t-1} + \xi_t \qquad \xi_t \sim NID\left(0,\sigma_\xi^2\right)$

latent seasonal: $s_{1,t} = -\sum_{j=1}^{s-1} s_{j,t-1} + \nu_t \qquad \nu_t \sim NID\left(0,\sigma_\nu^2\right)$

$$s_{2,t} = s_{1,t-1}$$
$$\vdots$$
$$s_{m-1,t} = s_{m-2,t-1}$$

This model is called BSM (Basic Structural Model). If we omit the drift component b_t in the BSM model, we mark this model as Level-BSM.

3.1.3. Quality of forecasting

Different criteria such as forecast error measurements, the speed of calculation, interpretability and others have been used to assess the quality of forecasting. Forecast error measures or forecast accuracy are the most important in solving practical problems. Typically, the common used forecast error measurements are applied for estimating the quality of forecasting methods and for choosing the best forecasting mechanism in case of multiple objects.

Training and test sets

It is important to evaluate forecast accuracy using genuine forecasts. Consequently, the size of the residuals is not a reliable indication of how large true forecast errors are likely to be. The accuracy of forecasts can only be determined by considering how well a model performs on new data that were not used when fitting the model. When choosing models, it is common practice to separate the available data $Y_n = \{y_1, \dots, y_n\}$ into two portions, training and test data,

training data: $Y_T = \{y_1, \dots, y_T\}$, test data: $Y_{test} = \{y_{T+1}, y_{T+2}, \dots, y_n\}$

where the training data is used to estimate any parameters of a forecasting method and the test data is used to evaluate its accuracy. Because the test data is not used in determining the forecasts, it should provide a reliable indication of how well the model is likely to forecast on new data. This data sub-division into regions can be seen in Fig. 5.

Forecast Accuracy

The forecast accuracy can be evaluated on the test set using residual diagnostics and forecast accuracy measures. A forecast "error" is the difference between an observed value and its forecast. Here "error" does not mean a mistake, it means the unpredictable part of an observation. It can be written as:

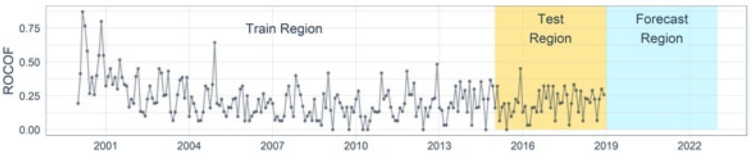


Fig. 5. Example of data sub-division into training, testing and forecasting region

$$e_t = y_t - f_t^{(m)}$$

where y_t is the measured value at time t, $f_t^{(m)}$ is predicted value at time t, obtained from the use of the forecast model m. Note that forecast errors are different from residuals in two ways. First, residuals are calculated on the training set while forecast errors are calculated on the test set. Second, residuals are based on one-step forecasts while forecast errors can involve multi-step forecasts. We can measure forecast accuracy by summarising the forecast errors in different ways – see bellow please.

Scale-dependent errors

The forecast errors are on the same scale as the data. Accuracy measures that are based only on et are therefore scale-dependent and cannot be used to make comparisons between series that involve different units. The two most commonly used scale-dependent measures are based on the absolute errors or squared errors:

Mean absolute error: $MAE = mean(|e_t|)$

Root mean squared error: $RMSE = \sqrt{mean(e_t^2)}$

When comparing forecast methods applied to a single time series, or to several time series with the same units, the MAE is popular as it is easy to both understand and compute. A forecast method that minimises the MAE will lead to forecasts of the median, while minimising the RMSE will lead to forecasts of the mean. Consequently, the RMSE is also widely used, despite being more difficult to interpret.

Percentage Errors

The percentage error is given by:

$$p_t = 100\frac{e_t}{y_t}$$

Percentage errors have the advantage of being unit-free, and so are frequently used to compare forecast performances between data sets. The most commonly used measure is:

Mean percentage error: $MPE = mean(p_t)$

Mean absolute percentage error: $MAPE = mean(|p_t|)$

Measures based on percentage errors have the disadvantage of being infinite or undefined if $y_t = 0$ for any t in the period of interest, and having extreme values if any y_t is close to zero. Another problem with percentage errors that is often overlooked is that they assume the unit of measurement has a meaningful zero.

Scaled Errors

Scaled errors were proposed by Hyndman & Koehler (2006) [21] as an alternative to using percentage errors when comparing forecast accuracy across series with different units. They proposed scaling the errors based on the training MAE from a simple forecast method.

For a non-seasonal time series, a useful way to define a scaled error uses naïve forecasts:

$$q_j = \frac{e_j}{\frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|}$$

Because the numerator and denominator both involve values on the scale of the original data, q_j is independent of the scale of the data. A scaled error is less than one if it arises from a better forecast than the average naive forecast computed on the training data. Conversely, it is greater than one if the forecast is worse than the average naive forecast computed on the training data. For seasonal time series, a scaled error can be defined using seasonal naive forecasts:

$$q_j = \frac{e_j}{\frac{1}{T-s}\sum_{t=s+1}^{T}|y_t - y_{t-s}|}$$

The mean absolute scaled error is simply

$MASE = mean(|q_j|)$.

Both evaluation metrics and residuals diagnostics are used. The most common evaluation metrics for forecasting are *RMSE*, which you may have used on regression problems; *MAPE*, as it is scale-independent and represents the ratio of error to actual values as a percent; and *MASE*, which indicates how well the forecast performs compared to a naïve average forecast.

4. Results of ROCOF modelling

In this section we bring results for ROCOF modelling using the respective above proposed new state space models.

4.1. SSOE models – innovations state space models for exponential smoothing

In this paper, the innovation state space models that capture various forms of the exponential smoothing methodology is used first. The method described by Hyndman et al [22] uses a state space framework for the automatic selection of exponential smoothening techniques for forecasting. The framework makes an assessment of best fit – comparing Akaike's Information Criterions (see [23, 24]).

ETS models

Based on the AIC criterion, the most appropriate ETS model for training data Y_T is first found. This model is ETS(A,Ad,A):

$$y_t = [1\,0\,1]\begin{bmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-m} \end{bmatrix} + \varepsilon_t \qquad \varepsilon_t \sim N(0, 0.116^2)$$

$$\begin{bmatrix} l_t \\ b_t \\ s_t \end{bmatrix} = \begin{bmatrix} 1 & 0.972 & 0 \\ 0 & 0.972 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-m} \end{bmatrix} + \begin{bmatrix} 2.59E-02 \\ 2.91E-03 \\ 1.19E-04 \end{bmatrix}\varepsilon_t$$

Even on the basis of all the data Yn, another type of ETS model was not chosen as optimal.

ETS(A,Ad,A):

$$y_t = [1\,0\,1]\begin{bmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-m} \end{bmatrix} + \varepsilon_t \qquad \varepsilon_t \sim N(0, 0.107^2)$$

$$\begin{bmatrix} l_t \\ b_t \\ s_t \end{bmatrix} = \begin{bmatrix} 1 & 0.964 & 0 \\ 0 & 0.964 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-m} \end{bmatrix} + \begin{bmatrix} 1.55E-02 \\ 1.39E-03 \\ 1.06E-04 \end{bmatrix}\varepsilon_t$$

Components of this state space model for both training and all data are presented in Fig. 6.

BATS models

Again, using the AIC criterion, we find the most appropriate BATS model for training data, which is model BATS(1, {0,0}, -, {12}) (see [23, 24]). Since $\omega = 1$, the observed data is not transformed. In addition, dumping parameter ϕ is zero, so the model becomes simpler as the drift component falls out. Moreover, the orders of both *AR* and *MA* processes are zero, so d_t is white noise ε_t. Components of this state space model for both training and all data are presented in Fig. 7.

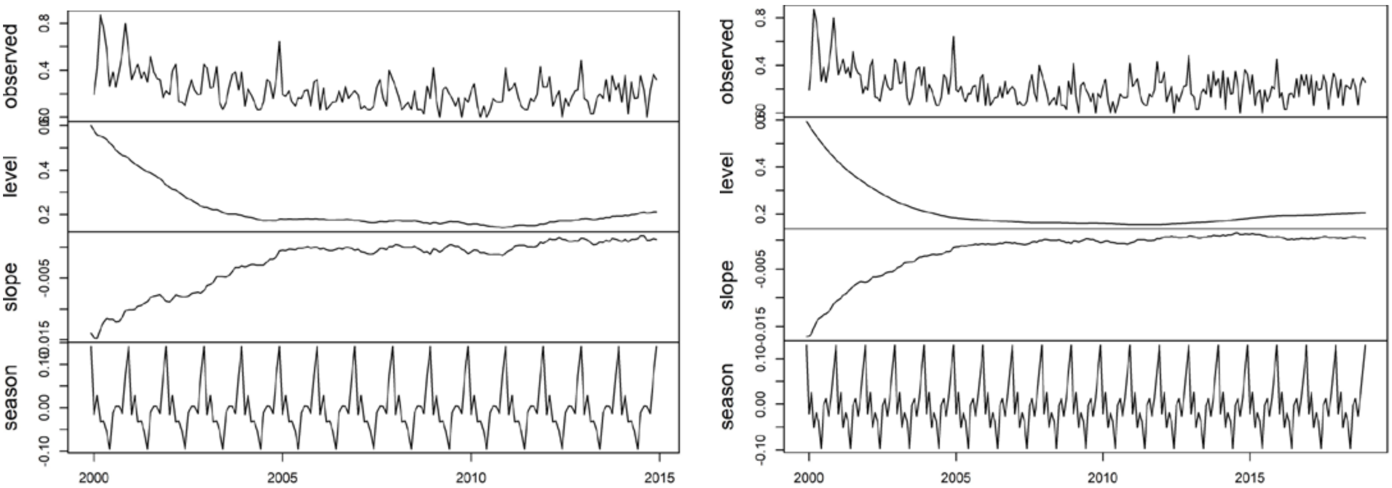


Fig. 6. Components of ETS method (left panel: training data, right panel: all data)

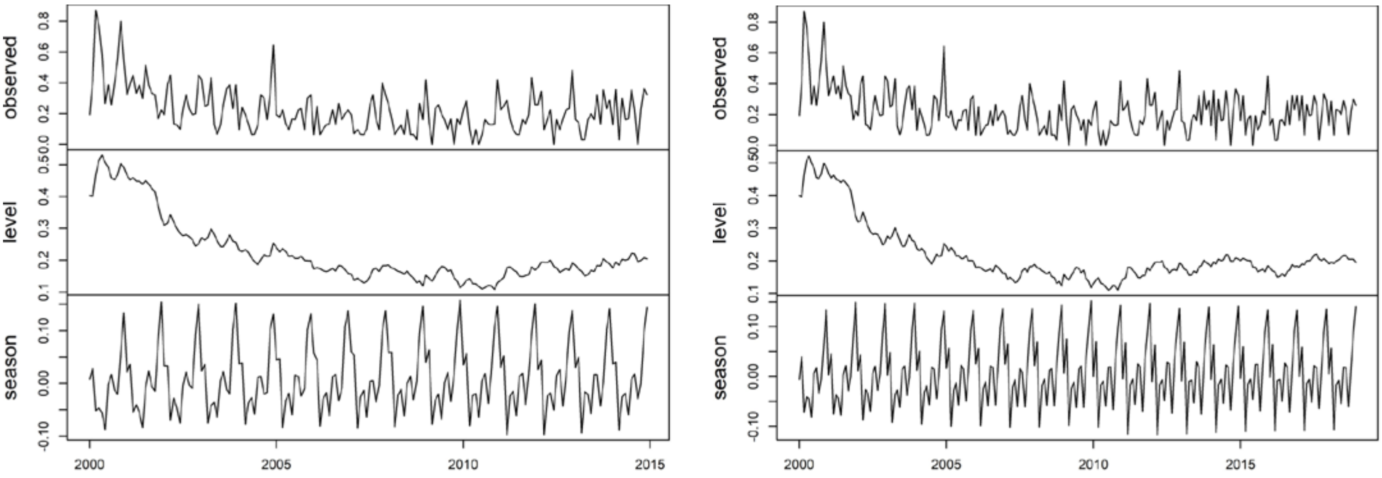


Fig. 7. Components of BATS method (left panel: training data, right panel: all data)

4.2. MSOE models – innovations state space models for exponential smoothing

The MSOE models are based on the state space form, the Kalman filter, and the associated smoother. The likelihood is constructed from the Kalman filter in terms of the one-step-ahead prediction errors and maximized with respect to the hyperparameters by numerical optimization. The score vector of the parameters can be obtained via a smoothing algorithm which is associated with the Kalman filter. Once the hyperparameters have been estimated, the filter is used to produce one-step-ahead prediction residuals which enables us to compute diagnostic statistics for normality, serial correlation, and goodness of fit. The smoother is used to estimate unobserved components, such as trends and seasonals, and to compute diagnostic statistics for detecting outliers and structural breaks.

BSM models

Next, the BSM model was fitted to the training data with the following results:

$$y_t = [10\,|\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0]\begin{bmatrix}\mu_{t-1}\\ s_{t-1}\end{bmatrix} + \varepsilon_t;\ \varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right)\ \sigma_\varepsilon = 1.07E-01$$

$$\begin{bmatrix}\mu_t\\ s_t\end{bmatrix} = \begin{bmatrix}F_\mu & 0\\ 0 & F_{seas}\end{bmatrix}\begin{bmatrix}\mu_{t-1}\\ s_{t-1}\end{bmatrix}\ \ \mu_t \sim N_2\left(0,\sum{}_\mu\right)\ \sum{}_\mu = \begin{bmatrix}\sigma_\eta^2 & 0\\ 0 & \sigma_\xi^2\end{bmatrix}$$

$$\sigma_\eta = 1.76E-05$$

$$\sigma_\xi = 5.43E-04$$

$$s_t \sim N_{11}\left(0,\sum{}_s\right)\ \sum{}_s = \begin{bmatrix}\sigma_\nu^2 & 0\\ 0 & 0\end{bmatrix}\ \sigma_\nu = 1.55E-02$$

where μt = (lt, bt), st = (s1,t, … , s11,t), $F_\mu = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$, $F_{seas} = \begin{bmatrix}-1 & -1 & \cdots & -1\\ 0 & & I_{10} & \end{bmatrix}$

and I10 is 10×10 matrix.

Next, the BSM model was fitted to the all of the available monthly data with the following results:

$$y_t = [10\,|\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0]\begin{bmatrix}\mu_{t-1}\\ s_{t-1}\end{bmatrix} + \varepsilon_t;\ \varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right)\ \sigma_\varepsilon = 9.84E-02$$

$$\begin{bmatrix}\mu_t\\ s_t\end{bmatrix} = \begin{bmatrix}F_\mu & 0\\ 0 & F_{seas}\end{bmatrix}\begin{bmatrix}\mu_{t-1}\\ s_{t-1}\end{bmatrix}\ \mu_t \sim N_2\left(0,\sum{}_\mu\right)\ \sum{}_\mu = \begin{bmatrix}\sigma_\eta^2 & 0\\ 0 & \sigma_\xi^2\end{bmatrix}\ \sigma_\eta = 2.99E-03$$

$$\sigma_\xi = 4.60E-04$$

$$s_t \sim N_{11}\left(0,\sum{}_s\right)\ \sum{}_s = \begin{bmatrix}\sigma_\nu^2 & 0\\ 0 & 0\end{bmatrix}\ \sigma_\nu = 1.50E-02$$

Components of this state space model for both training and all data are presented in Fig. 8.

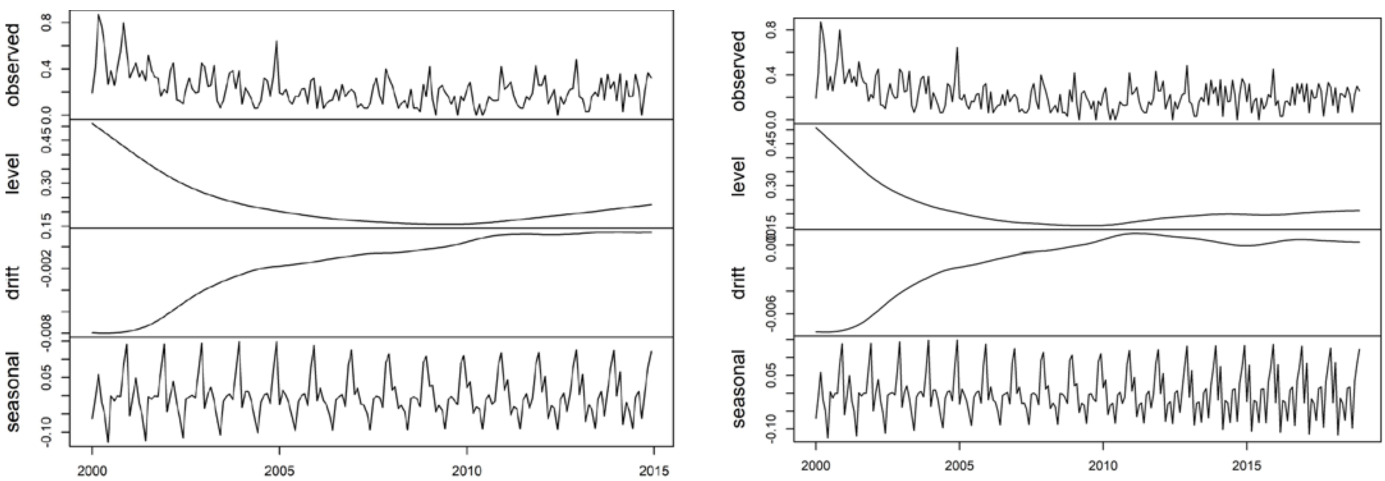


Fig. 8. Components of BSM method (left panel: training data, right panel: all data)

Level-BSM models

Next, the Level-BSM model was fitted to the training data with the following results:

$$y_t = [10\,|\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0]\begin{bmatrix} l_{t-1} \\ s_{t-1} \end{bmatrix} + \varepsilon_t\,;\, \varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right)\ \sigma_\varepsilon = 1.04E-01$$

$$\begin{bmatrix} l_t \\ s_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & F_{seas} \end{bmatrix}\begin{bmatrix} l_{t-1} \\ s_{t-1} \end{bmatrix}\ l_t \sim N\left(0,\sigma_\eta^2\right)\ \ \sigma_\eta = 1.71E-02$$

$$s_t \sim N_{11}\left(0,\sum{}_s\right)\ \sum{}_s = \begin{bmatrix} \sigma_\nu^2 & 0 \\ 0 & 0 \end{bmatrix}\sigma_\nu = 1.61E-02$$

where $s_t = (s_{1,t}, \ldots, s_{11,t})$, $F_{seas} = \begin{bmatrix} -1 & -1 & \cdots & -1 \\ 0 & & I_{10} & \end{bmatrix}$ and I_{10} is 10×10 matrix.

Next, the Level-BSM model was fitted to the all of the available monthly data with the following results:

$$y_t = [10\,|\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0]\begin{bmatrix} l_{t-1} \\ s_{t-1} \end{bmatrix} + \varepsilon_t\,;\ \varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right)\ \sigma_\varepsilon = 9.55E-02$$

$$\begin{bmatrix} l_t \\ s_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & F_{seas} \end{bmatrix}\begin{bmatrix} l_{t-1} \\ s_{t-1} \end{bmatrix}\ l_t \sim N\left(0,\sigma_\eta^2\right)\ \sigma_\eta = 1.56E-02$$

$$s_t \sim N_{11}\left(0,\sum{}_s\right)\ \sum{}_s = \begin{bmatrix} \sigma_\nu^2 & 0 \\ 0 & 0 \end{bmatrix}\ \sigma_\nu = 1.55E-02$$

Components of this state space model for both training and all data are presented in Fig. 9.

In all the graphs in Figures 6 – 9 we might observe very clearly specific development of the "level" value course and the affected ROCOF behaviour changes during the "seasonal" course.

5. Discussion to ROCOF modelling and performance measures

This and following part 5.1 are devoted to graphic and numerical results which help us to i) find the model which could fit a given type of the analyzed data best, ii) get an idea of the courses of single model types for both the training data and the testing data especially in the forecasting region, iii) numerically compare the results of single models for the courses – fitting, level, drift, forecast a 95% prediction intervals in the forecasting region for training and all data.

Based on the outcomes presented further there can be seen small but still existing difference in the applied models. Although one may say that divergence in these models is not significant it is not absolutely truth. Both of these models have their mathematical principles, therefore advantages and practical applicability especially in terms of these field data forms.

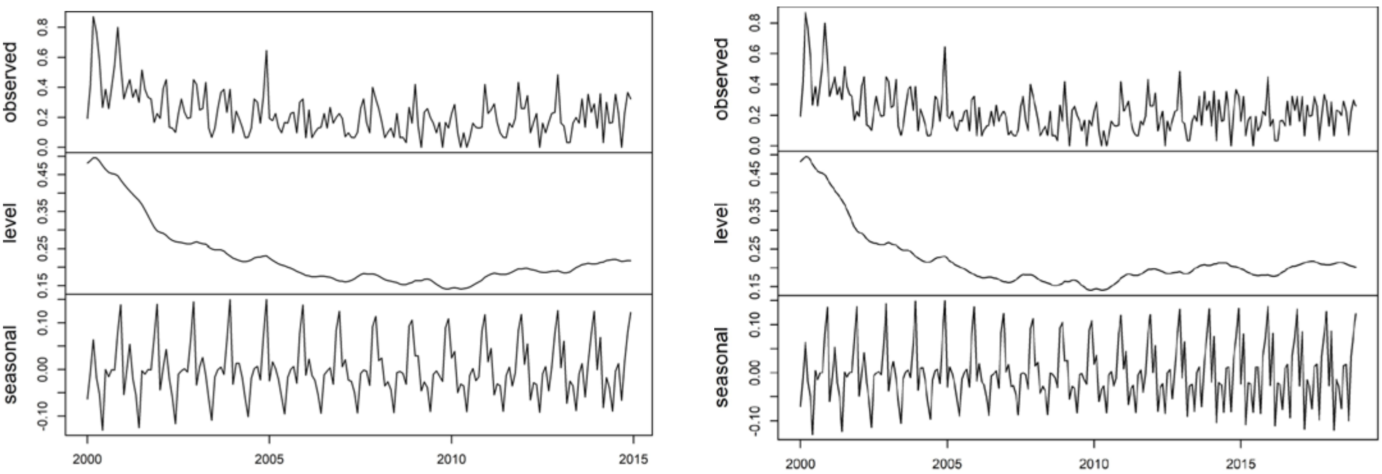


Fig. 9. Components of Level-BSM method (left panel: training data, right panel: all data)

Table 2. Forecast accuracy on test data with rating

Model	Type	RMSE	rRMSE	MAE	rMAE	MASE	rMASE
ets	SSOE	0.0886	3	0.0728	3	0.6480	3
bats	SSOE	0.0751	2	0.0637	2	0.5673	2
bsm	MSOE	0.0904	4	0.0742	4	0.6604	4
level-bsm	MSOE	0.0715	1	0.0595	1	0.5302	1

Table 3. Forecast accuracy on training data with rating

Model	Type	AIC	rAIC	BIC	rBIC	RMSE	rRMSE	MAE	rMAE	MASE	rMASE
ets	SSOE	178.3885	2	235.8617	2	0.1107	2	0.0842	2	0.7494	3
bats	SSOE	173.3949	1	179.7809	1	0.1110	3	0.0862	3	0.7433	2
bsm	MSOE	306.1965	3	318.9683	3	0.0980	1	0.0753	1	0.6706	1
level-bsm	MSOE	328.2960	4	337.8749	4	0.1223	4	0.0937	4	0.8342	4

Table 4. Forecast accuracy on all data with rating

Model	Type	AIC	rAIC	BIC	rBIC	RMSE	rRMSE	MAE	rMAE	MASE	rMASE
ets	SSOE	238.4869	1	300.2151	2	0.1032	2	0.0793	2	0.7613	3
bats	SSOE	243.7871	2	250.6458	1	0.1058	3	0.0814	3	0.6936	2
bsm	MSOE	492.1602	3	505.8776	3	0.0906	1	0.0689	1	0.6609	1
level-bsm	MSOE	514.8491	4	525.1372	4	0.1239	4	0.0956	4	0.9176	4

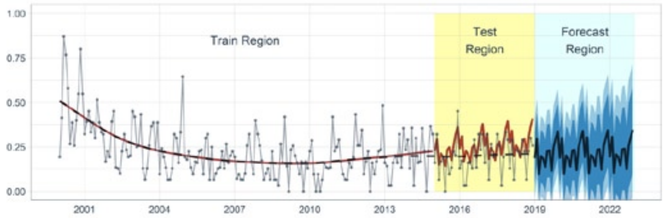


Fig. 10. Forecasting and testing – ETS model

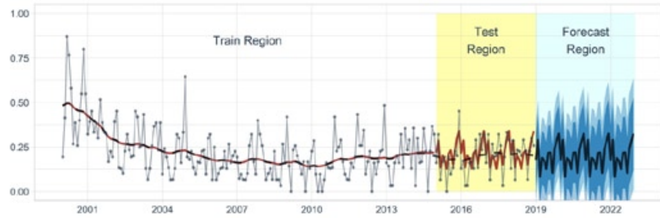


Fig. 11. Forecasting and testing – BATS model

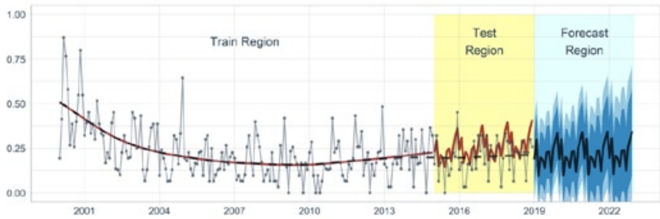


Fig. 12. Forecasting and testing – BSM model

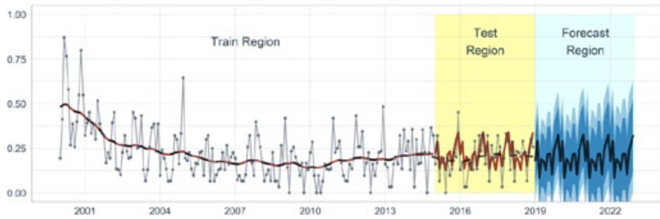


Fig. 13. Forecasting and testing – Level-BSM model

Table 5. Numerical values of ETS model components

Date	Nr. days	Nr. Failures	ROCOF	ETS Fit	ETS Level	ETS Drift	ETS Seasonal	ETS Forecast	ETS PI 95L	ETS PI 95U
1.1.2000	31	6	0.193548	0.570834	0.576737	-1.46E-02	-0.015703446			
1.2.2000	29	12	0.413793	0.590349	0.5579475	-1.47E-02	0.027815288			
1.3.2000	31	27	0.870968	0.511642	0.5529108	-1.33E-02	-0.031934835			
...										
1.10.2018	31	7	0.22580645	0.21918228	0.2049066	4.34E-04	0.014286635			
1.11.2018	30	9	0.3	0.27814934	0.205359	4.49E-04	0.072826575			
1.12.2018	31	8	0.25806452	0.33630317	0.2056702	3.24E-04	0.130503361			
1.1.2019	31	*~6*						0.184536	-0.02582	0.394893
1.2.2019	28	*~7*						0.232937	0.02258	0.443294
1.3.2019	31	*~5*						0.155278	-0.05508	0.365637
...										
1.10.2022	31	*~7*						0.2269827	0.014268325	0.4396971
1.11.2022	30	*~9*						0.2855796	0.072760766	0.4983983
1.12.2022	31	*~11*						0.3433112	0.130386422	0.5562359

Table 6. Numerical values of BSM model components

Date	Nr. days	Nr. Failures	ROCOF	BSM Fit	BSM Level	BSM Drift	BSM Seasonal	BSM Forecast	BSM PI 95L	BSM PI 95U
1.1.2000	31	6	0.193548	0.43750179	0.507319	-7.58E-03	-6.98E-02			
1.2.2000	29	12	0.413793	0.48954893	0.4999693	-7.58E-03	-1.04E-02			
1.3.2000	31	27	0.870968	0.55150313	0.4926839	-7.59E-03	5.88E-02			
...										
1.10.2018	31	7	0.22580645	0.244381	0.21077	2.61E-04	3.36E-02			
1.11.2018	30	9	0.3	0.293988	0.210966	2.59E-04	8.30E-02			
1.12.2018	31	8	0.25806452	0.333904	0.211154	2.59E-04	1.23E-01			
1.1.2019	31	~6						0.18671103	-0.036826694	0.4102488
1.2.2019	28	~8						0.2973334	0.073650738	0.5210161
1.3.2019	31	~3						0.09532622	-0.129205549	0.319858
...										
1.10.2022	31	~8						0.25667698	-0.109152479	0.6225064
1.11.2022	30	~9						0.30634686	-0.064033347	0.6767271
1.12.2022	31	~11						0.34633359	-0.027338535	0.7200057

5.1. Performance model measures

In this part we introduce numerical calculations of single measures which show the fitness of the proposed models. The forecast accuracy for the test data is put in Table 2, the forecast accuracy for the training data is put in Table 3, and the forecast accuracy for all the data is put in Table 4. The outcomes are always accompanied by rating of prioritization.

The modeling outputs along with prognoses are introduced in a graphical form first, see Fig. 10 – 13.

For each of the most suitable models of SSOE – ETS and MSOE – BSM groups we performed simulations within a forecast region in order to determine prediction intervals – PI (95% lower and upper (PI 95L and PI 95U). The graphical outcomes are put in Fig. 14 and 15, and the following numerical outcomes are put in Tables 5 and 6. The predicted numbers of water mains failures according to single models are put in *Italics*.

The introduced results show that graphical and numerical model forms are interesting, however, in their prognoses they complement each other significantly even when it comes to the predicted failure numbers.

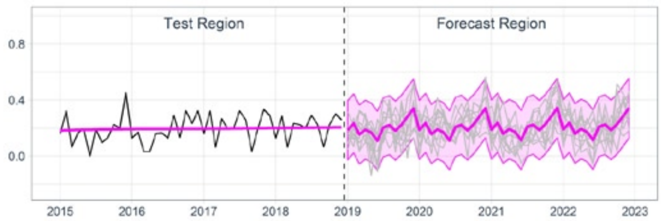


Fig. 14. Simulation inside the forecast region – ETS model – trend with 95% PI

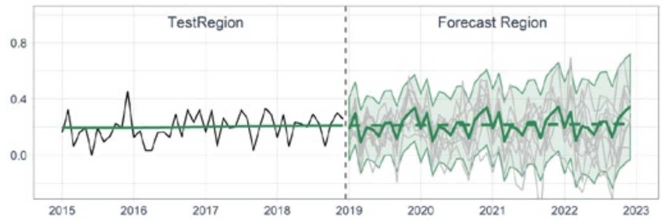


Fig. 15. Simulation inside the forecast region – BSM model – trend with 95% PI

6. Conclusions

In our article we introduce new and promising state-space models which seem to be very useful and suitable every time there is insufficient information on the system failures. The models SSOE and MSOE, or their representatives ETS and BSM, belong to the group of structural models which fit this type of analyzed field data very well, and also are suitable for forecasting failures and reliability measures development.

In our future work we would like to verify in practice, whether the achieved results agree and to what extent with our calculations. Luckily, as early as at the beginning of 2019, the first checks of our results conformed to our calculations significantly.

All the outcomes, both graphical and numerical, were acquired with the help of R Studio [31].

Acknowledgement

This article has been created with support of funds of partial development of UO-FVT-K202-MOBAUT.

References

1. Anderson B D O, Moore J B. Optimal filtering. Prentice Hall, New York: Englewood Cliffs 1979.
2. Antholzer S, Haltmeier M, Schwab J. Deep learning for photoacoustic tomography from sparse data. Inverse Problems in Science and Engineering 2019; 27(7): 987-1005, https://doi.org/10.1080/17415977.2018.1518444.
3. Box George E P, Cox D R. An analysis of transformations. Journal of the Royal Statistical Society 1964; 26(2): 211-252, https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.
4. Box George E P, Jenkins G M. Time Series Analysis: Forecasting and Control. San Francisco: Holden Day 1976.

5. Brockwell P J, Davis R A. Time Series: Theory and Methods, Berlin: Springer 1991. https://doi.org/10.1007/978-1-4419-0320-4
6. De Livera A M, Hyndman R J, Snyder R D. Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association 2011; 106(496): 1513-1527, https://doi.org/10.1198/jasa.2011.tm09771.
7. Dindarloo S. Reliability forecasting of a load-haul-dump machine: a comparative study of ARIMA and neural networks. Quality and Reliability Engineering International 2016; 32(4): 1545-1552, https://doi.org/10.1002/qre.1844.
8. Distefano S, Longo F, Trivedi K S. Investigating dynamic reliability and availability through state-space models. Computers & Mathematics with Applications 2012; 64(12): 3701-3716, https://doi.org/10.1016/j.camwa.2012.02.038.
9. Duan RX, Lin Y N, Zeng Y N, 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, https://doi.org/10.17531/ein.2018.4.7.
10. dos Santos TR, Gamerman D, Franco G D. Reliability Analysis via Non-Gaussian State-Space Models. IEEE Transactions on Reliability 2017; 66(2): 309-318, https://doi.org/10.1109/TR.2017.2670142.
11. Faes L, Porta A, Javorka M, Nollo G. Efficient computation of multiscale entropy over short biomedical time series based on linear state-space models. Complexity 2017: 1768264, https://doi.org/10.1155/2017/1768264.
12. Gardner J E S, McKenzie E. 1985. Forecasting trends in time series. Management Science 1985; 31(10): 1237-1246, https://doi.org/10.1287/mnsc.31.10.1237.
13. Glowacz A. Fault diagnosis of single-phase induction motor based on acoustic signals. Mechanical Systems and Signal Processing 2019; 117: 65-80, https://doi.org/10.1016/j.ymssp.2018.07.044.
14. Glowacz A. Fault Detection of Electric Impact Drills and Coffee Grinders Using Acoustic Signals. Sensors 2019; 19(2): 269, https://doi.org/10.3390/s19020269.
15. Gobbato M, Kosmatka J B, Conte J P. A recursive Bayesian approach for fatigue damage prognosis: An experimental validation at the reliability component level. Mechanical Systems and Signal Processing 2014; 45(2): 448-467, https://doi.org/10.1016/j.ymssp.2013.10.014.
16. Gobbato M, Conte J P, Kosmatka J B, Farrar C R. A reliability-based framework for fatigue damage prognosis of composite aircraft structures. Probabilistic Engineering Mechanics 2012; 29: 176-188, https://doi.org/10.1016/j.probengmech.2011.11.004.
17. Guo, Y J, Zhao Z B, Sun R B, Chen X F. Data-driven multiscale sparse representation for bearing fault diagnosis in wind turbine. Wind Energy 2019; 22(4): 587-604, https://doi.org/10.1002/we.2309.
18. Harvey A C. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge: University Press 1989, https://doi.org/10.1017/CBO9781107049994.
19. Hu Y W, Liu S J, Lu H T, Zhang H C. Remaining useful life model and assessment of mechanical products: a brief review and a note on the state space model method. Chinese Journal of Mechanical Engineering 2019; 32(1):15, https://doi.org/10.1186/s10033-019-0317-y.
20. Hyndman R J, Koehler A B, Ord J K, Snyder R D. Forecasting with Exponential Smoothing: The State Space Approach. Berlin Germany: Springer 2008, https://doi.org/10.1007/978-3-540-71918-2.
21. Hyndman R J, Koehler A B. Another look at measures of forecast accuracy. International Journal of Forecasting 2006; 22: 679-688, https://doi.org/10.1016/j.ijforecast.2006.03.001.
22. Hyndman R J, Koehler A B, Snyder R D, Grose S. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 2002; 18: 439-454, https://doi.org/10.1016/S0169-2070(01)00110-8.
23. Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2019). Forecast: Forecasting functions for time series and linear models. R package version 8.9, <URL: http://pkg.robjhyndman.com/forecast>.
24. Hyndman R J, Khandakar Y (2008). Automatic time series forecasting: the forecast package for R. Journal of Statistical Software 2008; 26(3): 1-22, https://doi.org/10.18637/jss.v027.i03.
25. Jimenez Cortadi A, Irigoien I, Boto F, Sierra B, Suarez A, Galar D. A statistical data-based approach to instability detection and wear prediction in radial turning processes. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20(3): 405-412, https://doi.org/10.17531/ein.2018.3.8.
26. Kontrec N Z, Milovanovic G V, Panic S R, Milosevic H. A reliability-based approach to nonrepairable spare part forecasting in aircraft maintenance system. Mathematical Problems in Engineering 2015: 731437, https://doi.org/10.1155/2015/731437.
27. Koucký M, Vališ D. Reliability of sequential systems with a restricted number of renewals. London: Taylor & Francis ltd. Risk, Reliability and Societal Safety, Vols. 1-3, ESREL 2007 Stavanger; 1845-1851.
28. Kozłowski E, Mazurkiewicz D, Żabinski T, Prucnal S, Sęp J. Assessment model of cutting tool condition for real-time supervision system. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21(4): 679-685, https://doi.org/10.17531/ein.2019.4.18.
29. Li H K, Zhou S, Kan H L, Cong M. Milling cutter condition reliability prediction based on state space model. Journal of Vibroengineering 2016; 18(7): 4312-4328, https://doi.org/10.21595/jve.2016.16648.
30. Pietrucha-Urbanik K, Studzinski A. Selected issues of costs and failure of pipes in an exemplary water supply system. Rocznik Ochrona Srodowiska 2016; 18: 616-627.
31. R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2015). URL https://www.R-project.org/ (accessed 16.06.03).
32. Rymarczyk T, Klosowski G. Innovative methods of neural reconstruction for tomographic images in maintenance of tank industrial reactors. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21(2): 261-267, https://doi.org/10.17531/ein.2019.2.10.
33. Snyder R D. Recursive estimation of dynamic linear models. Journal of the Royal Statistical Society 1985; 47(2): 272-276, https://doi.org/10.1111/j.2517-6161.1985.tb01355.x.
34. Snyder R. Discussion. International Journal of Forecasting 2006; 22(4): 673-676, https://doi.org/10.1016/j.ijforecast.2006.05.002.
35. Sousa H, Wang Y. Sparse representation approach to data compression for strain-based traffic load monitoring: A comparative study. Measurement 2018; 122: 630-637, https://doi.org/10.1016/j.measurement.2017.10.042.
36. Vališ D, Mazurkiewicz D. Application of selected Levy processes for degradation modelling of long range mine belt using real-time data. Archives of Civil and Mechanical Engineering 2018; 18(4): 1430-1440, https://doi.org/10.1016/j.acme.2018.05.006.

37. Vališ D, Žák L. Contribution to prediction of soft and hard failure occurrence in combustion engine using oil tribo data. Engineering Failure Analysis 2017; 82:583-598, https://doi.org/10.1016/j.engfailanal.2017.04.018.
38. Vališ D, Mazurkiewicz D, Forbelská M. Modelling of a Transport Belt Degradation Using State Space Model. In: Proceedings of the 2017 IEEE International Conference on Industrial Engineering & Engineering Management. Singapur: IEEE, 2017: 949-953, https://doi.org/10.1109/IEEM.2017.8290032.
39. Wang L, Lu Z R. Sensitivity-free damage identification based on incomplete modal data, sparse regularization and alternating minimization approach. Mechanical Systems and Signal Processing 2019; 120:43-68, https://doi.org/10.1016/j.ymssp.2018.10.013.
40. Whittle P. Hypothesis Testing in Time Series Analysis. Uppsala: Almquist and Wicksell 1951.

David VALIS
Department of Combat and Special Vehicles
University of Defence
Kounicova str. 65, 662-10 Brno, Czech Republic

Faculty of Transport and Computer Science,
University of Economics and Innovation,
Projektowa 4, 20-209 Lublin, Poland

Marie FORBELSKÁ
Department of Statistics and Operation Analysis
Mendel University in Brno
Zemědělská 1, 61300 Brno, Czech Republic

Zdeněk VINTR
Department of Combat and Special Vehicles
University of Defence
Kounicova str. 65, 662-10 Brno, Czech Republic

E-mails: david.valis@unob.cz, mforbelska@gmail.com, zdenek.vintr@unob.cz

Article citation info:
KOZŁOWSKI E, BORUCKA A, ŚWIDERSKI A. Application of the logistic regression for determining transition probability matrix of operating states in the transport systems. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 192–200, http://dx.doi.org/10.17531/ein.2020.2.2.

Edward KOZŁOWSKI
Anna BORUCKA
Andrzej ŚWIDERSKI

APPLICATION OF THE LOGISTIC REGRESSION FOR DETERMINING TRANSITION PROBABILITY MATRIX OF OPERATING STATES IN THE TRANSPORT SYSTEMS

ZASTOSOWANIE REGRESJI LOGISTYCZNEJ DO WYZNACZANIA MACIERZY PRAWDOPODOBIEŃSTW PRZEJŚĆ STANÓW EKSPLOATACYJNYCH W SYSTEMACH TRANSPORTOWYCH*

Transport companies can be regarded as a technical, organizational, economic and legal transport system. Maintaining the quality and continuity of the implementation of transport requisitions requires a high level of readiness of vehicles and staff (especially drivers). Managing and controlling the tasks being implemented is supported by mathematical models enabling to assess and determine the strategy regarding the actions undertaken. The support for managing processes relies mainly on the analysis of sequences of the subsequent activities (states). In many cases, this sequence of activities is modelled using stochastic processes that satisfy Markov property. Their classic application is only possible if the conditional probability distributions of future states are determined solely by the current operational state. The identification of such a stochastic process relies mainly on determining the probability matrix of interstate transitions. Unfortunately, in many cases the analyzed series of activities do not satisfy Markov property. In addition, the occurrence of the next state is affected by the length of time the system remains in the specified operating state. The article presents the method of constructing the matrix of probabilities of transitions between operational states. The values of this matrix depend on the time the object remains in the given state. The aim of the article was to present an alternative method of estimating the parameters of this matrix in a situation where the studied series does not satisfy Markov property. The logistic regression was used for this purpose.

Keywords: *logistic regression, transition probability matrix, Markov chains, transport system.*

Przedsiębiorstwa transportowe mogą być traktowane jako wyodrębniony pod względem technicznym, organizacyjnym, ekonomicznym i prawnym system transportowy. Zachowanie jakości i ciągłości realizacji zleceń przewozowych wymaga wysokiego poziomu gotowości pojazdów oraz personelu (szczególnie kierowców). Kontrolowanie i sterowanie realizowanymi zadaniami wspierane jest modelami matematycznymi, umożliwiającymi ocenę i określenie strategii dotyczącej podejmowanych działań. Wsparcie procesów zarządzania polega głównie na analizie sekwencji kolejnych, realizowanych czynności (stanów). W wielu przypadkach taki ciąg czynności jest modelowany za pomocą procesów stochastycznych, spełniających własność Markowa. Ich klasyczne zastosowanie możliwe jest tylko w przypadku, gdy warunkowe rozkłady prawdopodobieństwa przyszłych stanów są określone wyłącznie przez bieżący stan eksploatacyjny. Identyfikacja takiego procesu stochastycznego polega głównie na wyznaczeniu macierzy prawdopodobieństw przejść międzystanowych. Niestety w wielu przypadkach analizowane ciągi czynności nie spełniają własności Markowa. Dodatkowo, na wystąpienie kolejnego stanu wpływa długość interwału czasowego pozostania systemu w określonym stanie eksploatacyjnym. W artykule przedstawiono metodę konstrukcji macierzy prawdopodobieństw przejść pomiędzy stanami eksploatacyjnymi. Wartości tej macierzy zależą od czasu przebywania obiektu w danym stanie. Celem artykułu było zaprezentowanie alternatywnej metody estymacji parametrów tej macierzy w sytuacji, gdy badany szereg nie spełnia własności Markowa. Wykorzystano w tym celu regresję logistyczną.

Słowa kluczowe: *regresja logistyczna, macierz prawdopodobieństw przejść, łańcuchy Markowa, system transportowy.*

1. Introduction

The concept of a transport system, as defined by Grzywacz and Burnewicz [17] as well as Andrzejczak [3] is seen in this article as a segregated system of three subsystems: technical, organizational and economic-legal ones, creating a logical, internally balanced entirety, enabling to achieve a specific goal. This makes it possible to define the analyzed enterprise as the transport system, and assume the implementation of transport tasks as its operating goal.

Transport systems can be analyzed as multi-state sequences of subsequent planned and unplanned maintenance activities carried out by the transport system operator [27]. The construction of the models that describe them and allows the prediction of the operating state of the object used, allows planning of the maintenance strategy and control of the readiness of the machine [9, 21] and vehicles fleet [13, 26] etc. Modelling the functioning of technical objects using deterministic models is not always possible because the results (implementations) are affected by external disturbances (random factors), which make it

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

impossible to accurately predict subsequent states. In such cases, we model the behaviour of technical systems using probabilistic methods, in particular stochastic processes. An important class of stochastic processes are Markov processes. Some possibilities of applications of these processes are presented in papers [7, 24]. The essential condition of their use is satisfying Markov property: the future does not depend on the past when the present is known. Many analyzes assume a priori that the time series satisfies this property and its verification is omitted [55]. Only a few authors indicate the need for checking it [53] and eliminate cases that did not achieve this assumption [47]. An alternative for the systems that do not satisfy Markov property are classic reliability methods, allowing to determine empirical characteristics such as: renewal stream, renewal function, time to the next defect or intensity of the renewal stream [6, 14, 35] and calculating the main system assessment measures based on them [15, 23, 42]. In the literature, as part of similar studies, various models are presented [30, 43], including semi-Markov [7, 24] and also those using artificial neural networks [10, 33], factorization algorithm [31], fault trees [52] or reliability models [36, 42].

In many publications, the time of remaining in a particular operating state is not taken into account. The heterogeneity of the time interval between successive states may also cause non-fulfilment of Markov property. In this article, the logistic regression was used to estimate the conditional probabilities of the test object remaining in the individual operational states [49]. The logistic regression describes the relationship between a qualitative variable and one or more predictive variables [25, 46]. In the literature, logistic regression is used in medicine [5, 44], in computed tomography [46], to identify technical systems [25], in the area of corporate finance [39, 54], banking [1, 34] broadly understood investments [12, 29] and is used to assess the level of risk [2, 8, 45], in the social and demographic research [4, 41] and others [25, 40]. Regarding transport systems, logistic regression models are proposed primarily for assessing the road dangers demonstrated by the road accidents [18, 37], making the choices of routs in the transport network [32, 49] or analyzing the impact of selected factors on the implementation of transport processes [38, 48].

The paper shows the existence of a relationship between the duration of the operational state and the value of the probability of transition to the next state. In order to analyze the problem thoroughly and in detail, the introduction was firstly made, presenting the mathematical methods referred to in the article. The second chapter presents definitions regarding Markov chains and how to verify Markov property. The third chapter presents the method of estimating the transition probability matrix using logistic regression. Then, an example of the implementation of the proposed method using empirical data for a selected means of transport, carrying out transport tasks under the transport system (enterprise), has been presented. In the final stage, the results obtained are discussed, summaries of the analyzes carried out and directions for further research are indicated.

2. Markov chains

The state of an object is defined by its characteristic feature, a technical property that assigns it to a given operating system [50]. It is a vector whose components are physical values describing the object from the point of view of a given test [28]. In the literature, the state of a technical object is defined as the result of one and only one event in a series of experiments from a finite or countable set of pairs of mutually exclusive events [11, 54]. We use probability calculus tools and mathematical statistics to analyze technical systems [22, 53]. Let $(\Omega, \mathcal{F}, P)$ be a probabilistic space, N - a set of natural numbers, S - the space of the states of the analyzed phenomenon.

Definition 1 A sequence $\{X_t\}_{t\in N}$ of random variables $X_t : \Omega \to S$ for any $t \in N$ is called a stochastic process in discrete time [51, 54].

In the paper, we analyze the operating states in which the vehicles remain. The S set of operating states is a set of values of the stochastic process $\{X_t\}_{t\in N}$. At any $t \in N$ time, the object is in one of the possible states and $X_t(\omega) = x_t$, i.e. in the event of a ω random event occurring at the t moment, the system is in a state $x_t \in S$. In our research, we assume that the S set of states is a finite set and $S = \{s_1, s_2, \ldots, s_k\}$, $k \in N$, $2 \le k < \infty$. The $P(X_t = s_i) = p_i(t)$ value means the probability that the system at a moment $t \in N$ is in a state $s_i \in S$, $1 \le i \le k$, and $\sum_{i=1}^{k} p_i(t) = 1$.

Definition 2 A stochastic process $\{X_t\}_{t\in N}$ in discrete time is called a Markov chain if for each $n \in N$, of any moments $t_1, t_2, \ldots, t_n \in N$ satisfying the condition $t_1 < t_2 < \ldots < t_n$, and any $x_1, x_2, \ldots, x_n \in S$, the equality occurs [26, 47]:

$$P\left(X_{t_n} = x_n \middle| X_{t_{n-1}} = x_{n-1}, X_{t_{n-2}} = x_{n-2}, \ldots, X_{t_1} = x_1\right) = P\left(X_{t_n} = x_n \middle| X_{t_{n-1}} = x_{n-1}\right) \tag{1}$$

From the definition of the Markov chain it follows that the conditional distribution of the random variable X_n, for a given values $X_{t_0}, X_{t_1}, \ldots, X_{t_{n-1}}$ depends only on the last known value $X_{t_{n-1}}$. It is usually assumed that the t_i and t_{i+1} intervals are equal [16]. Below we assume that $t_n = n \in N$. If $\{X_t\}_{t\in N}$ is a heterogeneous Markov chain, then for any $t \in N$ and $1 \le i, j \le k$, the value:

$$P\left(X_t = s_j \middle| X_{t-1} = s_i\right) = p_{ij}(t) \tag{2}$$

we call the probability of transition from s_i state at the moment $t-1$ to the s_j state at moment t. Therefore, for the chains satisfying the Markov property (1), the conditional probability distributions of the future process states are determined only by its current state and moment t, regardless of the past (they are conditionally independent of the past states). The matrix $P(t) = \left[p_{ij}(t)\right]_{1\le i,j\le k}$ satisfying the condition $\sum_{j=1}^{k} p_{ij}(t) = 1$ for $t \in N$ and $1 \le i \le k$ is called the matrix of probabilities of transitions in one step at the moment t [7, 47, 51].

Definition 3 The Markov chain $\{X_t\}_{t\in N}$ is a homogeneous Markov chain, if the $p_{ij}(t)$ probabilities of transition do not depend on the moment $t \in N$.

Thus, for a homogeneous Markov chain $p_{ij}(t) = p_{ij}$ for $1 \le i, j \le k$ and any moment $t \in N$. The matrix $P = \left[p_{ij}\right]_{1\le i,j\le k}$ satisfying the condition $\sum_{j=1}^{k} p_{ij} = 1$, $1 \le i \le k$ we call the transition probability matrix in one step. For a homogeneous Markov chain, the probabilities of transition from a s_i state at a t moment to the s_j state at the $t+n$ moment is determined using the formula [13, 24]:

$$P\left(X_{t+n} = s_j \middle| X_t = s_i\right) = p_{ij}^{(n)} \tag{3}$$

where $\left[p_{ij}^{(n)}\right]_{1\le i,j\le k} = P^n$, $n \in N$ is the matrix of probability of transition in n steps.

Definition 4 If $\{X_t\}_{t\in N}$ is a homogeneous Markov chain and there is a distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ where $\pi_i \ge 0$, $1 \le i \le k$ and $\sum_{i=1}^{k} \pi_i = 1$ satisfying the equation:

$$\pi P = \pi \tag{4}$$

then the distribution π is called the stationary distribution of the homogeneous Markov chain.

The stationary property means that if at some $n \in N$ moment the chain reaches a stationary distribution, then for each subsequent moment greater than n the distribution will remain the same. We determine the stationary distribution by solving the system of equations [16, 22]:

$$\sum_{j=1}^{k} \pi_j \cdot p_{ij} = \pi_i \tag{5}$$

$$\sum_{i=1}^{k} \pi_i = 1 \tag{6}$$

and $\pi_i \ge 0$ for $1 \le i \le k$.

An important role in the studying of processes using Markov chains is played by its boundary properties, especially the boundary probability $p_j(n)$ and $p_{ij}^{(n)}$ at $n \to \infty$, which describe the probabilistic behaviour of the process after a long time [16, 22].

Theorem 1 (ergodic) Let $\{X_t\}_{t\in N}$ be a homogeneous Markov chain with a finite number of states $k < \infty$ $\left(k = \#S = \#\{i : s_i \in S\}\right)$, then:

a) a vector $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ exists such that $\pi_i \ge 0$ for $1 \le i \le k$ and $\sum_{i=1}^{k} \pi_i = 1$;

b) for any $1 \le i, j \le k$

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n)} \;;$$

c) π vector is the solution to the equation (6).

Below, the method of estimating the transition probability matrix for the homogeneous Markov chain and the way of verifying Markov property will be presented. Let $\{x_t\}_{0\le t\le n}$ be the realization of the Markov chain. The value $n_i = \#\{t : x_t = s_i, 0 \le t \le n\}$ means the number of moments for which the system remained in the state s_i for $1 \le i \le k$, where $\sum_{i=1}^{k} n_i = n$, while the value $n_{ij} = \#\{t : x_t = s_i, x_{t+1} = s_j, 0 \le t \le n-1\}$ means the number of transitions from the state s_i to the state s_j for $1 \le i, j \le k$ and $\sum_{j=1}^{k} n_{ij} = n_i$.

Assuming that the Markov property is satisfied, we estimate the transition probability matrix. The estimator of the transition probability from state s_i to state s_j we determine from the formula $\hat{p}_{ij} = \frac{n_{ij}}{n_i}$ for $1 \le i, j \le k$.

We use a χ^2 goodness of fit test to verify Markov property. At the significance level $\alpha \in (0,1)$ we create a working hypothesis:

$H_0 : P\left(X_t = x \mid X_{t-1} = y, X_{t-2} = z\right) = P\left(X_t = x \mid X_{t-1} = y\right)$ (the chain $\{X_t\}_{t\in N}$ satisfies Markov property)

and an alternative hypothesis:

$H_1 : P\left(X_t = x \mid X_{t-1} = y, X_{t-2} = z\right) \ne P\left(X_t = x \mid X_{t-1} = y\right)$ (the chain $\{X_t\}_{t\in N}$ does not satisfy Markov property),

where $x, y, z \in S$. As a measure of discrepancy between $P\left(X_t = x \mid X_{t-1} = y, X_{t-2} = z\right)$ and $P\left(X_t = x \mid X_{t-1} = y\right)$ distributions we choose the test statistics:

$$\chi_e^2 = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{v=1}^{k} \frac{\left(n_{ijv} - n_{ij}\hat{p}_{jv}\right)^2}{n_{ij}\hat{p}_{jv}} \tag{7}$$

which has a χ^2 distribution with k^3 degrees of freedom.

The value $n_{ijv} = \#\{t : x_t = s_i, x_{t+1} = s_j, x_{t+2} = s_v, 0 \le t \le n-2\}$ means the number of transitions from state s_i to state s_j and next to state s_v for $1 \le i, j, v \le k$. From the tables for the χ^2 distribution with k^3 degrees of freedom we determine the quantile of order $1-\alpha$, which we denote as $\chi^2\left(1-\alpha, k^3\right)$. If $\chi_e^2 < \chi^2\left(1-\alpha, k^3\right)$, then at the significance level α there are no grounds for rejecting the working hypothesis H_0, so we assume that the chain $\{X_t\}_{t\in N}$ satisfies Markov property. On the other hand, if $\chi_e^2 \ge \chi^2\left(1-\alpha, k^3\right)$, then at the significance level α we reject the working hypothesis H_0 in favour of the alternative hypothesis, thus the chain $\{X_t\}_{t\in N}$ does not satisfy Markov property.

3. Logistic regression

In many cases, the stochastic process $\{X_t\}_{t\in N}$ does not satisfy Markov property. The realization of the process $\{X_t\}_{t\in N}$ depends on additional factors. In transport and logistics systems, the time of remaining in a specific state directly affects the probability of transition to other states. Below, the authors used logistic regression to define the transition probability matrix, which depends on the time the object remains in a given state. In the case under consideration a random variable $X_t, t \in N$ describing the state of the system can get k possible realizations. Because we are considering moments for which the system state changes, so if at the moment $t \in N$ the system was in a state $s_i \in S$, then in the moment $t+\tau$ the system may take the states $S \setminus \{s_i\}$ (a random variable $X_{t+\tau}$ may get $k-1$ possible realizations). Determination of the transition probability is possible thanks to polynomial logistic regression [19, 25, 44, 46]. One of the levels should be taken as a reference. For each state $s_i \in S$, $1 \le i \le k$ we determine the probabilities of transition to the other states:

$$P\left(X_{t+\tau} = s_j \middle| X_t = s_i\right) = p_{ij}(\tau), \tag{8}$$

where $t,\tau \in N$ and $s_j \in S \setminus \{s_i\}$. From the set $S \setminus \{s_i\}$, we select the reference state $s_q \in S \setminus \{s_i\}$ and determine the logarithms of chances for the remaining states:

$$ln \frac{P\left(X_{t+\tau} = s_j \middle| X_t = s_i\right)}{P\left(X_{t+\tau} = s_q \middle| X_t = s_i\right)} = \beta_{ij}^0 + \beta_{ij}^1 \tau \tag{9}$$

for $s_j \in S \setminus \{s_i, s_q\}$. The values of structural parameters in the model (9) are determined using the maximum likelihood method [20, 25, 49]. Wald's test is used to assess the significance of model parameters.

We determine transition probabilities for states $s_j \in S \setminus \{s_i, s_q\}$ using the formula:

$$p_{ij}(\tau) = \frac{e^{\beta_{ij}^0 + \beta_{ij}^1 \tau}}{1 + \sum\limits_{\substack{1 \le v \le k \\ v \ne i,\, v \ne q}} e^{\beta_{iv}^0 + \beta_{iv}^1 \tau}}, \tag{10}$$

while for the reference state s_q the probability is:

$$p_{iq}(\tau) = \frac{1}{1 + \sum\limits_{\substack{1 \le v \le k \\ v \ne i,\, v \ne q}} e^{\beta_{iv}^0 + \beta_{iv}^1 \tau}}. \tag{11}$$

From the formulas (10) - (11) we obtain that the logarithm of the chances ratio for any two states $s_j, s_v \in S \setminus \{s_i, s_q\}$ is equal to:

$$ln \frac{P\left(X_{t+\tau} = s_j \middle| X_t = s_i\right)}{P\left(X_{t+\tau} = s_v \middle| X_t = s_i\right)} = ln \frac{p_{ij}(\tau)}{p_{iv}(\tau)} = \left(\beta_{ij}^0 - \beta_{iv}^0\right) + \left(\beta_{ij}^1 - \beta_{iv}^1\right)\tau \;.(12)$$

4. Estimation of transition probability matrix for the selected means of transport

The subject of the study was a Belgian distribution department operating for the benefit of hypermarket chains. Transport services are carried out every day, 24 hours a day, which is why it is important to schedule transport properly, taking into account the availability of employed staff (especially drivers), as well as the readiness of vehicles.

The study used data from the company's fleet management system that integrates, processes and archives readings from the vehicle's GPS transmitter, tachograph, CAN (Controller Area Network) and on-board computer. It allows to obtain data on the driver and the vehicle in real time, allows tracking of the position and movement of cars, visualization of the location of vehicles and trailers on the map, monitoring of driving and resting times, etc. The information concerned 69 Iveco Stralis EEV 460 trucks. The collected data was segregated and 10 operational states realized by heavy goods vehicles were analyzed. These activities are detailed in tab. 1.

Table 1. Operating states highlighted in the study

No.	Name of operational state
S1	Availability
S2	Driving
S3	Manipulation
S4	Repair
S5	Maintenance
S6	Parking
S7	Layover
S8	Off-loading
S9	Refuelling
S10	Loading

The study presented in the article was carried out for one randomly selected vehicle. Markov property were checked. The χ^2 test was used for this purpose. The test statistics was 2672.74, and $p-value = 2.2*10^{-16}$. This means that at the significance level $\alpha = 0.001$, the working hypothesis should be rejected in favour of the alternative hypothesis, therefore the analyzed stochastic process does not satisfy Markov property. Nevertheless, the transition probability matrix of realization of the process $\{X_t\}_{1 \le t \le n}$, $n = 6822$ was estimated (for comparison purposes), which is presented graphically in Fig. 1, while the values of this matrix are presented in Table 2.

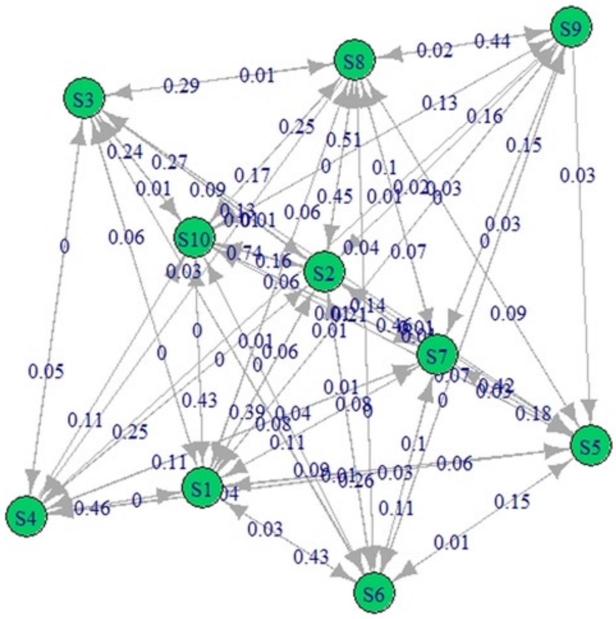


Fig. 1. Graph of the interstate transitions according to Markov chain

By solving equation (4), the boundary probabilities were estimated. The values of these probabilities are presented in Tab. 3.

Because Markov property was not satisfied for the analyzed data, the parameters of the transition probability matrix were estimated using the polynomial logistic regression model. The impact of the time

Table 2. Matrix of transition probabilities for the Markov chain

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	0	0.392	0.002	0.002	0.012	0.032	0.113	0.009	0.007	0.431
S2	0.060	0	0.128	0.003	0.015	0.015	0.142	0.453	0.022	0.162
S3	0.065	0.274	0	0.003	0.009	0.029	0.085	0.294	0	0.241
S4	0.456	0.246	0.053	0	0.035	0	0.105	0	0	0.105
S5	0.056	0.416	0.011	0.034	0	0.146	0.180	0.090	0	0.067
S6	0.433	0.264	0.082	0	0.014	0	0.111	0.005	0.005	0.087
S7	0.084	0.456	0.003	0.044	0.027	0.103	0	0.074	0.001	0.208
S8	0.062	0.510	0.008	0.009	0.005	0.044	0.097	0	0.019	0.247
S9	0.035	0.163	0	0.012	0.035	0.035	0.151	0.442	0	0.128
S10	0.004	0.735	0.008	0	0.012	0.001	0.064	0.173	0.003	0

Table 3. The values of boundary probabilities for the Markov chain

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
π_i	0.064	0.337	0.050	0.008	0.013	0.031	0.099	0.212	0.013	0.173

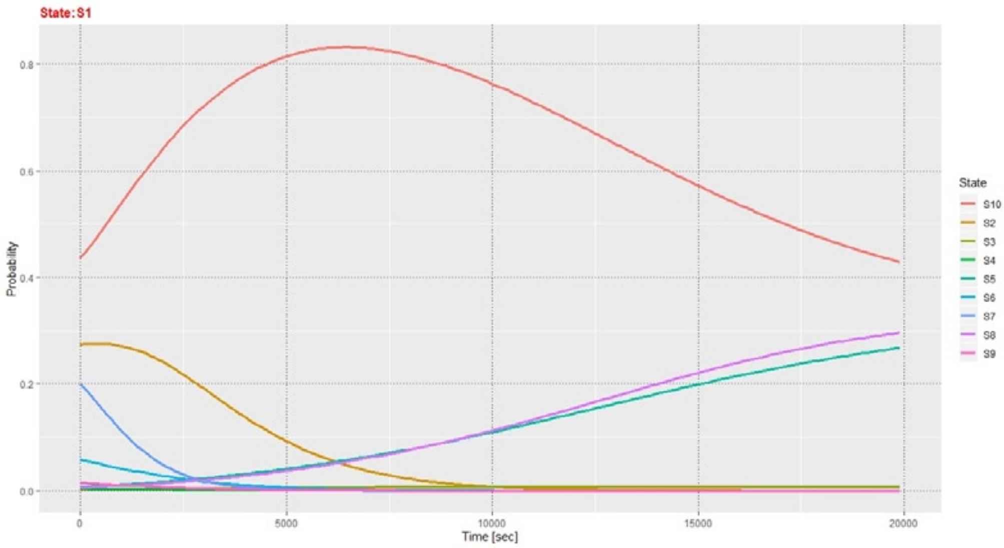


Fig. 2. Relationship between the transition probability from Availability state and its duration

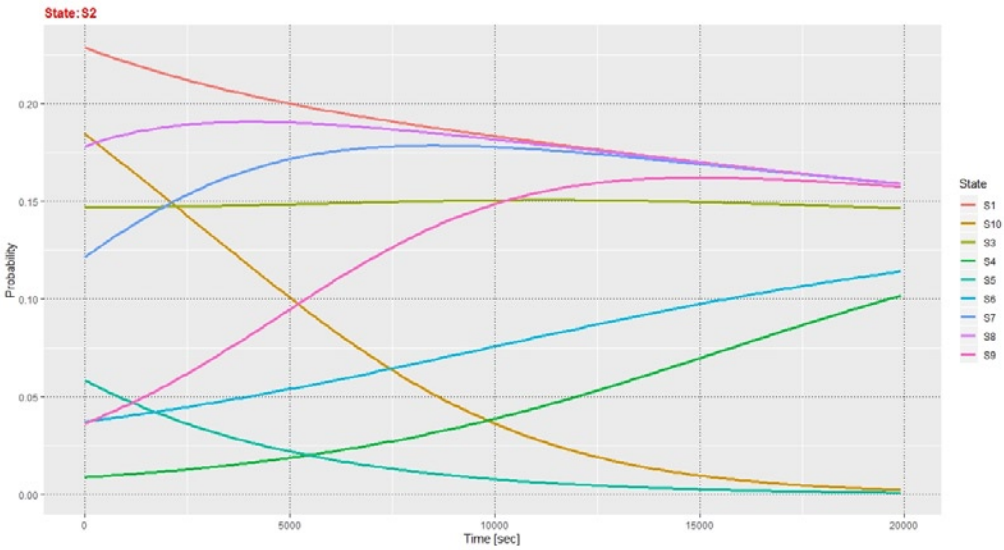


Fig. 3. Relationship between the transition probability from Driving state and its duration

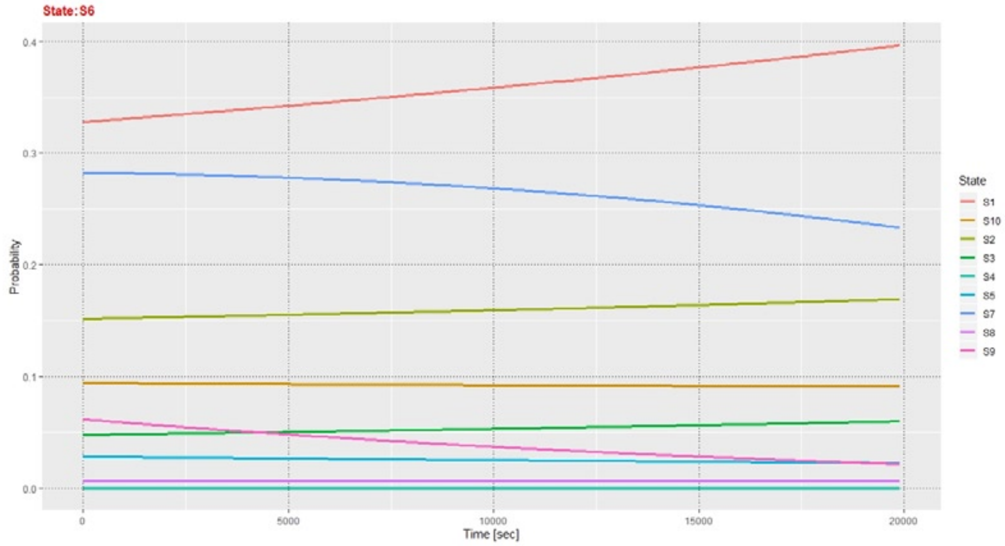


Fig. 4. Relationship between the transition probability from Parking state and its duration

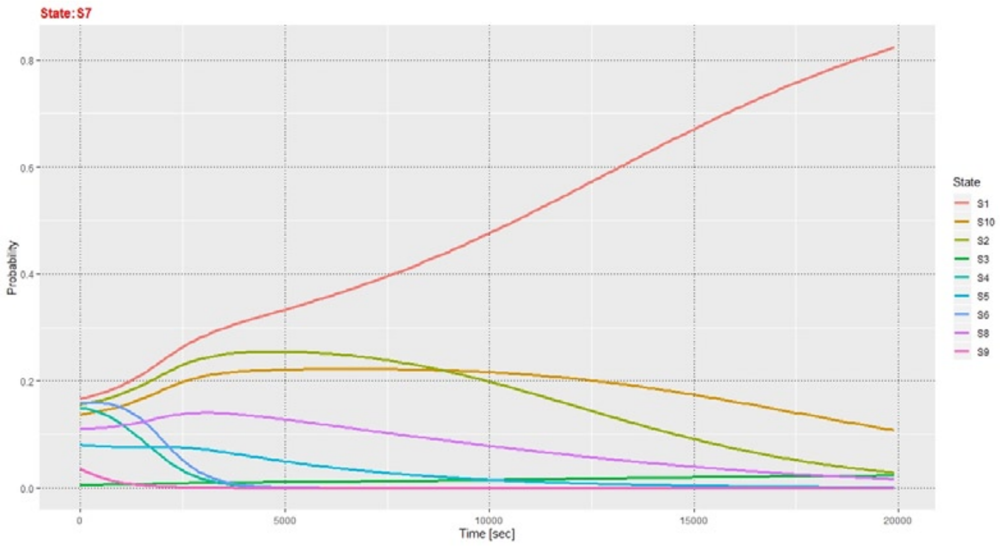


Fig. 5. Relationship between the transition probability from Layover state and its duration

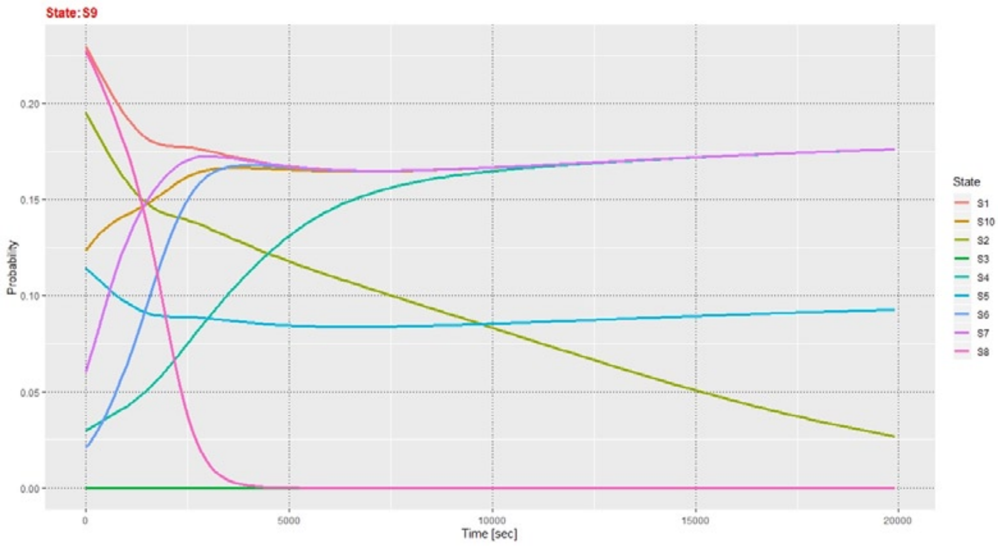


Fig. 6. Relationship between the transition probability from Refuelling state and its duration

while remaining in a specific operating state on the probability of transition to other states was investigated. It was assumed that the probability at the moment $t+\tau$ is a value conditionally dependent on the state in which the object was at the moment t and the duration length τ, as well as that after a time τ the system does not return to it. For each state, 8 logistic regression equations (9) were determined, which describe the relationships for nine possible transitions. The significance of structural parameters was examined by Wald test. As a result of using polynomial logistic regression for each system state a transition probability matrix was obtained, which depends on the duration time τ.

Then, the graphs were drawn illustrating the change in the transition probabilities depending on the duration time τ in a given operational state. For the selected states: availability, driving, parking, layover, refuelling, the relationships between the transition probabilities and the duration time are shown on Fig. 2 - Fig. 6.

The above graphs show the dependence of the transition probability value from a given operational state to the next, depending on the time the vehicle spends in it. From the graphs it is possible to see that the values of these probabilities are not constant, which shows impossibility of the use of the classical approach when estimating the transition probability matrix as for the Markov chain. The approach proposed by the authors shows how to determine the matrix of transition probabilities for the case when for a specific state the duration time significantly affects the values of the elements of this matrix. The variability of the transition probabilities is justified and reflects the specificity of the implementation of transport processes, which are partly determined by legal regulations concerning, for example, the driver's working time, as well as deadlines resulting from the operating strategy implemented in the company, regulating the periods of repairs and inspections.

The solutions presented are helpful in addition to developing a method that allows assessing the readiness of the system to carry out transport tasks. Operating states can be classified as states of suitability and unsuitability, and it is possible to determine the technical readiness factor as the sum of appropriate probabilities of reliability states.

5. Conclusion

The article estimates the matrix of transition probabilities to the identified operating states in which the tested vehicle was. The use of Markov chains is popular in such estimates, which requires the condition of the lack of memory of the analyzed process to be met. In the presented case this property was not fulfilled. In addition, it was shown that the probability of transition to a given operating state is conditionally dependent on the state in which the object was and the length of time spent in it. Therefore, an alternative method was proposed for their estimation. For this purpose, a polynomial logistic regression model was used. The transition probabilities were obtained, whose values for a given state differed depending on the length of time the vehicle stayed in the previous state.

The results obtained were compared with the values obtained according to the Markov chain - for which they are constant - showing that using so calculated transition probability matrix, when the Markov property is not met, may lead to the erroneous conclusions.

The proposed logistic regression model allows to conduct short-term forecasts regarding the implementation of the transport process. The assessment of the probability transition depending on previously carried out activities supports the process of scheduling transport tasks, as well as planning in the field of vehicle maintenance.

As part of further research, it is worth extending the proposed method by determining the values of estimators and assessing the boundary probabilities of transitions for individual states over a long period of time. This will allow a comprehensive evaluation of the system's functioning, as well as determining the level of readiness to carry out transport tasks. The division of operational states into states of suitableness and unsuitableness will allow to calculate the technical readiness coefficient as the sum of the respective boundary probabilities of the reliability states. The presented solution can also be used for modelling the driving cycles of heavy vehicles, which directly reflect the real operating conditions of the engine or components on the chassis dynamometer.

References

1. Abid L, Masmoudi A, Zouari-Ghorbel S. The Consumer Loan's Payment Default Predictive Model: an Application of the Logistic Regression and the Discriminant Analysis in a Tunisian Commercial Bank. Journal of the Knowledge Economy 2018; 9(3): 948-962, https://doi.org/10.1007/s13132-016-0382-8.
2. Adha M R, Nurrohmah S, Abdullah S. Multinomial Logistic Regression and Spline Regression for Credit Risk Modelling, Journal of Physics: Conference Series 2018; 1108, https://doi.org/10.1088/1742-6596/1108/1/012019.
3. Andrzejczak K. Metody prognozowania w modelowaniu eksploatacji środków transportu. Poznań: Wydawnictwo Politechniki Poznańskiej, 2013.
4. Ari E. Using Multinomial Logistic Regression to Examine the Relationship Between Children's Work Status and Demographic Characteristics, Ekonomi ve Yönetim Araştırmaları Dergisi 2016; 4(1): 77-93.
5. Bertens l C, Karel K G, Rutten F H, van Mourik Y, Hoes A W, Reitsma J B. A nomogram was developed to enhance the use of multinomial logistic regression modeling in diagnostic research. Journal of Clinical Epidemiology 2016; 71: 51-57, https://doi.org/10.1016/j.jclinepi.2015.10.016.
6. Bobrowski D. Modele i metody matematyczne teorii niezawodności w przykładach i zadaniach. Warszawa: Wydawnictwa Naukowo - Techniczne, 1985.
7. Borucka A, Niewczas A, Hasilova K. Forecasting the readiness of special vehicles using the semi-Markov model. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (4): 662-669, https://doi.org/10.17531/ein.2019.4.16.
8. Calster B V, Hoorde V K, Vergouwe Y, Bobdiwala S, Condous G., Kirk E., Bourne T, Steyerberg E W. Validation and updating of risk models based on multinomial logistic regression. Diagnostic and Prognostic Research 2017; 1:2, https://doi.org/10.1186/s41512-016-0002-x.
9. Cartella F, Lemeire J, Dimiccoli L, Sahli H. Hidden Semi-Markov Models for Predictive Maintenance, Mathematical Problems in Engineering 2015; 278120, https://doi.org/10.1155/2015/278120.
10. Ferdowsi A, Challita U, Saad W. Deep Learning for Reliable Mobile Edge Analytics in Intelligent Transportation Systems: An Overview, IEEE Vehicular Technology Magazine 2019; 14(1): 62-70, https://doi.org/10.1109/MVT.2018.2883777.
11. Fisz M. Rachunek prawdopodobieństwa i statystyka matematyczna. Warszawa: PWN, 1967.
12. Gautam V, Vaidya R R. Evidence on the determinants of investment-cash flow sensitivity Indian Economic Review December 2018; 53(1-2): 229-244, https://doi.org/10.1007/s41775-018-0021-3.

13. Girtler J, Ślęzak M. Application of the theory of semi-Markov processes to the development of a reliability model of an automotive vehicle, Archiwum Motoryzacji 2012; 2: 15-27, https://doi.org/10.5604/1234754X.1066721.
14. Gniedenko B.W., Bielajew J.K., Sołowiew A.D. Metody matematyczne w teorii niezawodności. Warszawa: Wydawnictwa Naukowo-Techniczne, 1965.
15. Gola A. Reliability analysis of reconfigurable manufacturing system structures using computer simulation methods, Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019, 21 (1): 90-102, https://doi.org/10.17531/ein.2019.1.11.
16. Grabski F. Semi-Markov Processes: Applications in System Reliability and Maintenance, Elsevier, 2014, https://doi.org/10.1016/B978-0-12-800518-7.00004-1.
17. Grzywacz W., Burnewicz J. Ekonomika transportu. Gdańsk: Wydawnictwa Komunikacji i Łączności, 1989.
18. Halim H, Ramli M I, Adisasmita S A, Aly S H, Prasetijonternational J A. Relationship model between accident factors and the traffic accident severity using logistic regression model, Journal of Engineering and Science Application 2018; 4(2), 169-181.
19. Hastie T., Tibshirani R., Friedman J., The elements of statistical learning, Springer-Verlag New York Inc., 2009, https://doi.org/10.1007/978-0-387-84858-7.
20. James G., Witten D., Hastie T., Tibshirani R., An introduction to statistical learning, Springer-Verlag GmbH, 2013, https://doi.org/10.1007/978-1-4614-7138-7.
21. Jasiulewicz-Kaczmarek M., Gola A., Maintenance 4.0 Technologies for Sustainable Manufacturing - an Overview, IFAC PapersOnLine 2019; 52-10: 91-96, https://doi.org/10.1016/j.ifacol.2019.10.005.
22. Jaźwiński J, Grabski F. Niektóre problemy modelowania systemów transportowych. Radom: Instytut Technologii Eksploatacji, 2003.
23. Jia X, Cui L, Xing L. New insights into reliability problems for supply chains management based on conventional reliability model. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (3): 465-470, https://doi.org/10.17531/ein.2018.3.16.
24. Knopik L, Migawa K. Semi-Markov system model for minimal repair maintenance. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (2): 256-260, https://doi.org/10.17531/ein.2019.2.9.
25. Kozłowski E, Mazurkiewicz D, Żabiński T, Prucnal S, Sęp J. Assessment model of cutting tool condition for real-time supervision system. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (4): 679-685, https://doi.org/10.17531/ein.2019.4.18.
26. Landowski B. Example of applying markov decision process o model vehicle maintenance process. Journal of KONES Powertrain and Transport 2013;20(4): 209-218, https://doi.org/10.5604/12314005.1137618.
27. Lewitowicz J, Zieja M, Szelmanowski A, Pazur A. New Measures of Operational Readiness for Multi-states Avionics Integrated Systems with Reduced Efficiency. In: Hanus R., Mazur D., Kreischer C. (eds) Methods and Techniques of Signal Processing in Physical Measurements. Lecture Notes in Electrical Engineering, 2019; 548: 165-180, https://doi.org/10.1007/978-3-030-11187-8_13.
28. Lewitowicz J., Kustroń K. Podstawy eksploatacji statków powietrznych. Warszawa: Wydawnictwo ITWL, 2003.
29. Linhares F S, Costa F M D, Beiruth A X. Earnings management and investment efficiency. Revista Brasileira de Gestão de Negócios 2018; 20(2): 295-310, https://doi.org/10.7819/rbgn.v20i2.3180.
30. Macián V, Tormos B, Herrero J. Maintenance management balanced scorecard approach for urban transport fleets. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (2): 226-236, https://doi.org/10.17531/ein.2019.2.6.
31. 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, https://doi.org/10.17531/ein.2018.2.14.
32. Murat Y S, Uludag N. Route choice modelling in urban transportation networks using fuzzy logic and logistic regression methods, Journal of Scientific & Industrial Research; 67(01): 19-27.
33. Nabian M A, Meidani H. Deep Learning for Accelerated Seismic Reliability Analysis of Transportation Networks, Computer-Aided Civil and Infrastructure Engineering 2018; 33: 443-458, https://doi.org/10.1111/mice.12359.
34. Nguyen T. Understanding Municipal Bond Ratings Using Ordered Multinomial Logistic Regression. Senior Independent Study Theses 2018; 8171.
35. Nicholson A, Schmöcker J, Bell M, Iida Y. Assessing Transport Reliability: Malevolence and User Knowledge. In The Network Reliability of Transport, Emerald Group Publishing Limited 2003; 1-22, https://doi.org/10.1108/9781786359544-001.
36. Niewczas A, Rymarz J, Debicka E. Stages of operating vehicles with respect to operational efficiency using city buses as an example. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (1): 21-27, https://doi.org/10.17531/ein.2019.1.3.
37. Nowakowska M. Spatial and temporal aspects of prior and likelihood data choices for Bayesian models in road traffic safety analyses. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19 (1): 68-75, https://doi.org/10.17531/ein.2017.1.10.
38. Odeck J, Bråthen S. A meta-analysis of DEA and SFA studies of the technical efficiency of seaports: A comparison of fixed and random-effects regression models. Transportation Research Part A: Policy and Practice 2012; 46(10): 1574-1585, https://doi.org/10.1016/j.tra.2012.08.006.
39. Oliva F L. A maturity model for enterprise risk management, International Journal of Production Economics 2016; 173: 66-79, https://doi.org/10.1016/j.ijpe.2015.12.007.
40. Osuji M N, Mejeha R O, Simonyan J B, Gbolagun A S. Determinants of Enterprise Preference of Cassava-based farmers by Gender in Southeast, Nigeria. Research Journal of Food and Nutrition 2018; 2(4): PP 45-52.
41. Park M, Wu A D. Investigating differential options functioning using multinomial logistic regression. International Journal of Quantitative Research in Education 2017; 4(1-2): 94 - 119, https://doi.org/10.1504/IJQRE.2017.086511.
42. Paś J, Klimczak T. Selected issues of the reliability and operational assessment of a fire alarm system. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (4): 553-561, https://doi.org/10.17531/ein.2019.4.3.
43. Paś J, Rosiński A. Selected issues regarding the reliability-operational assessment of electronic transport systems with regard to electromagnetic interference. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19 (3): 375-381, https://doi.org/10.17531/ein.2017.3.8.
44. Peter D, Hart A. Multinomial Logistic Regression Examination of TV Time and Two Different Measures of Obesity in U.S. Adults. American Journal of Public Health Research 2018; 6(5): 222-226, https://doi.org/10.12691/ajphr-6-5-3.
45. Raco A, Pesce A, Fraschetti F, D'Andrea G, Polli F M, Acqui M, Frati A. Risk of Postoperative Performance Status Worsening after Resection of Lesions Involving the Motor Pathway: A Multinomial Logistic Regression Model. Journal of Neurological Surgery, Part A: Central

European Neurosurgery 2018; 79(06): 453-463, https://doi.org/10.1055/s-0037-1617756.
46. Rymarczyk T.; Kozłowski E.; Kłosowski G.; Niderla K. Logistic Regression for Machine Learning in Process Tomography. Sensors 2019, 19, 3400, https://doi.org/10.3390/s19153400.
47. Sanusi W, Jemain A A, Zin W Z, Zahari M. The Drought Characteristics Using the First-Order Homogeneous Markov Chain of Monthly Rainfall Data in Peninsular Malaysia, Water Resources Management 2015; 29(5): 1523-1539, https://doi.org/10.1007/s11269-014-0892-8.
48. Schwartz J, Guasch J L, Wilmsmeier G, Stokenberga A. Logistics, Transport and Food Prices in LAC: Policy Guidance for Improving Efficiency and Reducing Costs. Latin America and Caribbean Region Environment and Water Resources occasional paper series 2009; 2.
49. Shao C, Xia J C, Lin T G, Goulias K G, Chen C. Logistic regression models for the nearest train station choice: A comparison of captive and non-captive stations, Case Studies on Transport Policy 2015; 3(4): 382-391, https://doi.org/10.1016/j.cstp.2015.06.002.
50. Sowa A. Stan obiektu jako wieloznaczne pojęcie we współczesnej eksploatacji technicznej. Czasopismo Techniczne. Mechanika 2012; 109(7-M): 269-278.
51. Stewart W.J. Probability, Markov Chains, Queues, and Simulation, 2009, https://doi.org/10.2307/j.ctvcm4gtc.
52. Szkoda M, Kaczor G. Reliability and availability assessment of diesel locomotive using fault tree analysis. Archives of Transport 2016; 40(4): 65-75, https://doi.org/10.5604/08669546.1225470.
53. Wang Z, Huang X, Liang Y J. Oil-gas reservoir lithofacies stochastic modeling based on one- to three-dimensional Markov chains. Journal of Central South University 2018; 25(6): 1399-1408, https://doi.org/10.1007/s11771-018-3835-3.
54. Zhu Y, Xie C, Sun B, Wang G-J, Yan X-G. Predicting China's SME Credit Risk in Supply Chain Financing by Logistic Regression, Artificial Neural Network and Hybrid Models. Sustainability 2016; 8(5): 1-17, https://doi.org/10.3390/su8050433.
55. Ziółkowski J., Borucka A. Markov model in logistic management of enterprise. Journal of Konbin 2016; 38(1): 271-290, https://doi.org/10.1515/jok-2016-0027.

Edward KOZŁOWSKI
Lublin University of Technology
Faculty of Management
ul. Nadbystrzycka 38, 20-618 Lublin, Poland

Anna BORUCKA
Military University Of Technology
ul. gen. Sylwestra Kaliskiego 2
01-476 Warsaw, Poland

Andrzej ŚWIDERSKI
Motor Transport Institute
ul. Jagiellońska 80
03-301 Warsaw, Poland

E-mails: e.kozlovski@pollub.pl, anna.borucka@wat.edu.pl, andrzej.swiderski@its.waw.pl

Article citation info:
MACIÁN V, TORMOS B, BASTIDAS S, PÉREZ T. Improved fleet operation and maintenance through the use of low viscosity engine oils: fuel economy and oil performance. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 201–211, http://dx.doi.org/10.17531/ein.2020.2.3.

Vicente MACIÁN
Bernardo TORMOS
Sophia BASTIDAS
Tomás PÉREZ

IMPROVED FLEET OPERATION AND MAINTENANCE THROUGH THE USE OF LOW VISCOSITY ENGINE OILS: FUEL ECONOMY AND OIL PERFORMANCE

POPRAWA EFEKTYWNOŚCI EKSPLOATACJI FLOTY DZIĘKI ZASTOSOWANIU OLEJÓW SILNIKOWYCH O NISKIEJ LEPKOŚCI: OSZCZĘDNOŚĆ PALIWA I WYDAJNOŚĆ OLEJU

For heavy-duty vehicles and road transportation, fuel consumption and associated CO_2 emissions have been of great concern, which has led to the development and implementation of technologies to reduce their impact on the environment. Low viscosity engine oils have arisen as one proven cost-effective solution to increase the engine efficiency; however, for the heavy-duty vehicle segment, engine protection against wear is a priority for end-users, and therefore there is some reluctance to the use of that new oil formulations. In this study, eight lubricant oils, representative of the HTHS viscosity reduction that heavy-duty oils have been undergoing and new API CK-4 and FA-4 categories, were evaluated for fuel economy, oil performance and engine wear, in a long-term test involving a fleet of 49 heavy-duty vehicles of four different engine technologies, some of them with diesel fuel and others with compressed natural gas. Results of fuel economy were positive for most of the buses' models. Regarding oil performance and wear, most of the formulations were found to be suitable for extended oil drain intervals (ODI); and although no alarming results were found, overall performance of the formulations of the fourth stage could lead to significant wear if the oil drain interval is extended. In this study, it should be noted that some of the information has been presented by the authors in other publications, here they are presented with the purpose of complementing the new results and summarize the entire test.

Keywords: *low viscosity engine oils, fuel economy, real working conditions, oil performance, engine wear.*

W przypadku pojazdów o dużej ładowności, i transportu drogowego w ogóle, ważny problem stanowi zużycie paliwa i związana z nim emisja CO2, które wymagają opracowywania i wdrażania technologii zmniejszających ich wpływ na środowisko. Jednym ze sprawdzonych i finansowo korzystnych rozwiązań w tym zakresie są oleje silnikowe o niskiej lepkości, które zwiększają wydajność silnika. Jednak w segmencie pojazdów ciężkich, priorytetem dla użytkowników końcowych jest ochrona silnika przed zużyciem, co pociąga za sobą niechęć do stosowania tych nowych preparatów olejowych. W pracy, przedstawiono badania ośmiu olejów smarowych o obniżonej lepkości wysokotemperaturowej HTHS reprezentatywnych dla produkowanych obecnie kategorii olejów do pojazdów ciężkich, z uwzględnieniem nowych kategorii oleju API CK-4 i FA-4. Oleje oceniano pod kątem oszczędności paliwa, wydajności oleju i zużycia silnika w badaniu długoterminowym obejmującym flotę 49 autobusów o silnikach opartych na różnych technologiach, z których część była zasilana olejem napędowym a część sprzężonym gazem ziemnym. Wyniki dotyczące oszczędności zużycia paliwa były pozytywne dla większości modeli badanych autobusów. Jeśli chodzi o wydajność oleju i zużycie silnika, większość preparatów okazała się być przystosowana do dłuższych okresów wymiany oleju; chociaż nie zaobserwowano niepokojących wyników, to jednak ogólna wydajność preparatów w czwartym etapie testu, mogłaby prowadzić do znacznego zużycia silnika przy wydłużeniu okresu wymiany oleju. Część przedstawionych danych publikowaliśmy już w innych pracach. Niniejszy artykuł stanowi uzupełnienie poprzednich wyników oraz podsumowanie całego badania.

Słowa kluczowe: *oleje silnikowe o niskiej lepkości, oszczędność paliwa, rzeczywiste warunki pracy, wydajność oleju, zużycie silnika.*

1. Introduction

Reduction of engine oils' viscosity has been one of the main cost-effective alternatives to reduce fuel consumption of internal combustion engines [19, 3] and therefore to reduce CO_2 emissions to levels required by standards and governments' laws [18, 6, 31, 35]. The evolution of the oil formulations has been accompanied by developments in the engine that also seek to improve the engine efficiency and reduce emissions; however, many of these solutions are either harmful to the lubricant oil or require it to work under more severe working conditions and contamination [22, 13, 38, 8]. In this way, new oil formulations are aimed to reduce the engine parasitic losses, associated with friction, but also ensure proper lubrication of the engine and wear protection.

In the last years, standards that classify the engine oils for the heavy-duty vehicles (HDVs) segment have been updated to account for changes in the working conditions of the vehicles and to comply with the environmental regulations. In this regard, the American Petroleum Institute (API) launched two new oil categories at the end of 2016 to define oils for fuel economy; these are API CK-4 with an unchanged HTHS viscosity limit above 3.5 cP, and API FA-4 category with a reduced viscosity between 2.9 and 3.2 cP [20]. The Society of Automotive Engineers (SAE) on the other hand, has released new oil viscosity grades for the HDVs and light-duty vehicles (LDVs) since 2013, the SAE 16 with HTHS viscosity of 2.3 cP, followed by SAE 12 and 8, released in 2015, with viscosity limits of 2 and 1.7 cP, respectively [25].

From the lubrication theory, these reduced viscosity values can help to decrease friction mechanical losses when the working conditions promote the appearance of hydrodynamic lubrication, and therefore friction is only determined by the shearing of the oil [22, 30]. Nonetheless, internal combustion engines comprise complex tribological pairs working under the different regimes of lubrication during one cycle. The piston-cylinder assembly, journal bearings and valve train, are therefore the main contributors to friction losses [30, 10, 24, 29]. In this way, reducing the oil viscosity comes with the risk of not being able to create a fluid film of lubricant, due to lower film thicknesses, and therefore promote mixed and boundary lubrication, where there is direct contact between the surfaces [21].

Wear of the engine components is a direct consequence of mixed and boundary lubrication, determining the performance and lifetime of the engine. Taking into account that HDVs usually work under low-medium speeds and high loads, an optimum balance between lubricant oil viscosity and wear protection must be found if fuel economy and greenhouse gas (GHG) emissions reduction is the goal, accompanied with the enhancement of maintenance practices, reduction of operation and maintenance costs and unscheduled downtimes. A recent work developed in a LDV, by instance, demonstrated that very low HTHS viscosity values are no longer able to significantly reduce fuel consumption and even increased the engine wear [26].

Low viscosity engine oils (LVEOs) have been extensively evaluated for fuel economy and performance, especially for LDVs under stationary and real working conditions [16, 11, 27, 26, 5]. For the HDVs segment, on the other hand, studies are more limited [35, 36, 32, 33] due to requirements and costs inherent to the vehicles operation. The study presented here aims to develop an exhaustive analysis of the effect of different engine oil formulations over fuel economy, performance and engine wear, under real working conditions of a public service HDVs fleet; and in this way highlight the importance of choosing the correct oil formulation, which in turn has a direct impact on the operation and maintenance of the fleet [23]. Attending to advances in the development of lubricant oils, with ever lower HTHS viscosity and the new API categories, a long-term test was developed, divided in four stages with a duration of one oil drain interval (ODI) each one, adding up more than 5 million km travelled. Taking into account that the oil performance varies with the engine design and working conditions, four different engine technologies were included in the test, three with diesel fuel and one with compressed natural gas (CNG). Regarding the oil formulations, they were evaluated a total of eight engine oils representative of the HTHS viscosity reduction that HDVs oils have been undergoing and different additives packages. Results demonstrated that reduction of HTHS viscosity and implementation of the new API CK-4 y FA-4 categories give positive results in fuel consumption reduction for three of the four buses' models. In terms of oil performance and engine wear, although no alarming results were found, it is possible to conclude that for new oil formulations, of the fourth stage, extending the ODI could lead to significant engine wear.

2. Methodology and materials

With the aim of developing a complete and comprehensive study of the effect of LVEOs in the HDVs segment, the work presented here comprises two sections. The first one to evaluate the effect of LVEOs over fuel consumption, and the second to assess the oil performance, degradation and engine wear, and consequently the potential effect on ODI (enlarging or decreasing) and thus in maintenance costs. Due to the multiple variables that accompany a real world test, such as environmental conditions, driving behavior, characteristics of the road (rolling resistance, elevation profile, traffic), number of passengers, that have an effect over fuel consumption, but that cannot be controlled, it was decided to develop a long-term test divided in stages (see Figure 1) of one ODI each one (30.000 km with a duration of about one year), completing a total of four stages. This test definition allowed obtaining a considerable amount of data with statistically significant results. Furthermore, the test definition, regarding oil formulation and buses distribution, was developed for each stage once results of the previous one were collected and analyzed. This was done with the aim of selecting the appropriate oil formulations for each bus model and prevent any possible engine damage.

Table 1. Characteristics of the bus models

Bus model	Diesel I	Diesel II	CNG	Diesel III
Model year	2008	2010	2007	2010
Length/width/height [m]	17.94/2.55/3	11.95/2.55/3	12/2.5/3.3	12/2.55/3.15
Vehicle approximate weight [tons]	17.5	12.7	12.1	11
Passenger capacity seated/stand	45/95	25/60	30/63	25/66
Engine displacement [cm^3]	11967	7200	11967	9300
Cylinders	6	6	6	5
Emissions certification level	Euro IV	Euro V	EEV	EEV
Power [kW]	220@2200 rpm	210@2200 rpm	180@2200 rpm	170@1900 rpm
Torque [Nm]	1600@1100 rpm	1100@1100 rpm	880@1000 rpm	1050@1500 rpm
Thermal loading [W/mm^2]	2.85	3.97	2.33	2.56
Oil sump volume [l]	31	29	33	31
EGR	NO	NO	NO	YES
Valve train configuration	OHV Roller follower (hardened steel)	OHV Cam follower (steel)	OHV Cam follower (steel)	OHV Cam follower (steel)
Engine oil specification recommended by manufacturer	API CJ-4 ACEA E4/E6/E7/E9			

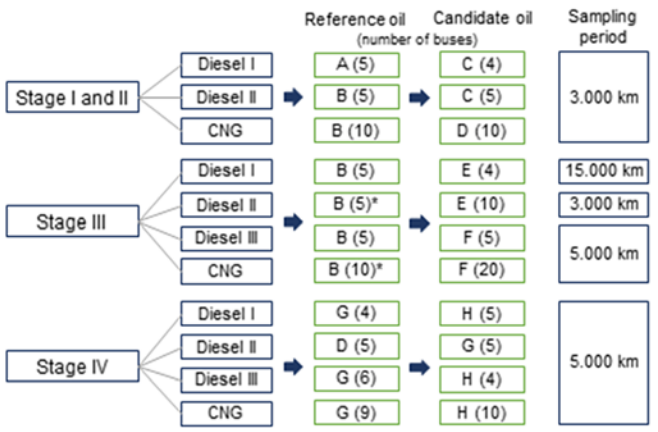


*Fig. 1. Distribution of the buses and oils for the four stages of the test. *Results of stages I and II were used as reference oil*

2.1. Fuel consumption test definition

Four engine models were selected for the test, three of them employing diesel fuel and one with CNG. Diesel I, Diesel II and Diesel III buses comply with the Euro IV, Euro V and EEV (Enhanced Environmentally friendly Vehicles) emissions certification level, respectively, while CNG buses comply with EEV. The main characteristics of the buses are presented in Table 1. In order to evaluate the LVEOs effect over fuel consumption, data of refueling and distance traveled by each bus was collected and recorded, in a daily basis, through GPS and information from the Computerized Maintenance Management System (CMMS) of the EMT of Valencia. Fuel consumption data was then calculated from these two parameters and averaged over the entire ODI period. In addition, average ambient temperature was recorded daily, as it has a significant effect on fuel consumption given, by instance, the use of A/C systems during summer season.

As it is shown in Figure 1, for the first and second stage of the test, 39 buses were employed, Diesel I, Diesel II and CNG. Following the recommendations of the buses OEMs it was decided to start with four formulations commercially available, splitting the group of buses in two, one using reference oil, and the other with an oil of lower viscosity. For the third stage, 10 Diesel III buses were included in the test and they were selected two oil formulations, as candidate oils, with HTHS viscosity values below the European Automobile Manufacturers' Association (ACEA) specifications for HDVs. Furthermore, data from the previous two stages were used for the analysis of results, that is, during the third stage, Diesel II and CNG buses only used candidate oil and results were compared with that of the reference oils used during the first and second stages. Given the novelty of the new API CK-4 and FA-4 oil categories, thought to improve fuel economy, it was decided to include two oil formulations complying with these API categories for the fourth stage of the test. In this way, except for Diesel II buses that continued to use a commercially available lubricant as reference oil, the rest of the bus models used API CK-4 oil as reference and API FA-4 as candidate.

Regarding the fuel used during tests, diesel fuel met the requirements of the standard UNE-EN 590 [4], while CNG fuel followed the requirements of the Commission Directive 2001/27/EC [7].

For each stage and once the ODI was completed, data recorded was submitted to Analysis of Variance (ANOVA). This is a statistical tool that allows evaluating the significance of the variables included in the analysis; these are daily temperature, oil mileage, month, oil formulation, service route and oil refill volume. As it was mentioned in Section 2, there exists variables inherent to the operation of the bus fleet that cannot be controlled and thus, the tests could not be randomized. The ANOVA analysis was therefore applied to each bus model separately avoiding the effect of different engine technologies, service routes, etc. on fuel consumption. Results of the ANOVA are percentage differences of the average fuel consumption between reference and candidate oil and their statistical significance with a 95% confidence.

2.2. Oil performance and degradation test definition

Oil formulations employed along the test are presented in Table 2 along with their main characteristics. Oils A, B, C and D are commercially available, while the rest of them are non-commercial new formulations for testing. Oil samples were collected following the

Table 2. Main properties of the oil formulations

Oil	A	B	C	D	E	F	G	H
SAE grade	15W40	10W40	5W30	5W30	5W30	5W30	5W30	5W30
API category	CI-4	CJ-4		CJ-4			CK-4	FA-4
API base oil group	I	III	III+IV	III+IV	III+IV	III+IV	III+IV	III+IV
kV@40°C [cSt]	108	96	71	68	55	54	68	55
kV@100°C [cSt]	14.5	14.4	12.5	11.7	9.8	9.4	12.6	10.5
HTHS@150°C [cP]	4.08	3.85	3.59	3.58	3.05	3.05	3.57	3.10
VI	>141	>145	>158	<169	>158	<169	168	165
TBN [mgKOH/g]	10	10	16	10	16	9	11	12
SAPS level	High	Mid	High	Mid	High	Mid	Mid	Mid
Calcium (Ca) [ppm]	1980	3357	5241	2329	3965	2282	1248	1312
Magnesium (Mg) [ppm]	704	15	27	100	25	61	847	895
Sodium (Na) [ppm]	nd	nd	nd	2.93	nd	nd	3.68	3.83
Barium (Ba) [ppm]	nd	nd	nd	10.03	nd	0.02	0.94	0.79
Phosphorus (P) [ppm]	731	1219	1064	712	960	712	715	764
Zinc (Zn) [ppm]	966	1534	1371	749	1132	827	784	835
Boron (B) [ppm]	195.10	4.50	301.95	3.63	897.82	12.36	311.85	333.69
Molybdenum (Mo) [ppm]	nd	nd	nd	nd	nd	nd	42.32	45.13
Used on stages	1,2	1,2,3	1,2	1,2,4	3	3	4	4

Table 3. Analytical techniques

Group	Parameter	Technique	Standard	Device
Degradation	Kinematic viscosity @100°C	Capillary viscometer	ASTM D-445	Cannon-Fenske opaque capillary viscometers
	TAN	Automatic potentiometric titrator	ASTM D-664	Thermo Scientific Orion 950 ROSS® FAST QC™ Titrator
	TBN		ASTM D-2896 (Fresh oil)	
			ASTM D-4739 (Used oil)	
	Oxidation	FT-IR Spectrometer	CMT-0080-11	iPal FTIR spectrophotometer, A2 Technologies
	Nitration			
	Aminic and antiwear additives		CMT-0120-12	
Wear	Wear metals and additives	ICP-OES Spectrometer	ASTM D-5185	iCap 7000 Series ICP Spectrometer

procedure depicted in ASTM D-4057 [1], with an interval between samples shown in Figure 1.

In order to evaluate the performance of the oil formulation and its condition along the ODI, a broad range of parameters were monitored. They are presented in Table 3 along with the technique, device employed and the standards that regularize their measurement procedure. These parameters have also been classified according to their purpose, that is, oil degradation and wear. Oxidation of the oil, aminic and antiwear additives were measured by FT-IR spectroscopy following an "in-house" methodology [17], based on ASTM D-7214. For the analysis of engine wear, inductively coupled plasma - optical emission spectrometry (ICP-OES) technique was employed, allowing to monitor the presence of wear metals and also those from the additive package in the oil formulation.

3. Results and discussion

3.1. Fuel consumption

Results of fuel consumption reduction due to the use of using LVEOs are presented here by bus model and for the four stages of the test. Results of stages I and II have been presented elsewhere in reference [15], where it was found that formulations with lower HTHS viscosity had significant benefits on fuel economy for the Diesel I, Diesel II and CNG buses, and that this potential is closely related with the engine's thermal load. Results of stage III, on the other hand, have been previously addressed in [32]; here LVEOs continued to prove their fuel economy potential, however, it was also concluded that for Diesel II buses, of higher thermal load, the use of formulations with HTHS lower than 3.5 cP, leads to the increase of fuel consumption.

In the bar plots presented in this section, bars with diagonal cross pattern represent the reference oil, while solid pattern is used for candidate formulations. Furthermore, fuel consumption difference between the reference and candidate oil is presented as percentage difference, therefore values with negative sign represent fuel saving, while positive sign means fuel consumption increase.

3.1.1. Diesel I buses

Six different oil formulations were tested in this bus model, being candidate oils E and H the ones with the lowest HTHS viscosity, 3.05 and 3.10 cP, respectively. Average fuel consumption at the end of each stage of the test is illustrated in Figure 2 along with the deviation of the measurements. Results of the ANOVA analysis are summarized in Table 4 with absolute and percentage differences in fuel consumption. Fuel savings were achieved for all the stages using a candidate oil of lower HTHS viscosity, although it is larger during the fourth stage. If HTHS viscosity values are compared between reference and candidate oil in the same stage, it can be seen that the smaller difference occurs in stage IV, suggesting that another parameter is helping to reduce fuel consumption, which would be the additives of the API FA-4 category.

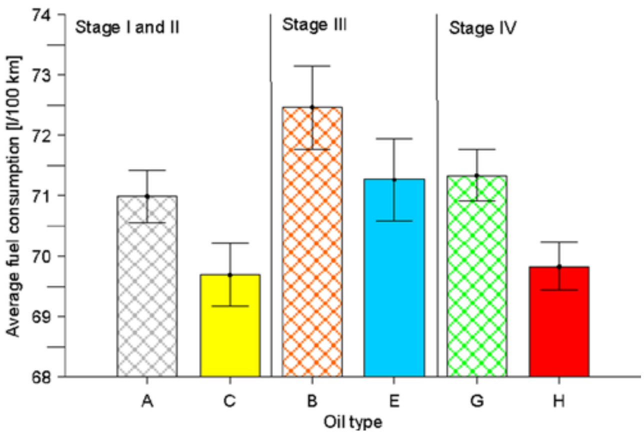


Fig. 2. Average fuel consumption and error bars for Diesel I buses along the test

Table 4. Fuel consumption for Diesel I buses

Reference - candidate oil	Absolute difference [l/100km]	Percentage difference [%]
A - C	1.30	-1.83
B - E	1.20	-1.65
G - H	1.51	-2.11

3.1.2. Diesel II buses

Average fuel consumption with the five oil formulations tested with this bus model are illustrated in Figure 3 and absolute and percentage differences in fuel consumption are in Table 5. For stages I and II candidate oil C gave non statistically significant fuel savings, less than 1%; therefore, and having analyzed the results of oil performance and degradation, it was decided to employ a formulation with a lower HTHS viscosity for stage III. Results with candidate oil E, which were compared with reference oil B of stages I and II, however, showed to greatly increase the fuel consumed by the buses in almost 6%. This result could be explained from the significant reduction of HTHS viscosity and the thermal load of this type of engine (Table 1); as the working conditions of a public service bus consist mostly of high loads and low engine speed, the appearance of mixed and boundary lubrication is likely to occur.

For stage IV on the other hand, it was decided to maintain HTHS viscosity in about 3.5 cP for the test oils D and G, the last one belonging to the new API CK-4 category. Results of fuel consumption gave an increase of 0.27% but its significance could not be proved statistically. Given that there is no difference in HTHS viscosity and the similarities in the rest of the oils' properties (see Table 2), this small fuel increase could be attributed to deviations in the measurements.

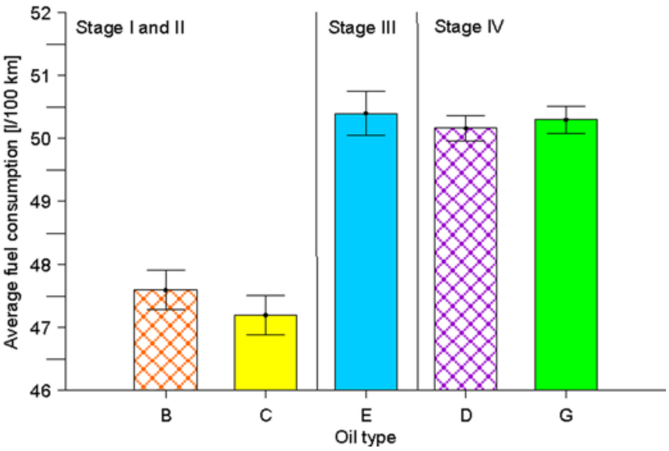


Fig. 3. Average fuel consumption and error bars for Diesel II buses along the test

Table 5. Fuel consumption differences for Diesel II buses

Reference - candidate oil	Absolute difference [l/100km]	Percentage difference [%]
B - C B - E D - G	0.43 2.84 0.14	-0.90 N.S 5.97 0.27 N.S

3.1.3. Diesel III buses

Diesel III buses were included in the test for stage III and IV. The average fuel consumption of the four oils tested in these buses is shown in Figure 4 and the differences in fuel consumption at the end of each stage are summarized in Table 6. As for Diesel I buses, fuel consumption savings gave greater results in stage IV than in stage III, even though the difference in HTHS viscosity values between reference and candidate oil are smaller. Results of both stages demonstrate that for this bus model it is possible to continue lowering the HTHS viscosity below 3 cP, however, it is also important to highlight the contribution of the additives of the API FA-4 category to fuel economy.

3.1.4. CNG buses

Average fuel consumption of the six oil formulations tested in CNG buses at the end of each stage are shown in Figure 5, while comparisons of fuel consumption reduction between reference and candidate oil are summarized in Table 7. For stages I, II and III, candidate oils gave the greatest fuel savings among all bus models, demonstrating the potential of LVEOs, especially when HTHS viscosity is reduced to 3.05 cP (oil F); a maximum fuel consumption reduction of 4.5% was achieved with this formulation. Furthermore it can also be seen, in stage I, that a small difference in HTHS viscosity can even give positive results for fuel economy. In stage IV, the percentage of fuel economy was reduced, although it is in the range of results obtained with Diesel I and III when comparing candidate oils G and H.

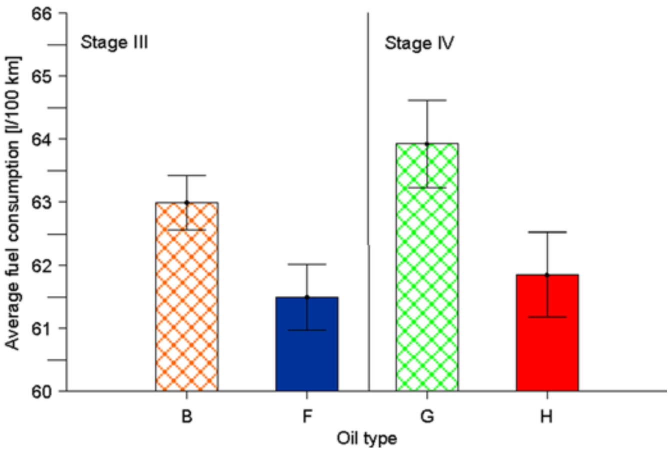


Fig. 4. Average fuel consumption and error bars for Diesel III buses along the test

Table 6. Fuel consumption differences for Diesel III buses

Reference - candidate oil	Absolute difference [l/100km]	Percentage difference [%]
B - F G - H	1.48 2.07	-2.35 -3.24

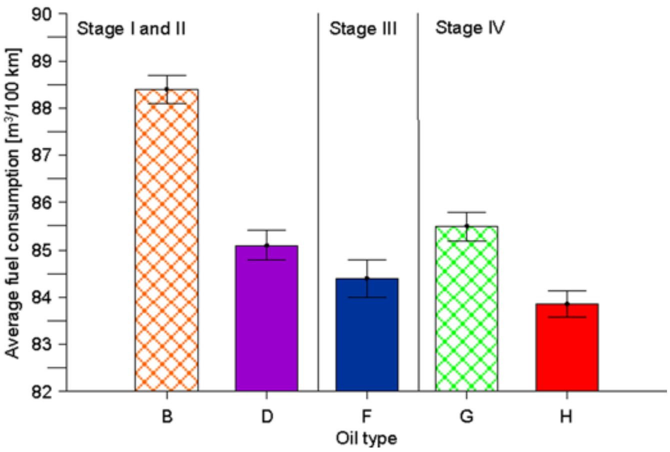


Fig. 5. Average fuel consumption and error bars for CNG buses along the test

Table 7. Fuel consumption differences for CNG buses

Reference - candidate oil	Absolute difference [l/100km]	Percentage difference [%]
B - D B - F G - H	3.30 4.00 1.64	-3.73 -4.52 -1.92

3.1.5. Fuel economy and thermal load

This section has been aimed to show the relation between the thermal load of the engine and the fuel economy potential of the different engine oil formulations. For this analysis, it is important to bear in mind that it is not possible to make a direct comparison between oils and engine technologies of different stages of the test, due to variables that cannot be controlled, such as ambient temperature and load (number of passengers), that may vary along the stages.

In Figure 6 it has been plotted the thermal load of the buses' engines (Table 1), defined as the maximum effective power over the piston area, against the fuel economy in percentage given by the candidate oil formulations. Here it can be seen that these parameters are strongly linked; for greater values of thermal load, the fuel economy potential of LVEOs tends to decrease, and even transform into fuel increase.

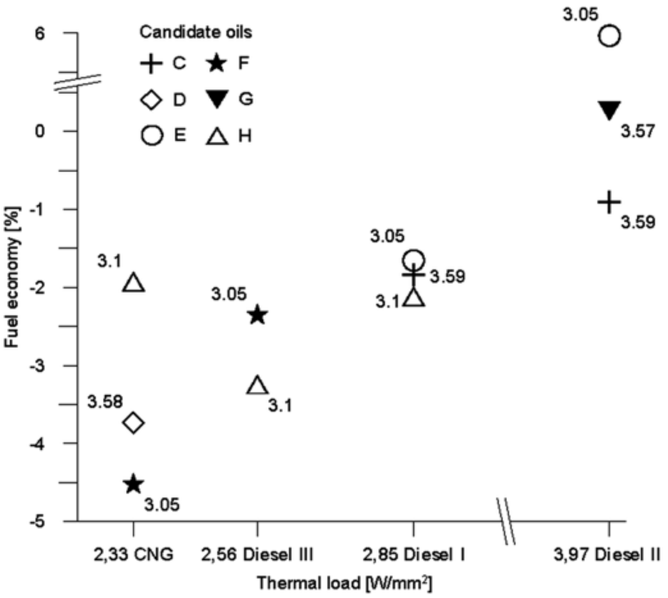


Fig. 6. Thermal load and fuel economy relationship

3.2. Oil performance and degradation

In this section are presented the most important results of oil performance, degradation and engine wear, obtained along the four stages of the test. Results of the parameters summarized in Table 3 are illustrated by engine model. Results of stages I and II have been previously presented in [14], and those of stage III in [33].

3.2.1. Kinematic viscosity

Kinematic viscosity was measured at 100ºC. Variations in this parameter have two main causes, one related with oxidation of the base oil [28], which in turn depends on the thermal load of the engine, and the other with the viscosity index improver (VII) added to the formulation. The effect of VII polymers consists in increasing the oil viscosity at elevated temperatures while keeping low resistance to flow in cold; however, this effect decreases with oil aging and is also affected by the high shear conditions presented in the normal engine operation [34, 37]. Results of kinematic viscosity at 100ºC are shown in Figure 7.

For Diesel I buses, oils A, B and C present a decrease in the kinematic viscosity along the ODI, suggesting shearing of the VII as the predominant factor of viscosity variation. On the other hand, oils G and H of the fourth stage and oil E of the third one, present an increase in the kinematic viscosity, as a result of its oxidation rates (see Figure 8). For oil E and H, a combination of the effect given by the VII shearing and base oil oxidation, resulted in a slighter increase of viscosity than oil G, which could have lower VII shearing.

For Diesel II buses and oils B and C, kinematic viscosity has a slight decrease along the ODI, as a result of the VII shearing that reduces the viscosity, accompanied by the opposite effect given by the oil oxidation. For the oil E the oxidation rate shown in Figure 8, was higher than C and its effect can be seen in the slighter variation of its kinematic viscosity.

For Diesel III and CNG buses, Figure 7 shows the prevalence of base oil oxidation for all the oil formulations. Stands out the results of oil F in Diesel III buses, its oxidations change along the ODI is lower than the other formulations, but the increase of its viscosity is higher. This situation suggests the presence of contaminants in the oil, such as soot.

In this study, the oxidation effect over the increase of kinematic viscosity was analyzed, however, there are another factors that also have an effect on this performance, such as external contaminants and combustion by-products.

3.2.2. Oxidation and aminic additives

Lubricant oils can oxidize when they are in contact with oxidizing atmospheres, as a consequence of blow-by by instance [9], and especially at elevated temperatures causing the oil molecules to break, rearrange and react [12]. This reaction causes oil thickening and thus loss of fluidity [2]; which in turn has a strong effect on the life of the oil. Furthermore, oxidation of the oil is affected by the presence of metals, such as iron and copper, which can be in contact with the oil if there are engine components with these metals, and in the form of wear debris [28]. Figure 8 shows the results of the oils' oxidation, by engine model, as a percentage change taking the fresh oil measurement as reference. As expected, this oil parameter increased along the ODI for all the oil formulations, and especially for oil B with a

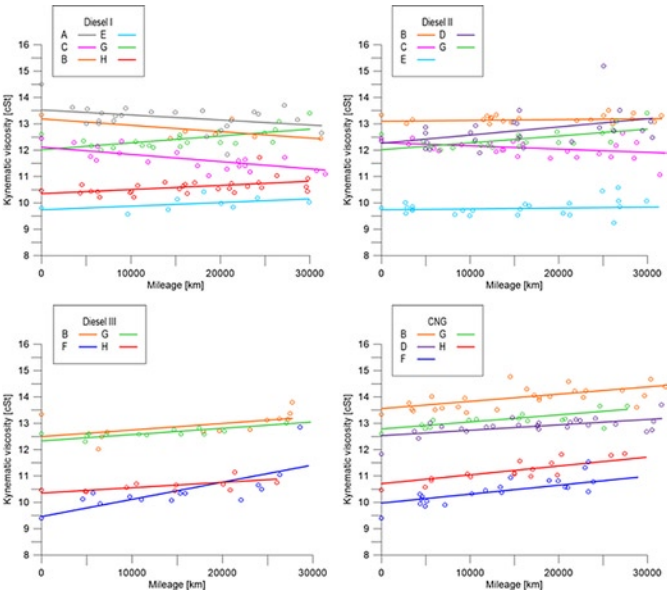


Fig. 7. Kinematic viscosity by engine model at 100ºC

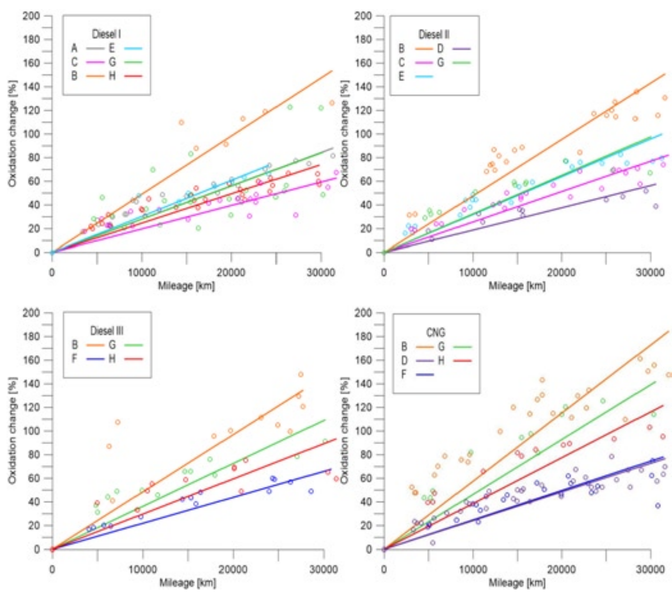


Fig. 8. Oxidation percentage change by engine model

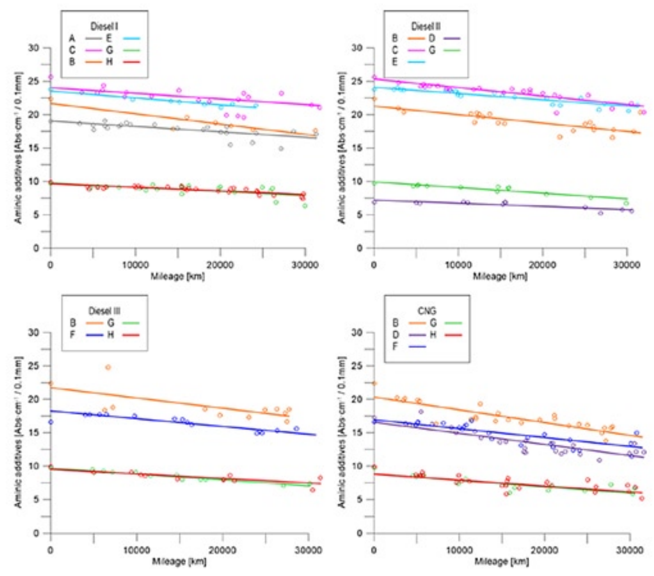


Fig. 9. Aminic additives by engine model

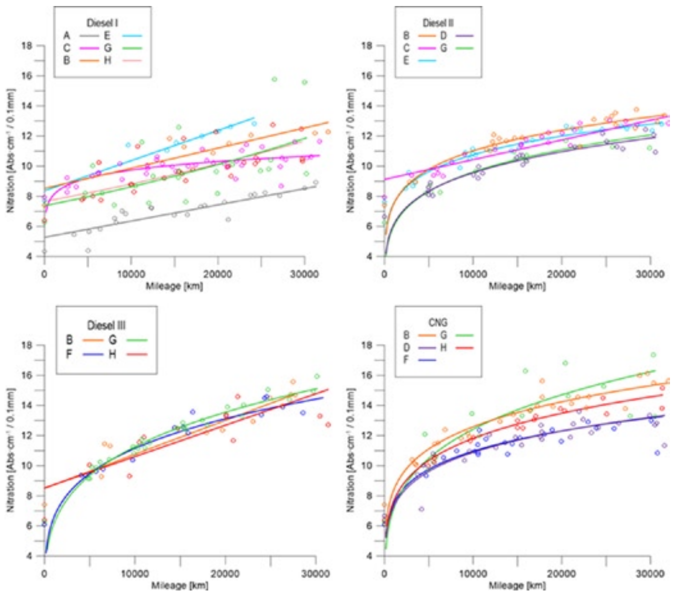


Fig. 10. Nitration by engine model

steeper slope in all the bus models. It can also be seen that the rate of oxidation was higher for the CNG buses because of the higher average combustion temperatures that they experience. For the Diesel III buses, the oxidation levels are also significant, possibly because of the use of EGR, which recirculates exhaust gases at elevated temperatures [13]. Oils G and H (API CK-4 and FA-4 oils) of the last stage present significant oxidation for all the bus models, especially if it is compared to previous formulations of lower oxidation resistance. This situation could be a consequence of the smaller quantity of aminic additive present in these formulations (shown in Figure 9), and the variation of the antiwear additive, as explained in Section 3.2.4.

In the following Figure 9, they are illustrated the results of aminic additives present in the oil formulations. Engine oils usually contain this additive along with ZDDP (zinc dithiophosphate) in order to delay the oxidation of the oil; its depletion along the ODI is clearly reflected in the oxidation rates of the oil. From these results, it can bee seen that the new API CK-4 and FA-4 oils contain smaller quantities of the additive, compared to the other formulations, but their depletion tendency is similar to some previous categories. Regarding oil B in Diesel I buses, the depletion of the aminic additive is more marked than the other formulations, which had an impact on its oxidation tendency.

3.2.3. Nitration

Nitration appears as a consequence of nitrogen dioxide (NOx) emissions from combustion reacting with the oil, it is closely related with oil oxidation in terms of its effects over the oil performance, that is, oil acidity, increase of viscosity and corrosive wear. This parameter was monitored by FT-IR spectrometer and results have been illustrated in Figure 10. Nitration along the ODI presented similar results to those of oxidation, although with smaller differences between the oil formulations. CNG and Diesel III buses presented the higher rates of nitration; for the former, the high temperatures reached during combustion, compared to diesel engines, promote the increase of NOx levels; while for the Diesel III engines, the use of EGR also introduces NOx compounds in the oil through the recirculated gases.

3.2.4. Antiwear additives

The antiwear additive used in the oil formulations, ZDDP, was monitored using a FT-IR spectrometer. The wavenumber range used for the measurements was between 1025 and 960 cm^{-1}, with two baseline points, one from 2200 to 1900 cm^{-1}, and the other from 650 to 550 cm^{-1}; results of antiwear additive are the measured area. Results of the content of this additive along the ODI are illustrated in Figure 11. Here, the depletion of the additive is marked for all the oil formulations; however for candidate oils of the fourth stage, G and H, the depletion of ZDDP is very significant; it can be seen that oils ran out of this additive before reaching the middle of the ODI for all the engine models, although it is more evident for CNG buses. Given that ZDDP is also used as anti-oxidant additive, the effect of its depletion can be observed on the oxidation rates of the oils, and in turn on the increase of TAN and wear of soft metals, such as copper and lead.

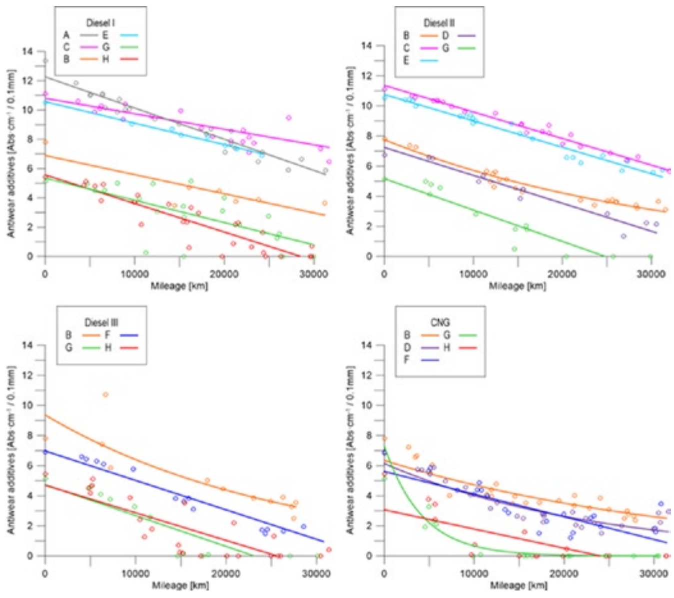


Fig. 11. Antiwear additives (ZDDP) by engine model

3.2.5. Total acid number (TAN) and total base number (TBN)

Results of TAN and TBN have been illustrated in Figure 12 and 13, respectively. It can be observed that the increase of acidic matter in the oil is significant for all the formulations and engine models. All the oil formulations presented marked decreases of TBN, and specifically for candidate oils of the fourth stage, G and H, TBN at the end of the ODI was about 50% lower than its initial value. In Figure 13, it can be noted that oils C and E have higher values of TBN from the

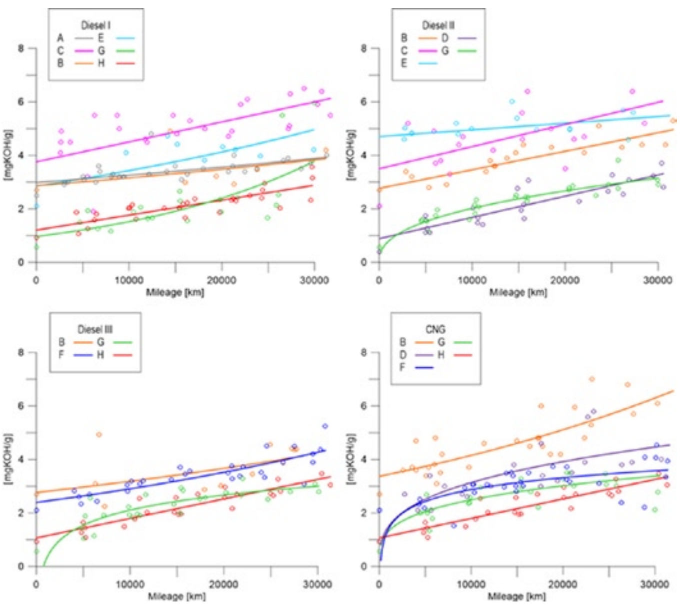


Fig. 12. TAN by engine model

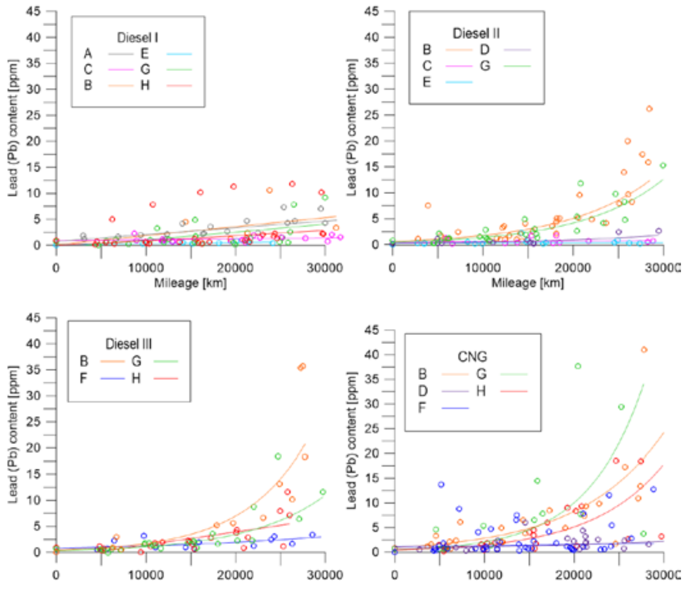


Fig. 14. Lead concentration by engine model

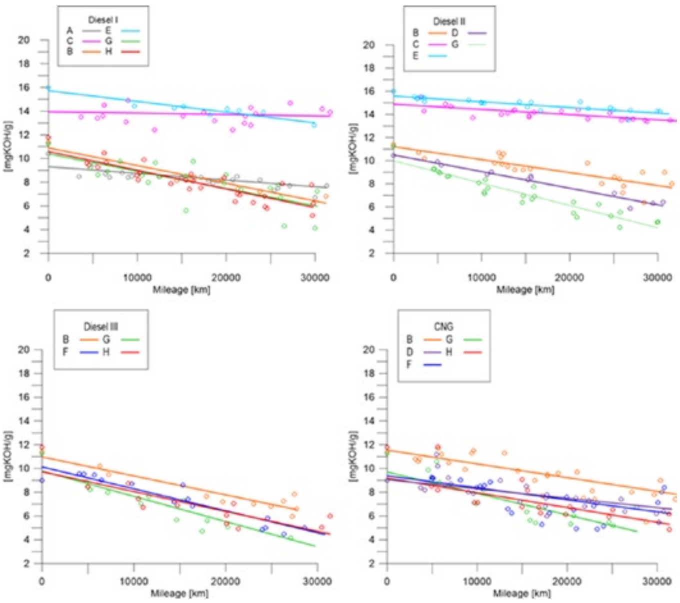


Fig. 13. TBN by engine model

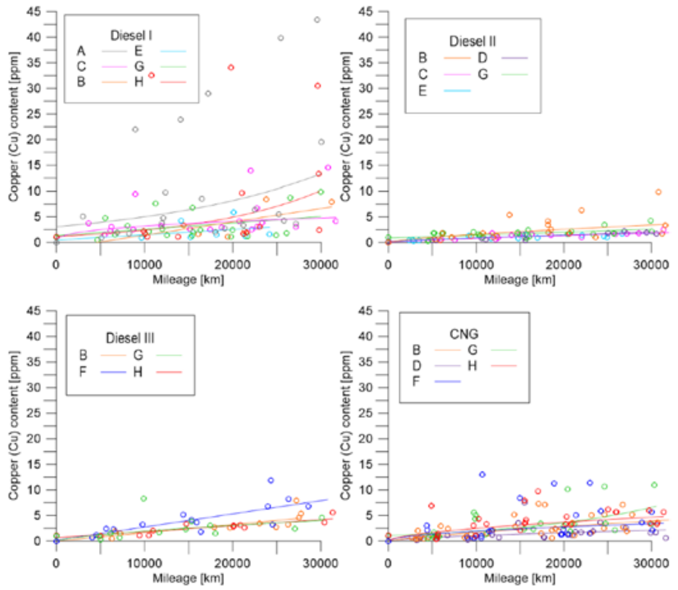


Fig. 15. Copper concentration by engine model

beginning of the test; this is due to the Ca-based detergent employed in these formulations (Table 2).

Results of this section, with high variations in the TAN measurements and the decrease of TBN, suggest the presence of corrosive wear [13] affecting lead (Pb) and copper (Cu), usually found in journal bearings and prone to corrosion [28]. The following Figures 14 and 15 show the concentrations of Pb and Cu, respectively, measured by ICP-OES spectrometer. Furthermore, in Table 8 and 9 are presented the mean and standard deviation (STD) of Pb and Cu concentrations, at the end of the ODI.

Regarding Pb concentration, its increase is significant for three of the engine models, Diesel II, Diesel III and CNG, and especially for oils B, G and H. Their increase can also be associated with the depletion of the antiwear additives, even before reaching 15.000 km, in some cases. Overall, the Cu content does not present abnormal results and the increasing rates are very similar between oil formulations. For Diesel I buses, oils A and H present some peak points however, given

Table 8. Mean and standard deviation of the Pb concentration at the end of the ODI

Oil formulation	Mean Pb concentration ± STD at 30.000 km[ppm]			
	Diesel I	Diesel II	CNG	Diesel III
A	4,3 ± 2,0			
B	3,4 ± 0,0	21,0 ± 13,5	21,8 ± 11,1	24,9 ± 11,0
C	1,3 ± 0,6	0,9 ± 0,5		
D		3,0 ± 1,7	3,0 ± 3,0	
E	0,6 ± 0,2	0,4 ± 0,2		
F			8,6 ± 10,5	3,9 ± 1,4
G	3,6 ± 3,6	8,7 ± 4,0	27,9 ±17,2	14,4 ± 7,9
H	6,6 ± 4,4		16,9 ± 10,8	7,1 ± 5,3

Table 9. Mean and standard deviation of the Cu concentration at the end of the ODI

	Mean Cu concentration ± STD at 30.000 km [ppm]			
Oil formulation	Diesel I	Diesel II	CNG	Diesel III
A	17,6 ± 14,0			
B	8,2 ± 0,3	4,4 ± 2,0	5,2 ± 3,6	5,3 ± 1,7
C	6,2 ± 4,6	2,0 ± 0,6		
D		2,0 ± 0,5	2,1 ± 1,3	
E	2,9 ± 2,1	1,9 ± 0,5		
F			4,5 ± 2,7	8,0 ± 2,6
G	5,5 ± 3,9	2,5 ± 1,0	6,4 ± 3,8	3,5 ± 0,9
H	20,3 ± 13,1		5,0 ± 1,1	4,0 ± 0,9

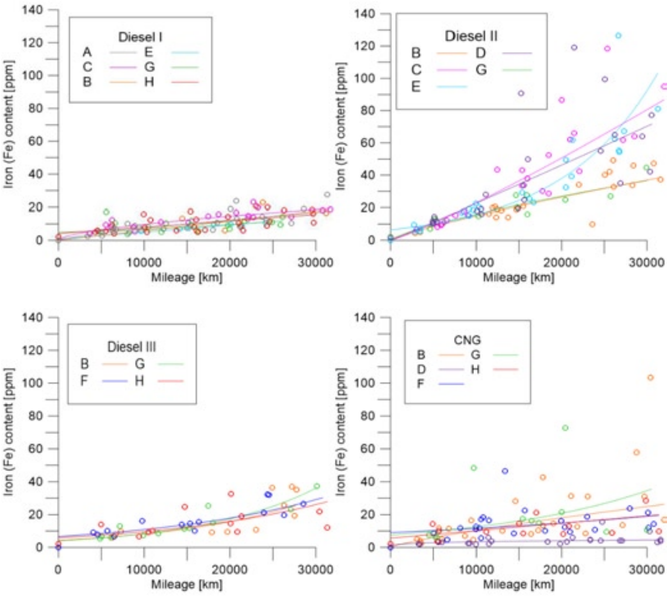


Fig. 16. Iron concentration by engine model

that the other wear metals, Pb and iron (Fe) (depicted in the following Section 3.2.6.), do not present high concentration values, its presence in the engine oil could be due to causes other than wear, such as external contamination.

3.2.6. Engine wear

To evaluate wear of the engine, the concentration of Fe in the oil was monitored by ICP-OES, and it can be seen in Figure 16 by engine model throughout the ODI. In the Table 10 are shown the results of Fe content in terms of wear ratio [ppm/10000 km] evaluated at the end of the ODI for all the bus models and their corresponding oil formulations. It stands out the significant increase of Fe content for the Diesel II buses and all the oil formulations, compared to the other bus models. This situation in the Diesel II buses arises from the combination of two main factors, their high thermomechanical stress, and the configuration of the valve train, which consists of steel OHV (over head valve) cam follower, leading to the increase of Fe debris in the oil. Overall, oil formulations of the fourth stage, G and H, showed to have

Table 10. Wear ratio evaluated at the end of the ODI

	Wear ratio [Fe ppm/10000 km]			
Oil formulation	Diesel I	Diesel II	CNG	Diesel III
A	5.84	-	-	-
B	5.61	15.13	5.05	10.41
C	6.28	29.52	-	-
D	-	26.83	3.92	-
E	4.58	26.93	-	-
F	-	-	6.98	10.25
G	3.70	12.77	11.84	9.80
H	5.21	-	6.40	8.13

a better performance in terms of Fe content than lubricants of previous stages, possibly due to the higher quality of the formulation.

4. Conclusions

a) The use of LVEOs in HDVs continues to be a proven alternative to reduce fuel consumption and therefore the carbon footprint of internal combustion engines. Four engine technologies were involved in the test where eight different oil formulations were evaluated. Results showed fuel consumption reduction for three of the four buses' models, demonstrating that the potential of LVEOs is closely linked to the mechanical and thermal stress of the engine. For the Diesel II buses, by instance, it is clear that the optimum HTHS viscosity value has a limit in about 3.5 cP, a lower viscosity results in a significant fuel consumption increase of about 6%. The use of the new formulations G and H that belong to the latest API CK-4 and FA-4 categories gave greater values of fuel economy for the Diesel I and Diesel III vehicles, than in the previous stages of the test. This could be a consequence of lower HTHS viscosity, for oil H, and the additives used in the new API categories.

b) Overall, from the oil analysis, it can be observed that the performance of LVEOs was as expected with no significant effects on engine wear. Regarding oil degradation, formulations with lower HTHS viscosity presented higher variations in measurements of TBN and TAN and in kinematic viscosity, which can be attributed to the increased demands placed on the lubricating oil due to lower oil film thickness.

c) For oils G and H, it is important to highlight the significant depletion of the antiwear additives, even before reaching 15.000 km. Given that the SAPS level of the formulations is low and the fuel used by the vehicles has low content in sulfur, the marked increase of TAN can be associated with other factors, such as oxidation. These previously mentioned conditions can lead to limit any possible extension of the ODI. Results of Fe content in these oils, however, showed to be lower than most of the previous formulations, possibly due to the additives and higher quality of these new formulations.

d) Results presented here, obtained from a public service HDVs fleet, show the importance of a comprehensive analysis of the oil formulations used in the vehicles, as it gives valuable information to make well-informed decisions on the maintenance program of the vehicles, reduce costs, both of maintenance and operation, and to reduce downtimes and repairs.

Acknowledgement

Author Sophia Bastidas would like to thank the support of the program Ayudas de Investigación y Desarrollo (PAID-01-17) of the Universitat Politécnica de Valéncia.

References

1. ASTM D4057-12 Standard Practice for Manual Sampling of Petroleum and Petroleum Products. West Conshohocken, Pennsylvania, 2018.
2. Bloch H P, Bannister K. Practical lubrication for industrial facilities. Lilburn: Fairmont Press, 2009.
3. Trucking into a greener future: the economic impact of decarbonizing goods vehicles in Europe, 2018. Downloadable at https://europeanclimate.org/trucking-into-a-greener-future/.
4. UNE-EN 590:2014+A1:2017 Combustibles para automoción. Combustibles para motor diesel (gasóleo). Requisitos y métodos de ensayo. Génova, España, 2017.
5. Dörr N, Agocs A, Besser C, Ristic A, Frauscher M. Engine oils in the field: A comprehensive chemical assessment of engine oil degradation in a passenger car. Tribology Letters 2019; 67(3):68, https://doi.org/10.1007/s11249-019-1182-7.
6. Reducing CO2 emissions from heavy-duty vehicles, 2019. Available at https://ec.europa.eu/clima/policies/transport/vehicles/heavy_en
7. Commission Directive 2001/27/EC, 2001. Downloadable at https://op.europa.eu/en/publication-detail/-/publication/6ec7ec8b-4cf0-4d88-94ee-58b03bd1a722/language-en
8. Golebiowski W, Wolak A, Zajac G. The influence of the presence of a diesel particulate filter (DPF) on the physical and chemical properties as well as the degree of concentration of trace elements in used engine oils. Petroleum Science and Technology 2019; 37(7):746–755, https://doi.org/10.1080/10916466.2018.1539751.
9. Hsu S M, Ku C S, Pei P T. Oxidative degradation mechanisms of lubricants. Aspects of Lubricant Oxidation. ASTM International 1986; 27-48, https://doi.org/10.1520/STP18372S.
10. James C J. Analysis of parasitic losses in heavy duty diesel engines. PhD thesis, Massachusetts Institute of Technology, 2012.
11. Jang J, Lee Y, Kwon O, Lee M, Kim J. The effect of engine oil on particulate matter, emissions and fuel economy in gasoline and diesel vehicle. SAE Technical Paper 2014-01-2837, 2014, https://doi.org/10.4271/2014-01-2837.
12. Kumbár V, Votava J. Differences in engine oil degradation in spark-ignition and compression-ignition engine. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2014; 16(4):622-628.
13. Lakshminarayanan P, Nayak N S. Critical component wear in heavy duty engines. Singapore: John Wiley & Sons, 2011.
14. Macián V, Tormos B, Miró G, Pérez T. Assessment of low-viscosity oil performance and degradation in a heavy duty engine real-world fleet test. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology 2016; 230(6):729–743, https://doi.org/10.1177/1350650115619612.
15. Macián V, Tormos B, Ruíz S, Ramírez L. Potential of low viscosity oils to reduce CO2 emissions and fuel consumption of urban buses fleets. Transportation Research Part D: Transport and Environment 2015; 39:76–88, https://doi.org/10.1016/j.trd.2015.06.006.
16. Macián V, Tormos B, Bermúdez V, Ramírez L. Assessment of the effect of low viscosity oils usage on a light duty diesel engine fuel consumption in stationary and transient conditions. Tribology International 2014; 9:132–139, https://doi.org/10.1016/j.triboint.2014.06.003.
17. Macián V, Tormos B, Gómez Y A, Salavert J M. Proposal of an FTIR methodology to monitor oxidation level in used engine oils: Effects of thermal degradation and fuel dilution. Tribology Transactions 2012; 55(6):872–882, https://doi.org/10.1080/10402004.2012.721921
18. Muncrief R, Rodriguez F. Briefing: A roadmap for heavy-duty engine CO2 standards within the European Union framework. The International Council on Clean Transportation, 2017. Downloadable at https://theicct.org/publications/roadmap-heavy-duty-engine-co2-standards-within-european-union-framework
19. Norris J, Escher G. Heavy duty vehicles technology potential and cost study. The International Council on Clean Transportation, 2017. Downloadable at https://theicct.org/publications/heavy-duty-vehicles-technology-potential-and-cost-study.
20. Porter R. New API certified CK-4 and FA-4 diesel engine oils are available beginning December 1, 2016. Available at https://www.api.org/news-policy-and-issues/news/2016/11/18/new-api-certified-diesel-engine-oils-are.
21. Priest M, Taylor C. Automobile engine tribology-approaching the surface. Wear 2000; 241(2):193–203, https://doi.org/10.1016/S0043-1648(00)00375-6.
22. Rahnejat H. Tribology and dynamics of engine and powertrain: fundamentals, applications and future trends. Cambridge: Elsevier, 2010.
23. Raposo H, Farinha J T, 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.
24. Richardson D E. Review of power cylinder friction for diesel engines. Journal of engineering for gas turbines and power 2000; 122(4):506–519, https://doi.org/10.1115/1.1290592.
25. SAE J300 Engine Oil Viscosity Classification. United States, 2015.
26. Sander D E, Allmaier H, Knauder C, Strömstedt F. Potentials and risks of reducing friction with future ultra-low-viscosity engine oils. MTZ worldwide 2018; 79(12):20–27, https://doi.org/10.1007/s38313-018-0124-3.
27. Singh S K, Singh S, Sehgal A K. Impact of low viscosity engine oil on performance, fuel economy and emissions of light duty diesel engine. SAE Technical Paper 2016-01-2316, 2016, https://doi.org/10.4271/2016-01-2316.
28. Stachowiak G W, Batchelor A W. Engineering Tribology, Fourth edition. Boston: Butterworth-Heinemann, 2014.
29. Taraza, D, Henein N, Bryzik W. Friction losses in multi-cylinder diesel engines. SAE Technical Paper 2000-01-0921, 2000, https://doi.org/10.4271/2000-01-0921.
30. Taylor C M. Engine tribology. Amsterdam: Elsevier, 1993.
31. Teter J. Trucks & buses (heavy-duty vehicles) Tracking Clean Energy Progress, 2019. Available at https://www.iea.org/tcep/transport/trucks/.
32. Tormos B, Ramírez L, Miró G, Pérez T. Real world fleet test to determine the impact of low viscosity engine oils from Heavy-Duty CNG and diesel buses - Part I: Fuel consumption. SAE Technical Paper 2017-01-2353, 2017, https://doi.org/10.4271/2017-01-2353.
33. Tormos, B., Ramírez, L., Miró, G., and Pérez, T. Real world fleet test to determine the impact of lower viscosity engine oils from Heavy-Duty CNG and diesel buses - Part II: Oil performance. SAE Technical Paper 2017-01-2351, 2017, https://doi.org/10.4271/2017-01-2351.
34. Totten G. Handbook of lubrication and tribology: Volume I application and maintenance, Second edition. Boca Raton: Taylor and Francis, 2006.

35. Van Dam W, Kleijwegt P, Torreman M, Parsons G. The lubricant contribution to improved fuel economy in heavy duty diesel engines. SAE Technical Paper 2009-01-2856, 2009, https://doi.org/10.4271/2009-01-2856.
36. Van Dam W, Miller T, Parsons G, Takeuchi Y. The impact of lubricant viscosity and additive chemistry on fuel economy in heavy duty diesel engines. SAE International Journal of Fuels and Lubricants 2012; 5(1):459-469, https://doi.org/10.4271/2011-01-2124.
37. Wolak A, Zajac, G. The kinetics of changes in kinematic viscosity of engine oils under similar operating conditions. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19(2):260-267, http://dx.doi.org/10.17531/ein.2017.2.14.
38. Wong V W, Tung S C. Overview of automotive engine friction and reduction trends–effects of surface, material, and lubricant-additive technologies. Friction 2016; 4(1):1–28, https://doi.org/10.1007/s40544-016-0107-9.

Vicente MACIÁN
Bernardo TORMOS
Sophia BASTIDAS
CMT-Motores Térmicos
Unversitat Politècnia de València
Camino de Vera S/N, 46022. Valencia, Spain

Tomás PÉREZ
REPSOL
Calle Agustín de Betancourt S/N, 28935. Móstoles, Spain

E-mails: vmacian@mot.upv.es, betormos@mot.upv.es, kabasmon@mot.upv.es, tperezg@repsol.com

Article citation info:
CHUDZIK A, WARDA B. Fatigue life prediction of a radial cylindrical roller bearing subjected to a combined load using FEM. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 212–220, http://dx.doi.org/10.17531/ein.2020.2.4.

Agnieszka CHUDZIK
Bogdan WARDA

FATIGUE LIFE PREDICTION OF A RADIAL CYLINDRICAL ROLLER BEARING SUBJECTED TO A COMBINED LOAD USING FEM

PROGNOZOWANIE TRWAŁOŚCI ZMĘCZENIOWEJ PROMIENIOWEGO ŁOŻYSKA WALCOWEGO PODDANEGO ZŁOŻONEMU OBCIĄŻENIU Z WYKORZYSTANIEM MES

The article presents the results of studies on the impact of a combined load of a radial cylindrical roller bearing for its predicted fatigue life. The distributions of maximum equivalent subsurface stresses and their depths, necessary during calculations of fatigue life, were determined using the finite element method, using the basic package of the ANSYS program. The calculations took into account the geometrical parameters of the bearing, including radial clearance and the shape of the rolling elements generators. The calculation results showed that the axial load of the radial cylindrical roller bearing and the tilt of the rollers associated with its operation reduces fatigue life. The obtained results were compared with the results of calculations according to the SKF catalogue method, obtaining satisfactory compliance.

Keywords: *rolling bearing; stress distribution; fatigue life; finite element method; combined load.*

W artykule zaprezentowano wyniki badań wpływu złożonego obciążenia promieniowego łożyska walcowego na jego prognozowaną trwałość zmęczeniową. Rozkłady maksymalnych zastępczych naprężeń podpowierzchniowych oraz głębokości ich występowania, niezbędne podczas obliczeń trwałości zmęczeniowej, określono za pomocą metody elementów skończonych, z wykorzystaniem pakietu podstawowego programu ANSYS. W obliczeniach uwzględniono geometryczne parametry łożyska, w tym luz promieniowy i kształt tworzących elementów tocznych. Wyniki obliczeń wykazały, że obciążenie osiowe promieniowego łożyska walcowego i przechylenie wałeczków towarzyszące jego działaniu powoduje zmniejszenie trwałości zmęczeniowej. Otrzymane wyniki porównano z wynikami obliczeń według katalogowej metody firmy SKF, otrzymując zadowalającą zgodność.

Słowa kluczowe: *łożysko walcowe, rozkład naprężeń, trwałość zmęczeniowa, metoda elementów skończonych, złożone obciążenie.*

Nomenclature

A	material constant
B	bearing width
C	dynamic load rating
D	bearing outside diameter
D_r	roller diameter
E	Young's modulus
F_r	radial load of the bearing
F_a	axial load of the bearing
$F_{ap\ max}$	permissible axial load
L_{10}	fatigue life
L_r	roller length
Q	resultant normal force in the roller-main race contact
Q_a	axial force per roller
Q_f	resultant normal force in the roller end flange contact
Q_r	radial force per roller
Z	depth of occurrence of maximal von Mises stresses along the x axis
Z_r	number of rollers in the bearing
c	exponent in the equation determining the survival probability
d	bearing bore diameter
d_{bi}	diameter of the inner ring raceway
d_{bo}	diameter of the outer ring raceway
e	Weibull slope
g	radial clearance in the bearing
h	exponent in the equation determining the survival probability
i	inner raceway (index)
j	number of roller
l	length of the roller-main race contact area
o	outer raceway (index)
r_b	radius of the main race
r_c	roller chamfer dimension
u	number of load cycles per one revolution
η	roller tilt angle
θ	roller skew angle
σ	maximal von Mises stress occurring along the x axis
ϕ	survival probability of the bearing element
ψ	angle measured along the bearing circumference
ψ_j	angle between rollers
ψ_{lim}	angle of the loaded zone

1. Introduction

In many cases, there is a need to precisely determine the fatigue life of bearing arrangements. Forecasting methods were presented in the catalogue of rolling bearings but they are not taking into account many factors affecting it. One of them is misalignment of the bearing's rings. In cylindrical rolling bearings it results in tilting rollers relative to the raceway and uneven pressure distribution in the contacts of the rollers with the raceways. Similar phenomenon can be observed in cylindrical roller bearings, which next to the radial load can carry a continuous, low axial load. These bearings include design variants NJ, NUP and NU with angular ring HJ. Occurring in the contacts of the rollers with the main raceways and auxiliary raceways (flanges), normal forces Q, Q_f and friction force T_f causes bearing tilt in two respectively perpendicular planes passing through the bearing axis. Bearing tilt in plane parallel to the bearing axis (angle θ), caused by friction force in roller contacts with flanges (Fig. 1), causes rollers precipitation from the right track and as a result occurrence of slips in roller contact with main raceways. Because of small values of θ angles, influence of phenomena associated with skewing on its fatigue life can be skipped [15]. Roller tilt in the plane passing through roller axis (angle η) causes occurrence of uneven pressure distribution in roller contacts with main raceways characterised by increased load on one end of the contact area. This phenomenon significantly reduces bearing fatigue life.

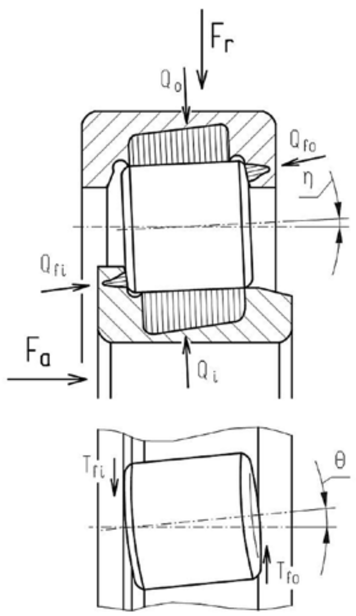


Fig. 1. Tilting η and skewing θ of the roller in the radial cylindrical roller bearing loaded with radial force F_r and axial force F_a

Until recently it was impossible to determine fatigue life of a NJ cylindrical roller bearing loaded with radial and axial force. Currently SKF company bearing catalogue [23] provides formulae allowing for approximate the fatigue life of such a loaded bearing.

According to the information in the catalogue [23] cylindrical single row bearings and cylindrical single row bearings with full number of rollers with flanges on the inner and outer ring can carry an axial load of value up to $F_a = 0.5\,F_r$. Because of mechanical limitations permissible axial load of radial cylindrical roller bearings of the dimension series 2 cannot be greater than $F_{ap\,max} = 0.0045\,D^{1,5}$. Basic nominal durability of the rolling bearing is determined from the formula:

$$L_{10} = \left(\frac{C}{P}\right)^{10/3}, \qquad (1)$$

where equivalent dynamic load of the NJ series 2 radial cylindrical roller bearing equals:

$$P = F_r \qquad \text{for}: F_a / F_r \le 0.2 \qquad (2)$$

$$P = 0.92\,F_r + 0.6\,F_a \qquad \text{for}: F_a / F_r > 0.2 \qquad (3)$$

Precise determination of cylindrical roller bearing fatigue life subjected to a combined load is only possible using computer technology. Forecasting method of the fatigue life requires consideration of local values of subsurface stress which affects fatigue life. It is necessary to determine the radial and axial load distribution on bearing individual elements beforehand. The most convenient way to complete this task is to use finite elements method. Properly built FEM solid model, replicating internal geometry of a bearing, including correction of roller generators but also taking into account bearing clearances and mutual tilting of mating elements of the bearing caused by complex loading and misalignment of the rings, would give one the most accurate information about both load distributions as well as pressure distributions in the contacts and corresponding subsurface stress distributions.

FEM calculations require a lot of computing power and are time consuming. This is the reason why FEM models were mostly used with ball bearings [11, 22]. In studies on determining load distribution in roller bearings, the authors usually used two-dimensional FEM models [6, 16, 19, 32], and thus were unable to account for the axial force acting on the bearing, or used three-dimensional models [10, 17, 18], but they did not deal with bearings subjected to complex loading, which greatly simplified the calculation model. In many cases, to simplify the three-dimensional bearing model, rollers or balls were replaced with elements defined by the model creator, so-called super elements. Super elements were used in the studies written by Golbach [8], Kania [12] and Claesson [5]. The last researcher developed simplified FEM models for tapered roller bearings and cylindrical roller bearings, including radial cylindrical roller bearings capable of carrying a complex load. Author [5] stressed that the use of full models of solid bearings in the finite element method is unprofitable from the point of view of the cost of calculations.

Researchers often used numerical methods other than the finite element method to find load distribution on rolling bearing components. Zhenhuan et al.[33] in their work they determined load distribution on high speed cylindrical roller bearings using the quasi-dynamic method. Cheng et al., did similarly, analyzing the impact of axial load on fatigue life of a radial cylindrical roller bearing [3]. In their work, they used the methods used in the models described by Brandlein and Fernlund [2, 7]. The authors examined the general conditions of contact of rollers with bearing raceways by introducing the concept of three bearing load zones. In the first of them, the rollers are in contact with the raceways along the entire roller generator length, in the second one on the roller generator part, and in the third one they are not in contact with the raceways, and thus, they do not carry load. However, the researchers did not present the distribution of axial forces on the rollers, focusing on determining the impact of uneven pressure distribution in the contacts of the rollers with the main raceways on bearing fatigue life. Another study that the authors did not use the finite element method is an article by Tong et al. dealing, among others, about the impact of bearing rings misalignment on its fatigue life [26]. The authors built a quasi-static cylindrical bearing model with four degrees of freedom, and numerically solved the system of bearing equilibrium equations, and then, by solving the Boussinesq problem for elastic half-space, determined the pressure distributions at the contacts enabling the calculation of the predicted fatigue life.

The authors of this article have also used a method that does not use finite elements in their previous works. The radial and axial load distributions were determined by solving the system of equations of rings and rollers by numerical way. Resultant normal forces Q occurring in the contacts of the rollers with the main raceways and the location of their application points were determined on the basis of the Palmgren formula [20, 21], while to calculate the resultant forces Q_f in the contacts of the roller faces with flanges, a calculation model was used, in which the contact area was divided into a number of layers of a defined, small width, and the pressure distribution under each of the layers was determined as for a rigid punch with an infinite length pressed under given angle in elastic half-space [27, 28]. Knowledge of load distributions enabled the authors to find subsurface stress distributions in the contacts and to calculate the predicted bearing fatigue life. The authors used this method to analyse the impact on fatigue life of radial cylindrical roller bearing correction roller generators [28], misalignment of bearing rings [29] and radial clearance in the bearing [4].

In all cited studies, the authors of this article used the finite element method to determine the distribution of subsurface stress, regardless of the solution of the Boussinesq problem for elastic half-space. For this purpose, three-dimensional models of the bearing fragment were built, which included a rolling element (roller) and a segment of one of the bearing rings [28, 29] or parts of both rings [4]. Consideration of the FEM model for a bearing fragment instead of the full solid model is commonly used by numerous researchers in a situation where the main purpose of the analysis is to study phenomena occurring in a single contact of rolling elements [1, 9, 24, 25, 30, 31]. The author of the work [13], who studied the influence of fitting and initial deformation on the resistance to motion in an angular-contact ball bearing, did a similar way. The author described in the work the analytical model of the angular contact ball bearing, but due to its complexity he used the finite element method to solve the problem, analysing a fragment of the bearing.

This paper discusses the impact of a combined load on the fatigue life of a radial cylindrical roller bearing. The task was carried out using the method described in works [27, 28], but only the finite element method was used to determine the subsurface stress distribution determining fatigue life.

2. Method for determining fatigue life of a radial cylindrical roller bearing

Predicted fatigue life of radial cylindrical roller bearing subjected to a combined load was calculated using the method described in the papers [27, 28], and also used in the works [4, 29]. This method is based on the basic assumptions of the commonly used fatigue life forecasting model developed by Lundberg and Palmgren [20, 21].

Determination of fatigue life L_{10} of a complete bearing requires an independent calculation of the durability of the inner L_{10i} and outer L_{10o} ring:

$$L_{10} = \left(L_{10i}^{-e} + L_{10o}^{-e} \right)^{-1/e} . \quad (4)$$

Durability of the inner ring, which is usually a rotating ring relative to the load, is determined from the relationship determining the inverse logarithm of the probability of durability φ_i:

$$ln \frac{1}{\varphi_i} = A \cdot u_i^e \, L_{10i}^e \, 2\pi \left[\frac{1}{2\pi} \int_0^{2\pi} \left(\int_0^l r_{bix} \, \sigma_{ix\psi}^c \, Z_{ix\psi}^{1-h} \, dx \right)^{1/e} d\psi \right]^e . \quad (5)$$

The durability of the outer ring, usually stationary, is calculated from a similar equation:

$$ln \frac{1}{\varphi_o} = A \cdot u_o^e \, L_{10o}^e \int_0^{2\pi} \int_0^l r_{box} \, \sigma_{ox\psi}^c \, Z_{ox\psi}^{1-h} \, dx \, d\psi , \quad (6)$$

where $\varphi_i = \varphi_o = 0.9$.

In the above formulae L_{10} is the number of revolutions of the bearing, u the number of load cycles per revolution, $\sigma_{x\psi}$ is maximum subsurface stress determined in accordance with the Huber-Mises-Hencke hypothesis (or in other words, von Mises stress), and $Z_{x\psi}$ is depth at which these stresses occur (Fig. 2). For the bearing to be considered, the material constant $A = 4.5 \cdot 10^{-40}$. The values of the exponents found in both equations are: $c = 31/3$, $h = 7/3$, $e = 9/8$. The method of deriving the formulae (5, 6) is presented in study [27].

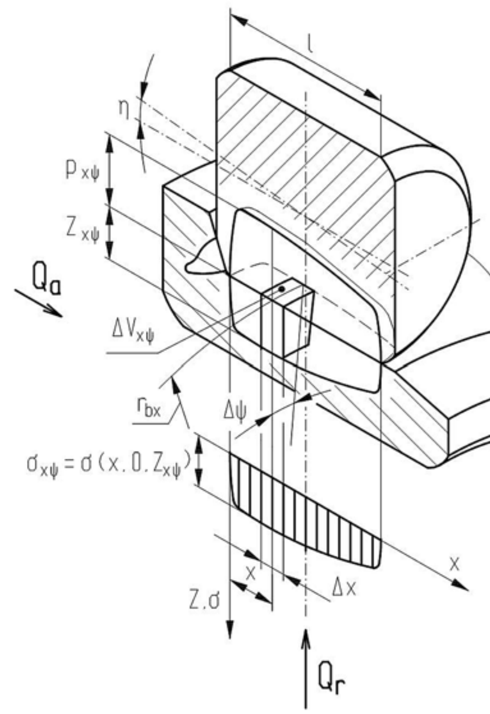


Fig. 2. Distributions of pressure and subsurface stresses deciding on fatigue life of radial cylindrical roller bearing subjected to combined load

Equations (5) and (6) were solved by numerical integration using a computer program ROLL2 [27]. Earlier, the distribution of maximum substitute subsurface stresses σ were found using the finite element method in the contact area between the rollers and raceways and the depth of their occurrence Z. To determine them, it was necessary to know the values of forces acting in contacts of subsequent rollers with the main raceways of the bearing as well as the tilt angles of the rollers $\eta = \eta_i = \eta_o$. They were determined using a computer program ROLL1, built based on the method described in the works [27, 28].

3. Determination of subsurface stresses and their depth of occurrence using FEM

Numerical calculations were done using the ANSYS program. Calculations for a complete bearing model requires a computer with very high computing power. Therefore, the calculations were carried out for the model comprising a single roller and a fragment of the outer or inner race. For this purpose, a numerical 3D solid model mapping the internal geometry of the bearing was built, which includes modified logarithmic correction of roller generators. The method of determining the value of forces acting on the rollers and the angles of inclination of the rollers relative to the raceways necessary to perform the numerical calculations are described in chapter 5.

Symmetry conditions of the tested bearing enabled the authors of the work to model half the roller and the corresponding part of the raceway of the inner or outer ring. Part of the raceway under consideration resulted from the dimensions of the bearing and the number of rollers in the tested bearing. Key point in the researching rolling bearings is the roller contact zone with the raceway. Strongly non-linear phenomena occur there, which is why it is very important to create contact pairs and proper setting of contact parameters. FEM systems do not allow to geometrically model curvilinear objects. The roller cross section, i.e. the circle, is a broken line, this is due to the fact that the examined 3D model is divided into finite elements. For this reason, the results of calculations obtained by numerical method strictly depend on the number of elements resulting from the division of the examined object into finite elements which in the case of contact phenomena is particularly important. For this reason, the division into elements in the area of anticipated contact has been increased and an uneven division of contact elements in this area has been used. The contact model was used for calculations in the contact zone: TARGE surface type to CONTA surface type of the contact. While analysing the character of the roller contact with the bearing raceways, calculations assumed the roller surface as the contact surface. The surface of the raceway was assumed as the target surface. The contact surface was modelled using contact elements Conta175 and the target surface was modelled using elements Targe170. For selected roller volumes in the contact zone, the edge is divided into equal 0.05 mm length parts. The edge of selected raceway volumes was divided into equal parts with a length of 0.1 mm. Because the length of the assumed element in the contact zone (0.05 mm) is less than half the width of the contact surface, which ranges between 0.1 mm up to 0.2 mm, the division assumed was considered appropriate to ensure reliable results.

In the numerical model of the non-contact part of the bearing, 8-node solid elements of the type SOLID185 were used. The numerical model under study was divided into 752540 solid finite elements. The distance between the raceway nodes and the roller are adjustable using angle which eliminated any possible shape errors. The calculations included symmetry conditions and degrees of freedom resulting from real working conditions. ALLDOF type displacement has been taken from the outer surfaces of the outer and inner race. The lateral surfaces of the outer and inner race are deprived of the possibility of displacement in the direction of the z axis. Conditions of symmetry of the model relative to the plane y-z were imposed. The coefficient of contact stiffness was assumed for calculations FKN=1.5 calculated according to the method described in paper [14]. The Augmented Lagrange method was used. Other values of the coefficients characterising the contact were left as default.

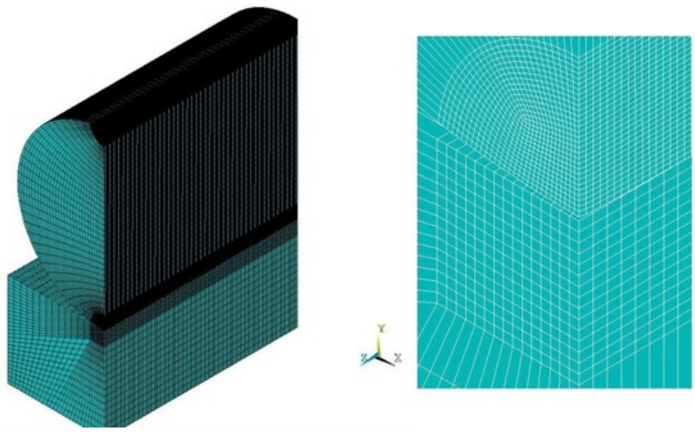


Fig. 3. 3D numerical model of the roller-raceway contact

4. Researched object

The researched object was a radial cylindrical roller bearing NJ 213 ECP that can carry an axial load [23]. The parameters of the analysed bearing are shown in Table 1.

Table 1. Parameters of the NJ 213 ECP cylindrical roller bearing [23]

Bearing bore diameter	d = 65 mm
Bearing outside diameter	D = 120 mm
Bearing width	B = 23 mm
Diameter of the outer ring raceway	d_{bo} = 108.5 mm
Diameter of the inner ring raceway	d_{bi} = 78.5 mm
Roller diameter	D_r = 15 mm
Roller length	L_r = 15 mm
Roller chamfer dimension	r_c = 0.5 mm
Number of rollers in the bearing	Z_r = 16
Dynamic load rating	C = 122000 N
Permissible axial load	$F_{ap\ max} \leq$ 5915 N

It was assumed that the profile of roller generators describes the curve corresponding to the modified logarithmic correction with the same parameters as correction of bearing rollers being the subject of analyses in the works [4, 29] (Fig. 4).

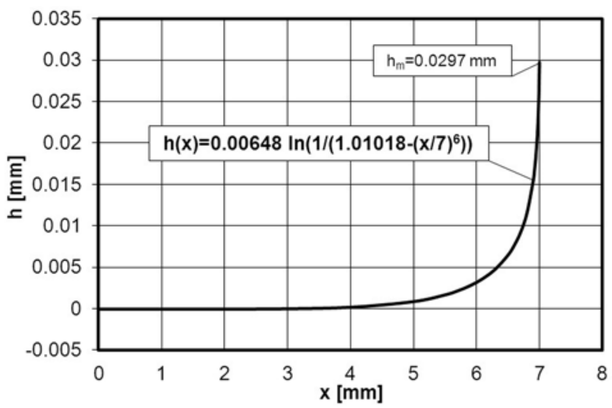


Fig. 4. Roller generator profile with a modified logarithmic correction [4, 29]

5. Radial and axial load distributions

Calculations of fatigue life of a radial cylindrical roller bearing were carried out for two values of radial load F_r, forming 10% and 20% of the dynamic bearing capacity of bearing C. First value, $F_r = 0.1\ C$, is similar to the radial load to which radial cylindrical roller bearings of the type NJ and NUP, used in the axle boxes of a typical passenger and freight bogie, are subjected. Second one, $F_r = 0.2\ C$, corresponds to the average radial load for general purpose cylindrical roller bearings, such as analysed NJ 213 ECP bearing. The axial force used in the calculations was of equal value in both cases, not exceeding the permissible value of axial load $F_{ap\ max}$. In the first case it was $F_a = 0.4\ F_r$, in second $F_a = 0.2\ F_r$. For comparison, calculations were also made only for radial loading, i.e. the case $F_a = 0$. It was assumed that the axes of both bearing rings remain parallel. The methodology used allows one to set any clearance value, both positive and negative. The value of radial clearance in the bearing $g = 0$ was taken for calculations.

Table 2. Radial and axial load distributions on NJ 213 ECP bearing rollers

Radial load		$F_r = 0.1\ C = 12200$ N			$F_r = 0.2\ C = 24400$ N		
Axial load		$F_a = 0.4\ F_r = 4880$ N			$F_a = 0.2\ F_r = 4880$ N		
ψ_{lim} [°]		147.78			116.8		
Roller No.	ψ_j [°]	Q_{rj} [N]	Q_{aj} [N]	η_j [']	Q_{rj} [N]	Q_{aj} [N]	η_j [']
1	0	3125	591	1.587	6224	625	1.566
2	22.5	2865	587	1.589	5701	620	1.566
3	45	2138	572	1.599	4241	603	1.569
4	67.5	1148	528	1.647	2159	569	1.588
5	90	456	327	1.986	453	324	1.973
6	112.5	131	117	2.411	12	12	2.704
7	135	14	14	2.711	0	0	0
8	157.5	0	0	0	0	0	0
9	180	0	0	0	0	0	0
Radial load		$F_r = 0.1\ C = 12200$ N			$F_r = 0.2\ C = 24400$ N		
Axial load		$F_a = 0$			$F_a = 0$		
ψ_{lim} [°]		90			90		
Roller No.	ψ_j [°]	Q_{rj} [N]	Q_{aj} [N]	η_j [']	Q_{rj} [N]	Q_{aj} [N]	η_j [']
1	0	3114	0	0	6228	0	0
2	22.5	2852	0	0	5704	0	0
3	45	2119	0	0	4238	0	0
4	67.5	1071	0	0	2142	0	0
5	90	0	0	0	0	0	0
6	112.5	0	0	0	0	0	0
7	135	0	0	0	0	0	0
8	157.5	0	0	0	0	0	0
9	180	0	0	0	0	0	0

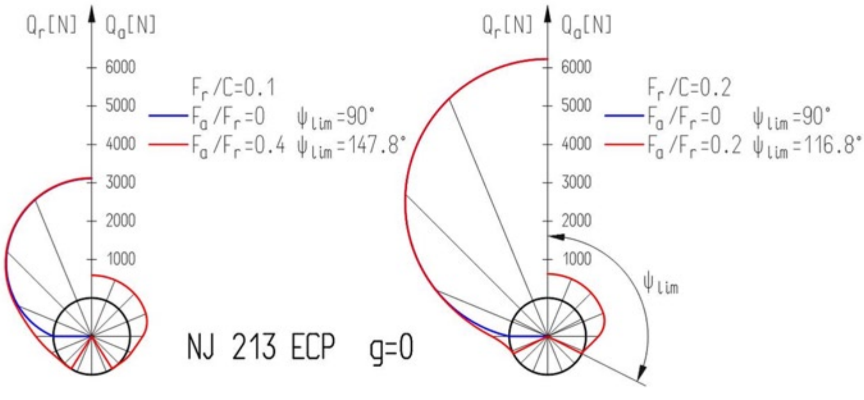


Fig. 5. The distributions of radial and axial load on the rollers of NJ 213 ECP bearing

The results of calculations of load distributions, obtained using the program ROLL1 [27], is shown in Table 2 and illustrated in figures 5 and 6.

As can be seen from the table and what can be seen in Figure 7, along with the increase in axial load, due to the rollers tilting (Fig. 8), the size of the loaded zone of bearing rollers increases. Its value is determined by the angle ψ_{lim}. It can also be seen that only rollers loaded with radial force can simultaneously carry axial load.

6. Distribution of maximum subsurface stress and their depth of occurrence

To calculate the predicted bearing fatigue life, knowledge of the maximum subsurface stress distributions along the roller-raceway contact line σ_{xj} and the depth distributions of these maximum stresses Z_{xj}, where j is the number of the next roller, is required. Distributions of maximum subsurface stresses were found using FEM method. For calculations, it was assumed that the rolling elements are made of elastic-ideally plastic material. The material properties were determined by Young's modulus $E = 208$ GPa, Poisson's ratio $\nu = 0.3$ and the tensile yield strength $\sigma_o = 1950$ MPa.

The ANSYS program allows one to illustrate stress distributions, values and locations of maximum and minimum stress in the form of maps. Sample stress maps for rollers number 1, 3, 5, 6 and loads $F_r = 0.2\ C$, $F_a = 0.2\ F_r$ are shown in figures 7, 8, 9 and 10. The program also allows one to get charts showing the maximum stress distributions along the roller contact line with the raceway, but it is not possible to directly obtain the distribution of the depth of maximum stresses. Basic ANSYS package only allows finding subsurface stress distribution in any plane

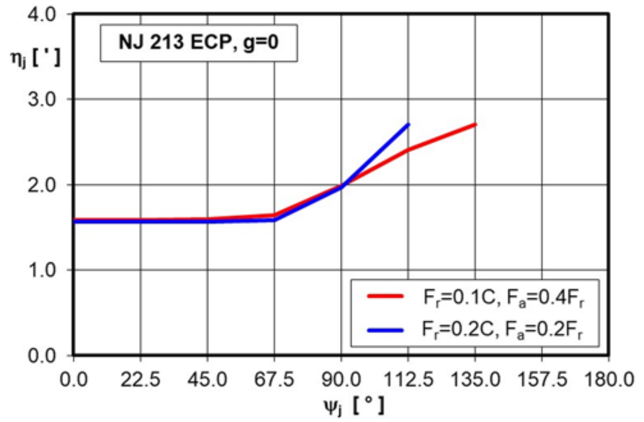


Fig. 6. Roller tilt angle η_j as a function of the angle ψ_j that determines the position of the roller in the bearing

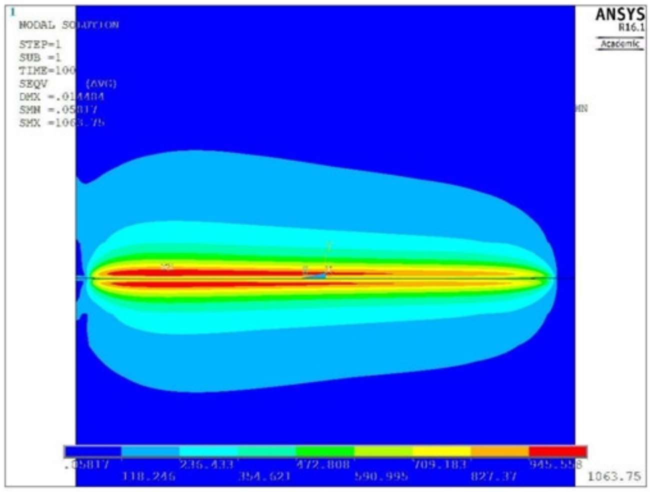


Fig. 7. Distributions of von Mises stresses below the roller-inner ring raceway contact surface; roller No.1, F_r = 0,2 C, F_a = 0,2 F_r

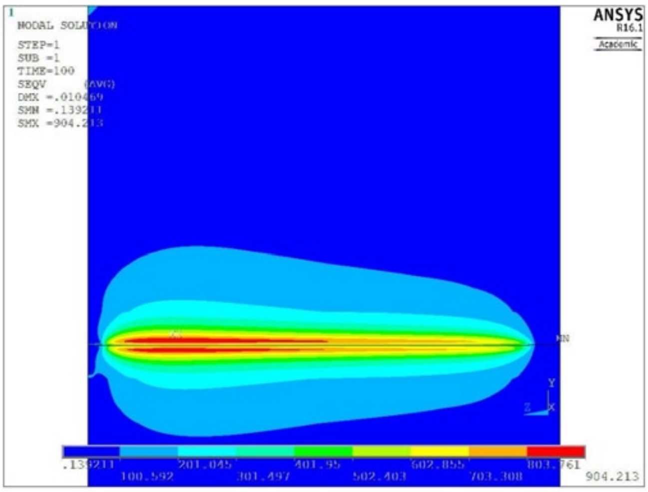


Fig. 8. Distributions of von Mises stresses below the roller-inner ring raceway contact surface; roller No.3, F_r = 0,2 C, F_a = 0,2 F_r

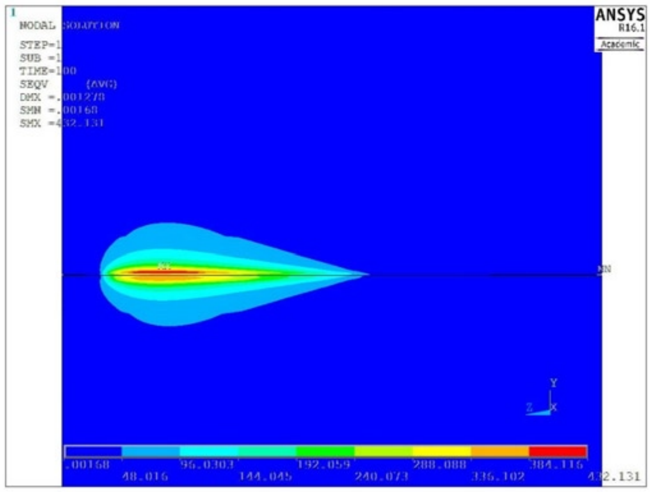


Fig. 9. Distributions of von Mises stresses below the roller-inner ring raceway contact surface; roller No.5, F_r = 0,2 C, F_a = 0,2 F_r

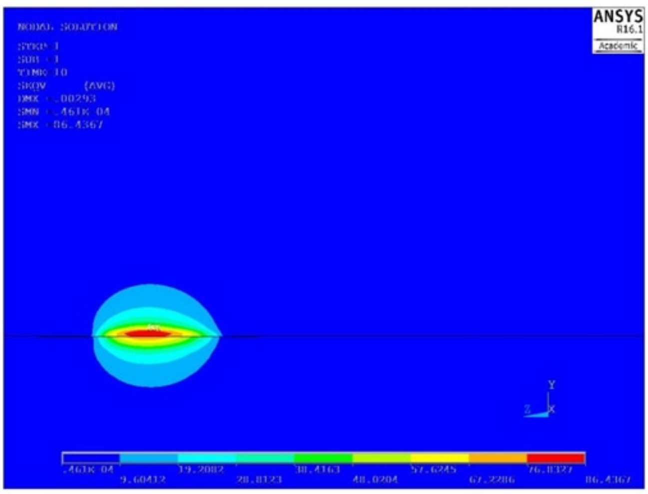


Fig. 10. Distributions of von Mises stresses below the roller-inner ring raceway contact surface; roller No.6, F_r = 0,2 C, F_a = 0,2 F_r

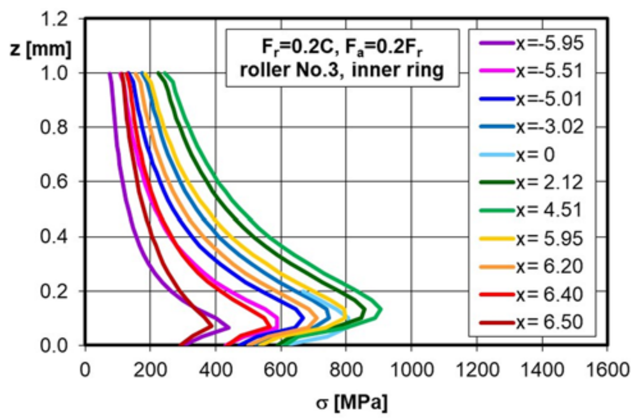


Fig. 11. Distribution of von Mises stresses in in selected cross-sections of the roller-inner ring raceway contact

specified by the user. By examining changes in subsurface stresses in selected sections, the distribution of the depth of occurrence of maximum subsurface stresses along the contact line of rolling elements can be determined. The same procedure was used to search for the distribution of the occurrence of maximum subsurface stress at study [4]. An example of a graph that allows finding the depth of occurrence of maximum subsurface stress is shown in Figure 11.

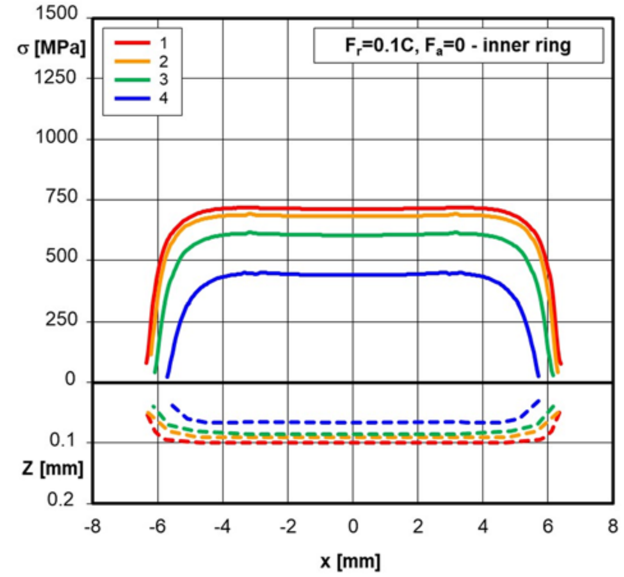


Fig. 12. Distributions of maximal von Mises stresses σ and the depth of their occurrence Z in contacts of subsequent rollers with the inner ring raceway; F_r = 0,1 C, F_a = 0

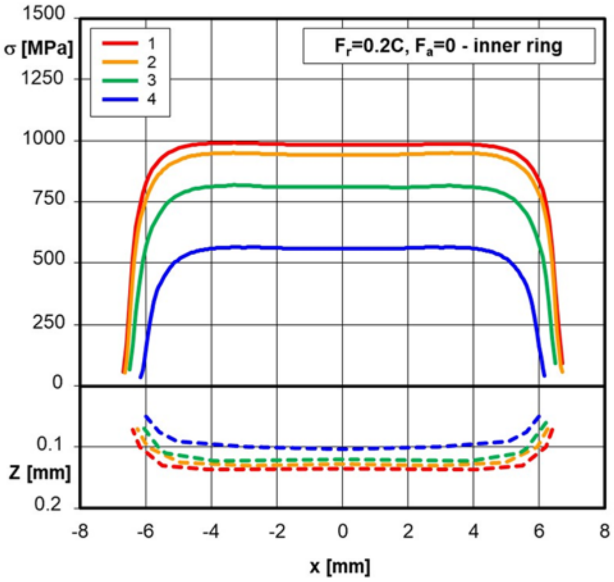


Fig. 13. Distributions of maximal von Mises stresses σ and the depth of their occurrence Z in contacts of subsequent rollers with the inner ring raceway; F_r = 0,2 C, F_a = 0

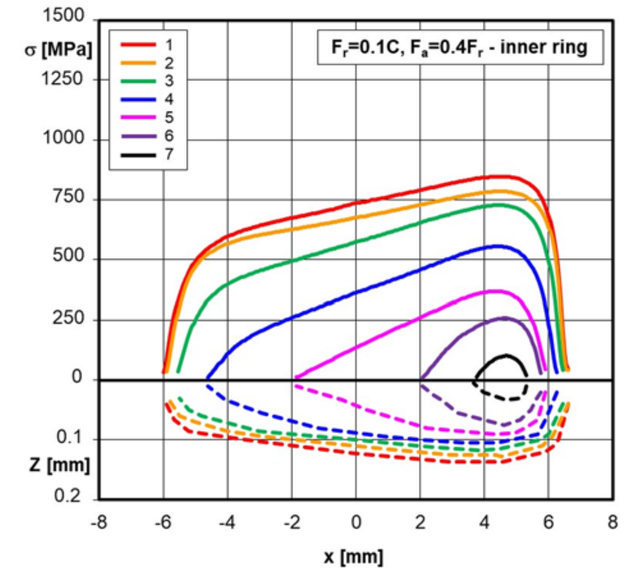


Fig. 14. Distributions of maximal von Mises stresses σ and the depth of their occurrence Z in contacts of subsequent rollers with the inner ring raceway; F_r = 0,1 C, F_a = 0,4 F_r

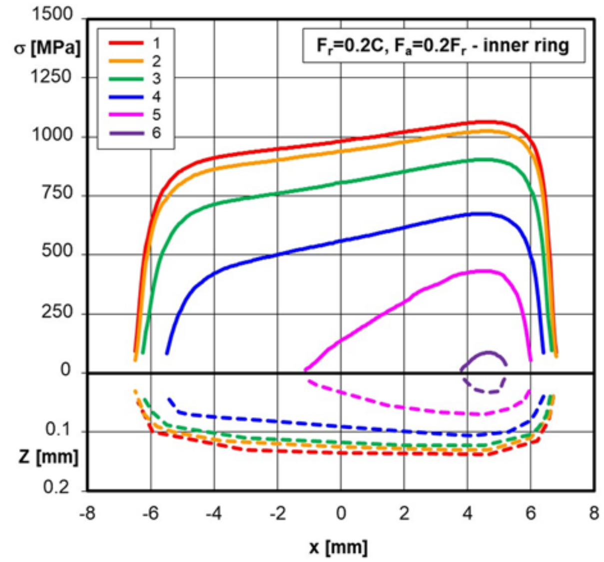


Fig. 15. Distributions of maximal von Mises stresses σ and the depth of their occurrence Z in contacts of subsequent rollers with the inner ring raceway; F_r = 0,2 C, F_a = 0,4 F_r

In figures 12, 13, 14 and 15 distributions of the maximum subsurface stress σ and the depth of their occurrence Z in roller contacts with the inner ring race obtained by FEM are presented. The numbers next to the coloured lines shown in the legend correspond to the numbers of consecutive load-bearing rollers (compare: Table 2).

Two first charts present distributions for cases where the bearing is only loaded with radial force: $F_r = 0.1\ C$ (Fig. 12) and $F_r = 0.2\ C$ (Fig. 13). The stress distribution σ and the depth of their occurrence Z are characterised by symmetry. Thanks to the logarithmic correction of the forming rollers, no stress accumulation occurs in the contacts but the length of the contact area of even the most loaded roller is less than the length of the forming roller l (Fig. 2) and decreases as the radial force on the roller decreases. As one can see, the logarithmic correction protects against pressure accumulation at the edges of the rollers, but prevents the roller working surface from being fully used.

The next two graphs illustrate stress distribution σ and their depth of occurrence Z for the bearing subjected, next to the radial load $F_r = 0.1\ C$ i $F_r = 0.2\ C$, to the axial load equal respectively: $F_a = 0.4\ F_r$ (Fig. 14) and $F_a = 0.2\ F_r$ (Fig. 15). Roller tilt relative to raceways due to axial load results in uneven stress distribution in the contacts, characterized by an increase in stress at one end of the contact field. In the load-bearing rollers area, the size of which is determined by the angle ψ_{lim}, two areas can be distinguished. In the first area, the rollers are in contact with the raceways virtually along the entire generator length, although in this case as well as for $F_a = 0$, the contact field length is always smaller than the length l. In the second area, the contact only exists on the roller generator part, in extreme cases, for rollers lying near the end of the loaded zone, the contact area is located at the very edge of the roller working surface (roller number 7 for $F_r = 0.1\ C$ and $F_a = 0.4\ F_r$, roller number 6 for $F_r = 0.2\ C$ and $F_a = 0.4\ F_r$, Fig. 10, 11). The authors of the study [3] noted the occurrence of similar bearing loaded zones.

7. Fatigue life of a bearing subjected to combined load

Distributions of maximum subsurface stress and depth of their occurrence at the contacts of rollers with inner and outer ring raceways obtained using FEM allowed for calculation of the predicted fatigue life of the NJ 213 ECP bearing.

In Table 3 the results of calculations of the predicted fatigue life of the tested bearing for the considered load cases were summarised. The durability of a bearing loaded only with radial force was compared with the durability of a bearing subjected to a combined load. In addition, the table presents the results of calculations of fatigue life according to formulae (1), (2) and (3) in accordance with the SKF catalogue method [23].

The two methods used to calculate the predicted fatigue life, numerical using stress information obtained using FEM and the catalogue method, give comparable results. Numerical method at lower radial loads (F_r = 0.1 C) gives results of the predicted durability slightly overstated compared to the catalogue method, and at higher loads (F_r = 0.2 C) understated. The axial load of the radial cylindrical roller bearing and the associated tilting of the rollers reduces fatigue life. The numerical method allows one to capture a decrease in durability even with an axial load that is a small part of the radial load (F_r = 0.2 C, F_a = 0.2 F_r, decrease in durability by 11% compared to durability for F_a = 0). In such cases, the catalogue method ignores the impact of axial load on fatigue life. The impact of axial load is taken into account by the catalogue method only after exceeding the threshold value, i.e. for $F_a / F_r > 0.2$. Then the predicted fatigue life of the bearing decreases abruptly. For the case F_r = 0,1 C and F_a = 0,4 F_r bearing life determined according to the catalogue method is reduced by 39% compared to the life of F_a = 0. The predicted fatigue life determined by the numerical method decreases to a lesser extent. For the present case, the decrease in durability compared to durability for F_a = 0 is 17%.

Numerical method using previously determined distributions of maximum subsurface stress and their depth in contact of rolling elements, enables more precise forecasting of fatigue life of a radial cylindrical roller bearing subjected to combined loading than allowed by catalogue methods. In addition, it allows one to take into account other factors affecting bearing durability, such as correction of forming rolling elements, or bearing clearance. The practical application of the proposed method reduces the need for time-consuming calculations of subsurface stresses using FEM. However, this time can be greatly reduced by using the Boussinesq solution for elastic half-space to determine the stress [4, 28, 29], which gives results comparable to FEM.

8. Summary

The article is dedicated to forecasting fatigue life of a radial cylindrical roller bearing type NJ loaded with radial and axial force. For forecasting, a numerical method using the maximum subsurface stress distributions and their depth in contact of rolling elements determined using the finite element method was used. Distribution of radial and axial forces for the analysed load cases, as well as the distribution of maximum stresses and their location along the contact line with the raceways of subsequent load-bearing rollers are presented. It has been demonstrated that the axial load of a radial cylindrical roller bearing and the associated tilting of rollers reduces fatigue life. The results of calculations of fatigue life obtained using the proposed method were compared with the results obtained according to the catalogue method, obtaining sufficient compliance.

The numerical methodology used in this work allows to include in the calculations of durability the impact of local changes in subsurface stresses in the contacts of mating elements, thanks to which it is possible to take into account many factors usually overlooked in engineering methods, described in catalogues of rolling bearings.

The most important factor is the combined bearing load, which impact on fatigue life is included in the catalogue methods in an approximate way in the form of dependence on the equivalent bearing load. The numerical method allows to accurately determine the predicted fatigue life of a bearing subjected to combined loading for any load combination.

Radial clearance is another important factor affecting the life of a cylindrical roller bearing. Radial clearance equal to zero was assumed in the calculations whose results were presented in the article, but it was possible to set a different value of the clearance, both positive and negative.

The last factor that affects the fatigue life of the bearing and which was included in the research is the correction of roller generators. Defining a generator profile is required when performing numerical calculations of pressure distributions. This makes it difficult to calculate the life of bearings with unknown profile geometry, however it allows one to optimise the correction parameters that create for the greatest fatigue life of the bearing operating under certain conditions.

References

1. Benchea M, Iovan-Dragomir A, Cretu S. Misalignment effects in cylindrical roller bearings. Applied Mechanics & Materials 2014; 658-659: 277-282, https://doi.org/10.4028/www.scientific.net/AMM.658.277.
2. Brandlein J. The fatigue life of axially loaded cylindrical roller bearings. FAG, Ball and Roller Bearing Engineering 1972; 1: 7-11.
3. Cheng W W, Shih S, Grace J, et al. Axial load effect on contact fatigue life of cylindrical roller bearings. ASME Journal of Tribology 2004; 126: 242-247, https://doi.org/10.1115/1.1614823.
4. Chudzik A, Warda B. Effect of radial internal clearance on the fatigue life of the radial cylindrical roller bearing. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (2): 211-219, https://doi.org/10.17531/ein.2019.2.4.
5. Claesson E. Modelling of roller bearings in ABAQUS. Chalmers, Applied Mechanics, Master's Thesis, Göteborg, Sweden, 2014: 1-38.
6. Demirhan N, Kanber B. Stress and displacement distribution on cylindrical roller bearing rings using FEM. Mechanics Based Design of Structures and Machines 2008; 36: 86-102, https://doi.org/10.1080/15397730701842537.
7. Fernlund I, Synek V. Influence of axial loads on the life of cylindrical roller bearings. SKF, The Ball Bearing Journal 1967; 151: 21-26.
8. Golbach H. Integrated non-linear FE module for rolling bearing analysis. INA reprint "Proceedings of NAFEMS WORLD CONGRESS '99 on Effective Engineering Analysis", Vol. 2, Newport, Rhode Island, USA, 25-28 April 1999: 1-12.
9. Göncz P, Ulbin M, Glodež S. Computational assessment of the allowable static contact loading of a roller-slewing bearing's case-hardened raceway. International Journal of Mechanical Sciences 2015; 94 95: 174-184, https://doi.org/10.1016/j.ijmecsci.2015.03.006.
10. Guo Y, Parker R G. Stiffness matrix calculation of rolling element bearings using a finite element/contact mechanics model. Mechanism and Machine Theory 2012; 51: 32-45, https://doi.org/10.1016/j.mechmachtheory.2011.12.006.
11. Kang Y, Shen P, Huang C, Shyr S, Chang Y. A modification of the Jones-Harris method for deep-groove ball bearings. Tribology International

2006; 39: 1413-1420, https://doi.org/10.1016/j.triboint.2005.12.005.
12. Kania L. Modelling of rollers in calculation of slewing bearing with the use finite elements. Mechanism and Machine Theory 2006; 41: 1359-1376, https://doi.org/10.1016/j.mechmachtheory.2005.12.007.
13. Kosmol J. Extended model of angular bearing - influence of fitting and pre-deformation. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (3): 493-500, https://doi.org/10.17531/ein.2019.3.16.
14. Harris TA, Kotzalas MN. Rolling Bearing Analysis, Fifth Edition -2 Volume Set. CRC Press; 2007, https://doi.org/10.1201/9781482275148.
15. Krzemiński-Freda H, Warda B. The effect of roller end-flange contact shape upon frictional losses and axial load of the radial cylindrical roller bearing. Proc. of the 15th Leeds-Lyon Symp. on Tribology, 6th-9th September 1988: 287-295, https://doi.org/10.1016/S0167-8922(08)70205-0.
16. Laniado-Jácome E: Numerical model to study of contact force in a cylindrical roller bearing with technical mechanical event simulation. Journal of Mechanical Engineering and Automation 2011; 1 (1): 1-7, https://doi.org/10.5923/j.jmea.20110101.01.
17. Lee J, Pan J. Closed-form analytical solutions for calculation of loads and contact pressures for roller and ball bearings. Tribology International 2016; 103: 187-196, https://doi.org/10.1016/j.triboint.2016.06.042.
18. Li S. A mathematical model and numeric method for contact analysis of rolling bearings. Mechanism and Machine Theory 2018; 119: 61-73, https://doi.org/10.1016/j.mechmachtheory.2017.08.020.
19. Liu J, Shi Z, Shao Y. An investigation of a detection method for a subsurface crack in the outer race of a cylindrical roller bearing. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19 (2): 220-228, https://doi.org/10.17531/ein.2017.2.8.
20. Lundberg G, Palmgren A. Dynamic capacity of rolling bearings. Acta Polytech Scand Mech Eng 1947; 1(3): 1-52.
21. Lundberg G, Palmgren A. Dynamic capacity of roller bearings. Acta Polytech Scand Mech Eng 1952; 2(4): 96-127.
22. Łazarz B., Peruń G., Bucki S.: Application of the finite-element method for determining the stiffness of rolling bearings. Transport Problems 2008; 3 (3): 33-40.
23. Łożyska toczne SKF. PUB BU/P1 10000 PL, 2014.
24. Poplawski J V, Peters S , Zaretsky E V. Effect of roller profile on cylindrical roller bearing life prediction - Part II Comparison of roller profiles. Tribology Transactions 2001; 44 (3): 417-427, https://doi.org/10.1080/10402000108982476.
25. Shah Maulik J, Darji P H. Fatigue life improvement through reduction of edge pressure in cylindrical roller bearing using FE analysis. International Journal For Technological Research in Engineering 2014; 1 (10): 1069-1074.
26. Tong V C, Kwon S W, Hong S W. Fatigue life of cylindrical roller bearings. Proc IMechE Part J: J Engineering Tribology 2016; 231: 623-636, https://doi.org/10.1177/1350650116668767.
27. Warda B. Wykorzystanie istniejących teorii zmęczenia powierzchniowego do prognozowania trwałości złożonych węzłów tocznych. Zeszyty Naukowe PŁ, Łódź 2009; 1055 (386): 1-159.
28. Warda B, Chudzik A. Fatigue life prediction of the radial roller bearing with the correction of roller generators. International Journal of Mechanical Sciences 2014; 89: 299-310, https://doi.org/10.1016/j.ijmecsci.2014.09.015.
29. Warda B, Chudzik A. Effect of ring misalignment on the fatigue life of the radial cylindrical roller bearing. International Journal of Mechanical Sciences 2016; 111-112: 1-11, https://doi.org/10.1016/j.ijmecsci.2016.03.019.
30. Xia X, Zhu S, Jia C, Niu R. Study of interval of arc modification length of cylindrical roller using ANSYS. International Journal of Engineering and Science 2012; 1 (1): 8-13.
31. Xia X, Shang Y, Zhu S. Effects of misalignment in cylindrical roller bearings on contact between roller and inner ring. IOSR Journal of Engineering 2013; 3 (5): 1-8, https://doi.org/10.9790/3021-03540108.
32. Zhao H: Analysis of load distributions within solid and hollow roller bearings. ASME Journal of Tribology 1998; 120: 134-139, https://doi.org/10.1115/1.2834176.
33. Zhenhuan Y, Liqin W, Le G, Chuanwei Z. Effects of tilted misalignment on loading characteristics of cylindrical roller bearings. Mechanism and Machine Theory 2013; 69:153-167, https://doi.org/10.1016/j.mechmachtheory.2013.05.006.

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

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

E-mails: agnieszka.chudzik@p.lodz.pl, bogdan.warda@p.lodz.pl

Article citation info:
LIU Y, WANG Y, FAN Z, HOUZ, ZHANG S, CHEN X. Lifetime prediction method for MEMS gyroscope based on accelerated degradation test and acceleration factor model. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 221–231, http://dx.doi.org/10.17531/ein.2020.2.5.

Yao LIU
Yashun WANG
Zhengwei FAN
Zhanqiang HOU
Shufeng ZHANG
Xun CHEN

LIFETIME PREDICTION METHOD FOR MEMS GYROSCOPE BASED ON ACCELERATED DEGRADATION TEST AND ACCELERATION FACTOR MODEL

METODA PROGNOZOWANIA CZASU PRACY ŻYROSKOPU MEMS NA PODSTAWIE TESTU PRZYSPIESZONEJ DEGRADACJI I MODELU WSPÓŁCZYNNIKA PRZYSPIESZENIA

The reliability analysis of MEMS gyroscope under long-term operating condition has become an urgent requirement with the enlargement of its application scope and the requirement of good durability. In this study we propose a lifetime prediction method for MEMS gyroscope based on accelerated degradation tests (ADTs) and acceleration factor model. Firstly, the degradation characteristic (bias instability) is extracted based on Allan variance. The effect of temperature stress on the degradation rate of bias instability is analyzed, and it shows that the degradation rate of bias instability would increase with the increase of the temperature. Secondly, the ADTs of MEMS gyroscope are designed and conducted, the degradation model of MEMS gyroscope is established based on the output voltage of MEMS gyroscope and Allan variance. Finally, the acceleration factor model of MEMS gyroscope under temperature stress is derived, and the lifetime of the MEMS gyroscope is predicted based on two group tests data under high stress level. The results show that the lifetime calculated by the acceleration factor model and mean lifetime under high stress levels is close to the mean lifetime calculated by the linear equation at normal temperature stress.

Keywords: *MEMS gyroscope, temperature stress, accelerated degradation test, acceleration factor model, Allan variance.*

Analiza niezawodności żyroskopu MEMS w warunkach długotrwałej pracy stała się pilną koniecznością wraz z rozszerzeniem zakresu jego zastosowania i wprowadzeniem wymogu dobrej trwałości. W niniejszym artykule, zaproponowano metodę prognozowania czasu pracy żyroskopu MEMS w oparciu o testy przyspieszonej degradacji i model współczynnika przyspieszenia. W pierwszej kolejności, wyznaczono charakterystykę degradacji (niestabilność wskazań) na podstawie wariancji Allana. Analizowano wpływ naprężenia cieplnego na szybkość degradacji w zakresie niestabilności wskazań. Analiza wykazała, że szybkość degradacji wzrastała wraz ze wzrostem temperatury. Następnie, opracowano i przeprowadzono testy przyspieszonej degradacji żyroskopu MEMS, a model jego degradacji ustalono na podstawie napięcia wyjściowego żyroskopu i wariancji Allana. Na koniec, wyprowadzono model współczynnika przyspieszenia dla żyroskopu MEMS w warunkach naprężenia cieplnego, a żywotność żyroskopu prognozowano na podstawie danych z dwóch testów grupowych przeprowadzonych w warunkach wysokiego naprężenia. Wyniki pokazują, że czas pracy obliczony na podstawie modelu współczynnika przyspieszenia i średni czas pracy przy wysokich poziomach naprężeń są zbliżone do średniego czasu pracy obliczonego na podstawie równania liniowego przy normalnym naprężeniu cieplnym.

Słowa kluczowe: *żyroskop MEMS, naprężenie cieplne, test przyspieszonej degradacji, model współczynnika przyspieszenia, wariancja Allana.*

1. Introduction

MEMS gyroscopes, due to its lower mass, lower power consumption, and smaller volume, have become one of the most promising inertial components, widely used in commercial, military and other fields [13], such as inertial navigation devices, portable electronic products and so on. However, MEMS gyroscopes often undergo high temperature condition, and physical properties of silicon materials would be significantly affected by temperature. How to predict the lifetime of the MEMS gyroscope has become an urgent problem to be solved in engineering [32].

Accelerated degradation test (ADT) collects performance degradation data of products at accelerated stress levels, and uses these data to extrapolate reliability and lifetime of products for the use conditions [20, 23, 33, 34]. Mulloni studied the lifetime prediction of RF-MEMS switches based on ADT [22]. The reliable lifetime of the smart electricity meter was predicted by an accelerated degradation test and degradation model [17, 35]. Wang [33] made a research on the lifetime prediction of self-lubricating spherical plain bearings based on the ADT. An ADT of light bars was tested under different stresses, and a response model based on an inverse power (exponential) law for the failure time under different stresses was then calculated to predict the lifetime under operating conditions [10]. A progressive-stress accelerated degradation test with a non-linear degradation path was proposed to obtain timely information of the product's lifetime distribution for highly reliable products [6]. Chen studied the accelerated degradation reliability modeling and test data statistical analysis of aerospace electrical connector [5]. Lifetime prediction technology based on ADT has become an inevitable requirement for reliability prediction of MEMS gyroscopes under time and cost constraints [4, 31].

The selection of accelerated stress is the first step in the implementation of ADT. Many scholars have studied the effect of temperature on the performance of MEMS device. Jin-Won Joo studied the global and local deformations of packages caused by temperature, and studied the influence of temperature on frequency shift of MEMS gyroscope [15]. Chun-Lin Lu analyzed the thermal stress for a CMOS-MEMS microphone with various metallization and materials [8]. Saeedivahdat et al. studied the effect of thermal stress on frequency response for a diaphragm in MEMS microphone by numerical model [29]. Therefore, temperature stress may be the most likely accelerated stress of MEMS gyroscope in ADTs.

Extraction of degradation characteristic is the key step to establish degradation model [25, 30]. There are many noise terms in the output voltage of MEMS gyroscope due to its working principle and processing technology, which makes the extraction of the effective degradation characteristics more complicated. Many scholars have studied the noise terms by power spectral density (PSD), time domain analysis methods (autocorrelation function, Gauss Markov process and autoregressive model) and the Kalman filter method. However, these methods have their own inherent shortcomings [16, 19, 36].

Allan variance is an important means to analyze the random noise characteristics of gyroscope, including laser gyroscope, fiber optic gyroscope and MEMS gyroscope. The method is a time domain analysis method proposed by the American National Bureau of Standards, which has been an IEEE standard method for parameter analysis of gyroscopes [9]. Allan variance usually describe the noise of gyroscope as several stochastic characteristics: quantization noise, angular random walk, bias instability, rate random walk and rate ramp. Some stochastic characteristics often directly reflect the performance of gyroscope, such as angular random walk and bias instability [1-3].

Many scholars have studied the performance of laser gyroscope and fiber optic gyroscope based on Allan variance. Lawrence analyzed the performance of ring laser gyroscope based on Allan variance [18]. Pin analyzed the noise characteristics of fiber optic gyroscope based on weighted least squares algorithm and Allan variance [26]. Grantham analyzed the random error of inertial sensors using Allan variance [11]. Chikovani evaluated the performance of laser gyroscopes, and fiber optic gyroscopes from different countries based on Allan variance [7]. However, the research on performance study of MEMS gyroscope is limited.

The rest of this document is organized as follows. In Section 2, we extract the degradation characteristic (bias instability) of MEMS gyroscope based on Allan variance, and analyze the effect of temperature stress on the degradation rate of bias instability. In Section 3, a universal lifetime prediction method of MEMS gyroscope is proposed. In Section 4, a case study is provided to illustrate the application and utility of the proposed method. Finally, the conclusions are presented in Section 5.

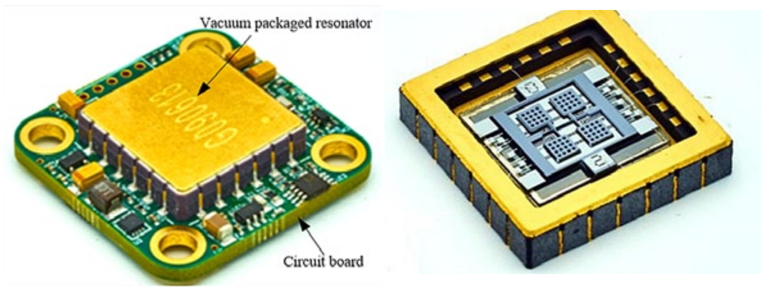
2. Theory

2.1. Structure and operating principle of MEMS gyroscope

Almost all MEMS gyroscope use vibrating mechanical elements to sense rotation, ‘‘Butterfly’’ MEMS gyroscope is a typical vibrating MEMS gyroscope based on the Coriolis effect [14]. As shown in Figure1, it has a four-mass full-differential structure with high sensitivity, good linearity and low cost, and is widely used in commercial fields.

"Butterfly" MEMS gyroscope mainly contains the circuit board and the vacuum packaged resonator, as shown in Figure 1(a). The image of resonator is shown in Figure 1(b), which consists of a single crystal silicon structure and a Pyrex glass base with patterned electrodes [14]. The silicon structure is manufactured by anisotropic wet etching process, which mainly includes four proof masses, a slanted suspension beam, and four trapezoidal cantilever beams, as shown in Figure 2(a). The excitation electrodes and detection electrodes are fabricated on Pyrex glass base under the proof mass, as shown in Figure 2(b), and it is connected with silicon structure by anodic bonding to form driving capacitance and detecting capacitance. The cross section of the slanted suspension beam (A-A) is composed of (100) and (111) crystal planes with an angle of 54.74°, as shown in Figure 2(c).

The slanted beam with an asymmetric cross section so that vertical electrostatic forces bend the beams both vertically and horizontally, as shown in Figure 2(c). The excitation mode is the flexural vibration of the slanted beam; the detection mode is the torsional vibration of the slanted beam. The excitation mode and the detection mode are two perpendicular degenerate modes of the axially symmetric elastic body. There will be a sinusoidal oscillation in detection axis direction due to the Coriolis Effect when an angular rate inputs.



(a) MEMS gyroscope (b) Unpackaged resonator

Fig. 1. EG136A type MEMS gyroscope , which consists of the circuit board and the vacuum packaged resonator

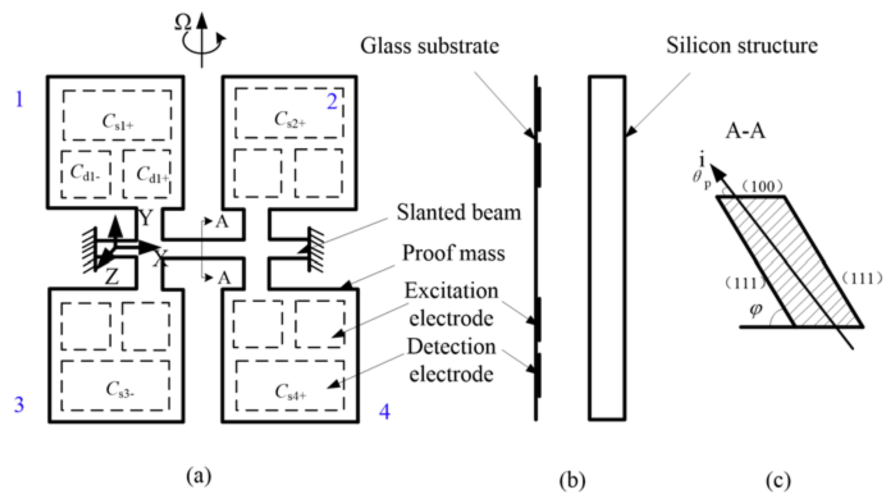


Fig. 2 Silicon structure and the electrode layout of ‘’Butterfly’’ MEMS gyroscope

Electrostatic force from the excitation electrodes makes the four proof masses vibrate with stable amplitude and phase on its resonant frequency by automatic gain control methods. There will be an oscillation in detection axis (Z direction) due to the Coriolis Effect if there is an angular rate in the sensitive axis (Y direction), which reflects the value of the input angular rate.

2.2. The principle of Allan variance

Allan Variance is a method of describing the root mean square of random drift error. Sampling the angular rate of MEMS gyroscope with sampling time τ_0, and the total sample size is M. Then m ($m=\tau_0$, $2\tau_0$, ..., $M/2\tau_0$]) data points are made as one cluster, so we can get $J=[M/m]$ clusters, where $[x]$ represents the floor of x. J is the number of sampled groups. The time of each group $\tau=m\tau_0$ is the correlation time, and the average value of the defined angular rate can be expressed as:

$$\bar{w}_k(m)=\frac{1}{m}\sum_{i=1}^{m} w_{(k-1)m+i} \qquad k=1,2,\cdots\left[\frac{M}{2}\right] \tag{1}$$

The Allan variance defined by the angular rate can be expressed as:

$$\sigma^2(\tau)=\frac{1}{2(J-1)}\sum_{k=1}^{J-1}\left[\bar{w}_{k+1}(m)-\bar{w}_k(m)\right]^2 \tag{2}$$

where $\langle y\rangle$ represents the average of y.

Allan variance can usually subdivide the random noise characteristics of the inertial device (MEMS gyroscope) into five typical characteristics, namely quantization noise, angular random walk, bias instability, and rate random walk and rate ramp. It can be expressed as [28]:

$$\begin{cases}\sigma_Q^2=3Q^2/\tau^2\\ \sigma_N^2=N^2/\tau\\ \sigma_B^2=2\ln 2/\pi\, B_s^2\\ \sigma_K^2=\kappa^2\tau/3\\ \sigma_R^2=R^2\tau^2/2\end{cases} \tag{3}$$

where Q is a random drift caused by quantization noise. The initial output of the inertial sensor is analog data, which is quantized into digital data for calculation, thus quantization noise is generated at the output of the inertia device [26]. N is the white noise generated during the measurement process, which appears as an angular random walk at the output of the MEMS gyro and belongs to high frequency noise [19]. B_s is a random drift caused by bias instability, which is a kind of low-frequency noise generated by the sensor electronic components or the surrounding environment [26]. κ is a random drift caused by random walk. The source of this error is still unclear. R is the random drift caused by the rate ramp, which is caused by the monotonic change of the sensor for a long time [26].

It is generally considered that five typical characteristics are independent of each other, so the total variance can be expressed as [24]:

$$Y=X\beta+e \tag{4}$$

where:

$$Y=\begin{bmatrix}\sigma_1^2\\ \sigma_2^2\\ \vdots\\ \sigma_m^2\end{bmatrix},\ X=\begin{bmatrix}\tau_1^{-2} & \tau_1^{-1} & 1 & \tau_1^{1} & \tau_1^{2}\\ \tau_2^{-2} & \tau_2^{-1} & 1 & \tau_2^{1} & \tau_2^{2}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \tau_m^{-2} & \tau_m^{-1} & 1 & \tau_m^{1} & \tau_m^{2}\end{bmatrix},\ \beta=\begin{bmatrix}3Q^2\\ N^2\\ \left(2\ln 2/\pi\, B_s\right)^2\\ \kappa^2/3\\ R^2/2\end{bmatrix}\ \text{and,}\ e=\begin{bmatrix}w_1\\ w_2\\ \vdots\\ w_m\end{bmatrix} \tag{5}$$

Five unknown parameters Q, N, B_s, κ , R can be obtained by the least squares algorithm:

$$\beta=\left(X^TX\right)^{-1}X^TY \tag{6}$$

2.3. The effect of temperature stress on the degradation rate of bias instability

Bias instability of MEMS gyroscope in a certain period of time can be calculated by Allan variance. In order to characterize the degradation characteristic changing with time, we calculate the bias instability using daily data, so the bias instability in D days are B_s^1 , B_s^2 ..., B_s^D respectively.

The five characteristics of the MEMS gyroscope under long-term operation (D days) can be expressed as:

$$\beta_t=\left(X^TX\right)^{-1}X^TY_t \qquad (t=1,2,3,...,D) \tag{7}$$

where:

$$Y_t=\begin{bmatrix}\sigma_{t1}^2\\ \sigma_{t2}^2\\ \vdots\\ \sigma_{tm}^2\end{bmatrix},\ X=\begin{bmatrix}\tau_1^{-2} & \tau_1^{-1} & 1 & \tau_1^{1} & \tau_1^{2}\\ \tau_2^{-2} & \tau_2^{-1} & 1 & \tau_2^{1} & \tau_2^{2}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \tau_m^{-2} & \tau_m^{-1} & 1 & \tau_m^{1} & \tau_m^{2}\end{bmatrix}\ \text{and,}\ \beta_t=\begin{bmatrix}3Q_t^2\\ N_t^2\\ \left(2\ln 2/\pi\, B_s^t\right)^2\\ \kappa_t^2/3\\ R_t^2/2\end{bmatrix} \tag{8}$$

where σ_{tm}^2 is the calculated Allan variance of the t-th day. Q_t is the quantization noise of the t-th day, N_t is the angular random walk of the t-th day, B_s^t is the bias instability of the t-th day, κ_t is the rate ramp of the t-th day, and R_t is the rate random walk of the t-th day.

Equation (7) can also be simplified as:

$$\beta_t=A_{5\times m}Y_t \qquad (t=1,2,3,...,D) \tag{9}$$

In order to study the bias instability B_s in the vector β_t, it assumed that the elements of the third row in the matrix A are a_{31}, a_{32}, $a_{33...}$ a_{3m}, so equation (9) can be further simplified as:

$$\left(2\ln 2/\pi\, B_s(t)\right)^2=\left[a_{3,1},a_{3,2},a_{3,3},a_{3,m}\right]\begin{bmatrix}\sigma_{t,1}^2\\ \sigma_{t,2}^2\\ \vdots\\ \sigma_{t,m}^2\end{bmatrix}=a_{3,1}\sigma_{t,1}^2+a_{3,2}\sigma_{t,2}^2+\cdots+a_{3,m}\sigma_{t,m}^2 \tag{10}$$

The bias instability change with time can be expressed as (equation (11) is verified in Section 4.3):

$$B_s^t = B_s(t) = Kt + b \tag{11}$$

where K is degradation rate of the bias instability, and b is the intercept.

Substituting equation (11) into equation (10), equation (10) can also be expressed as:

$$\left[0.6643B_s(t)\right]^2 = \left[0.6643(Kt+b)\right]^2 = a_{3,1}\sigma_{t,1}^2 + a_{3,2}\sigma_{t,2}^2 + \cdots + a_{3,m}\sigma_{t,m}^2 \tag{12}$$

Partial derivatives for temperature in both sides of equation (12), we can get:

$$0.8826(Kt+b)\frac{\partial K}{\partial T} = \frac{\partial\left(a_{3,1}\sigma_{t,1}^2 + a_{3,2}\sigma_{t,2}^2 + \cdots + a_{3,m}\sigma_{t,m}^2\right)}{\partial T} \tag{13}$$

Equations (7)-(9) show that the third row elements a_{31}, a_{32}, a_{33}... a_{3m} in matrix A are only related to the correlation time τ, so the right term of the equation (13) can be written as:

$$\frac{\partial\left(a_{3,1}\sigma_{t,1}^2 + a_{3,2}\sigma_{t,2}^2 + \cdots + a_{3,m}\sigma_{t,m}^2\right)}{\partial T} = a_{3,1}\frac{\partial\sigma_{t,1}^2}{\partial T} + a_{3,2}\frac{\partial\sigma_{t,2}^2}{\partial T} + a_{3,3}\frac{\partial\sigma_{t,3}^2}{\partial T} + \cdots + a_{3,m}\frac{\partial\sigma_{t,m}^2}{\partial T} \tag{14}$$

Substituting equation (2) into equation (14), equation (14) can be expressed as:

$$\begin{aligned} & a_{3,1}\frac{\partial\sigma_{t,1}^2}{\partial T} + a_{3,2}\frac{\partial\sigma_{t,2}^2}{\partial T} + a_{3,3}\frac{\partial\sigma_{t,3}^2}{\partial T} + \cdots + a_{3,m}\frac{\partial\sigma_{t,m}^2}{\partial T} \\ & = \frac{a_{3,1}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\bar{w}_{k+1}(1) - \bar{w}_k(1)\right]^2}{\partial T} + \frac{a_{3,2}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\bar{w}_{k+1}(2) - \bar{w}_k(2)\right]^2}{\partial T} \\ & + \frac{a_{3,m}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\bar{w}_{k+1}(m) - \bar{w}_k(m)\right]^2}{\partial T} \end{aligned} \tag{15}$$

The first term in equation (15) denotes the square of the difference between adjacent arrays when the correlation time τ=1. Substituting equation (1) into the first item of equation (15), we can get:

$$\begin{aligned} & \frac{a_{3,1}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\bar{w}_{k+1}(1) - \bar{w}_k(1)\right]^2}{\partial T} \\ & = \frac{a_{3,1}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[w_{k+1} - w_k\right]^2}{\partial T} \end{aligned} \tag{16}$$

It is assumed that all the parameters (electrical parameters and structural parameters) of the MEMS gyroscope do not change during the operating process, only the environment temperature changes from T to $T+\Delta T$, resulting in the MEMS gyroscope's output (angular rate) changing from w_k to w_{k+1}, so equation (16) can be expressed as:

$$\begin{aligned} & \frac{a_{3,1}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\bar{w}_{k+1}(1) - \bar{w}_k(1)\right]^2}{\partial T} \\ & = \frac{a_{3,1}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[w_{k+1}(T+\Delta T) - w_k(T)\right]^2}{\partial T} \end{aligned} \tag{17}$$

The second term in equation (15) denotes the square of the difference between adjacent arrays when the correlation time τ=2. Substituting equation (1) into the first item of equation (15), we can get:

$$\begin{aligned} & \frac{a_{3,2}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\bar{w}_{k+1}(2) - \bar{w}_k(2)\right]^2}{\partial T} \\ & = \frac{a_{3,2}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\frac{1}{2}\sum_{i=1}^{2}w_{2k+i} - \frac{1}{2}\sum_{i=1}^{2}w_{2(k-1)+i}\right]^2}{\partial T} \end{aligned} \tag{18}$$

Similarly, it is assumed that all the parameters (electrical parameters and structural parameters) of the MEMS gyroscope do not change during the working process, only the environment temperature changes from T to $T+\Delta T$, resulting in the MEMS gyroscope's output (angular rate) changing from w_{2k+i} to $w_{2(k-1)+i}$, so equation (18) can be expressed as:

$$\begin{aligned} & \frac{a_{3,2}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\bar{w}_{k+1}(2) - \bar{w}_k(2)\right]^2}{\partial T} \\ & = \frac{a_{3,2}}{8(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\sum_{i=1}^{2}w_{2k+i}(T+\Delta T) - \sum_{i=1}^{2}w_{2(k-1)+i}(T)\right]^2}{\partial T} \end{aligned} \tag{19}$$

So the m-th term in equation (18) can be expressed as:

$$\begin{aligned} & \frac{a_{3,m}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\bar{w}_{k+1}(m) - \bar{w}_k(m)\right]^2}{\partial T} \\ & = \frac{a_{3,m}}{2m^2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\sum_{i=1}^{m}w_{km+i}(T+\Delta T) - \sum_{i=1}^{m}w_{(k-1)m+i}(T)\right]^2}{\partial T} \end{aligned} \tag{20}$$

Substituting equation (17) equation (19) and equation (20) into equation (15), we can get:

$$\begin{aligned} & a_{3,1}\frac{\partial\sigma_{t,1}^2}{\partial T} + a_{3,2}\frac{\partial\sigma_{t,2}^2}{\partial T} + a_{3,3}\frac{\partial\sigma_{t,3}^2}{\partial T} + \cdots + a_{3,m}\frac{\partial\sigma_{t,m}^2}{\partial T} \\ & = \frac{a_{3,1}}{2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[w_{k+1}(T+\Delta T) - w_k(T)\right]^2}{\partial T} \\ & + \frac{a_{3,2}}{2\cdot 2^2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\sum_{i=1}^{2}w_{2k+i}(T+\Delta T) - \sum_{i=1}^{2}w_{2(k-1)+i}(T)\right]^2}{\partial T} \\ & +\ldots+ \frac{a_{3,m}}{2m^2(J-1)}\sum_{k=1}^{J-1}\frac{\partial\left[\sum_{i=1}^{m}w_{km+i}(T+\Delta T) - \sum_{i=1}^{m}w_{(k-1)m+i}(T)\right]^2}{\partial T} \end{aligned} \tag{21}$$

It can be known that the solution of partial derivative is independent of the value of k from equation (21), without a loss of generality, let k = 3, we prove that the partial derivative is greater than 0. The proof process is as follows.

Let:

$$x(T+\Delta T)=\sum_{i=1}^{m} w_{3m+i}(T+\Delta T)$$
$$x(T)=\sum_{i=1}^{m} w_{2m+i}(T) \tag{22}$$

$\frac{\partial x(T)}{\partial T}>0$ due to $\frac{\partial w}{\partial T}>0$ [26], so:

$$\frac{\partial\left[x(T+\Delta T)-x(T)\right]^2}{\partial T}=2\left[x(T+\Delta T)-x(T)\right]\frac{\partial x(T+\Delta T)-\partial x(T)}{\partial t}>0 \tag{23}$$

So we can get:

$$\frac{\partial\left[\sum_{i=1}^{m} w_{3m+i}(T+\Delta T)-\sum_{i=1}^{m} w_{2m+i}(T)\right]^2}{\partial T}>0 \tag{24}$$

Because the correlation time τ is positive, the third row elements of matrix A are all positive, so:

$$0.8826(Kt+b)\frac{\partial K}{\partial T}>0 \tag{25}$$

Because the bias instability is larger than 0, so:

$$\frac{\partial K}{\partial T}>0 \tag{26}$$

It indicates that the degradation rate of bias instability would increase with the increase of the temperature. So bias instability can be the degradation characteristic of MEMS gyroscope, and the temperature can be the accelerated stress of MEMS gyroscope in ADTs.

3. A universal lifetime prediction method of MEMS gyroscope

3.1. Model assumptions

The degradation of the product is affected by temperature stress. The degradation process of the product can be accelerated by the temperature stress higher than the use or storage conditions. In ADTs, the stress level of the accelerated stress is L. The setting of stress level of accelerated stress does not change the degradation mechanism of the product, that is, the degradation mechanism of the product in ADTs is consistent with that in normal use.

It is assumed that the theoretical degradation trajectory and accelerated degradation trajectory of the product satisfy the linear degradation model, which can be expressed as:

$$B_{s,ijk}(t|S_i)=K_i t+b_i+\varepsilon_{ijk},\quad i=1,2,...,M,\ j=1,2,...,n_i,k=1,2,...,l_i \tag{27}$$

where $B_{s,ijk}$ is the characteristic value (bias instability) of the k-th measurement of the j-th specimen under the i-th stress level, t is the time, ε_{ijk} is the random measurement error, K_i is the degradation rate under the i-th stress level, and b_i is the intercept under the i-th stress level.

3.2. The description of the universal method for lifetime prediction of MEMS gyroscope

On the basis of the above model assumptions, a universal method for lifetime prediction of MEMS gyroscope is proposed, as shown in Figure 3. Firstly, a long-term test platform for MEMS gyroscope is constructed, and the ADT scheme is designed. The degradation characteristic (bias instability) of MEMS gyroscope is extracted based on the output voltage of the MEMS gyroscope and Allan variance. In order to characterize the degradation trend of MEMS gyroscope, we propose the following steps:

Step 1, Convert the output voltage of the MEMS gyroscope to angular rate in each day, and then calculate the root mean square value of the angular rate in each day based on equation (2). Step 2, The bias instability of MEMS gyroscope in every day can be calculated by the least squares method, i.e. equation (6). Step 3, The degradation trend of bias instability in D days can be obtained. Step 4, Fit the degradation trend of bias instability with linear function.

Then the degradation model of MEMS gyroscope can be established. Because the accelerated stress in ADTs is temperature, so we use Arrhenius model as the acceleration model [17, 27, 35], and the degradation rate of the MEMS gyroscope can be expressed by:

$$\frac{\mathrm{d}B_s(t)}{\mathrm{d}t}=K=A\exp(-E/k_bT) \tag{28}$$

where k_B is the Boltzmann constant, which is 8.6171×10^{-5}eV/°C. T is the Kelvin temperature with a unit of K, A is an unknown parameter and E is the activation energy.

Taking the natural logarithm on both sides of equation (28) results in:

$$\ln\left(\frac{\mathrm{d}B_s(t)}{\mathrm{d}t}\right)=\ln A-E/k_bT \tag{29}$$

So, the activation energy E can be expressed as:

$$E=-k_b\frac{\partial\left[\ln\left(\frac{\mathrm{d}B_s(t)}{\mathrm{d}t}\right)\right]}{\partial\left(\frac{1}{T}\right)} \tag{30}$$

Integrating on the both two sides of equation (28) results in:

$$\int_0^{B_s^{cir}}\mathrm{d}(B_s)=A\exp(-E/k_bT)\int_o^{TF}\mathrm{d}t \tag{31}$$

where TF is the failure time and B_s^{cir} is failure threshold of bias instability.

So the failure time TF can be expressed as:

$$TF=\frac{B_s^{cir}}{A\exp(-E/k_bT)} \tag{32}$$

The acceleration factor model of MEMS gyroscope can be expressed as:

$$AF = \frac{TF_0}{TF_1} = \frac{\exp\left(-E / k_b T_1\right)}{\exp\left(-E / k_b T_0\right)} = \exp\left(\frac{E}{k_b}\left(\frac{1}{T_0} - \frac{1}{T_1}\right)\right) \quad (33)$$

where TF_0 and TF_1 are failure time under normal stress and accelerated stress, T_0 and T_1 are normal temperature level and high temperature level.

As shown in equation (33), the acceleration factor model of MEMS gyroscope contains two unknown parameters, which can be estimated by least square method. Finally, the lifetime under normal stress level can be calculated by acceleration factor model and failure data under high stress level.

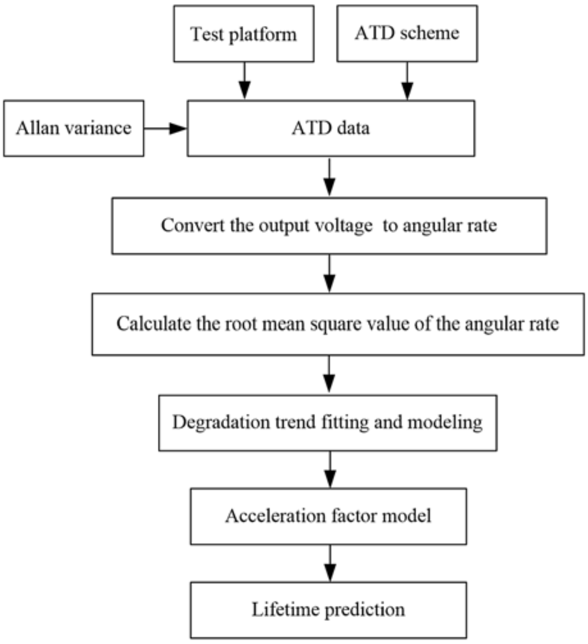


Fig. 3. The method diagram of lifetime prediction for MEMS gyroscope

4. Application Case

4.1. Design of ATD scheme and construction of test platform

According to GJB150.1-86, two accelerated stress levels were selected as S_1=45°C, S_2=60°C, and the normal temperature stress S_0=25°C. The number of specimens under each stress level was 4. Twelve MEMS gyroscopes, numbered 1#, 2#, 3#,..., 11#, 12#, were selected from the same batch MEMS gyroscopes. 1-4# MEMS gyroscopes were tested at 25℃ for 23 days, 5-8# MEMS gyroscopes were tested at 45℃ for 23 days, and 9-12# MEMS gyroscopes were tested at 60℃ for 23 days.

The test platform consists of E3631A power supply, NI PXLe-1071 data acquisition card, computer, constant temperature and humidity test chamber and multiple data lines, as shown in Figure 4. The E3631A power supply provides 5V voltage for the MEMS gyroscopes. The NI PXLe-1071 acquisition card can collect the output voltage of four MEMS gyroscopes simultaneously. It should be noted that Figure 4 only lists the test environment at normal stress level. The remaining two accelerated stress tests also used the test platform shown in Figure 4, only changing the temperature of the constant temperature and humidity test chamber.

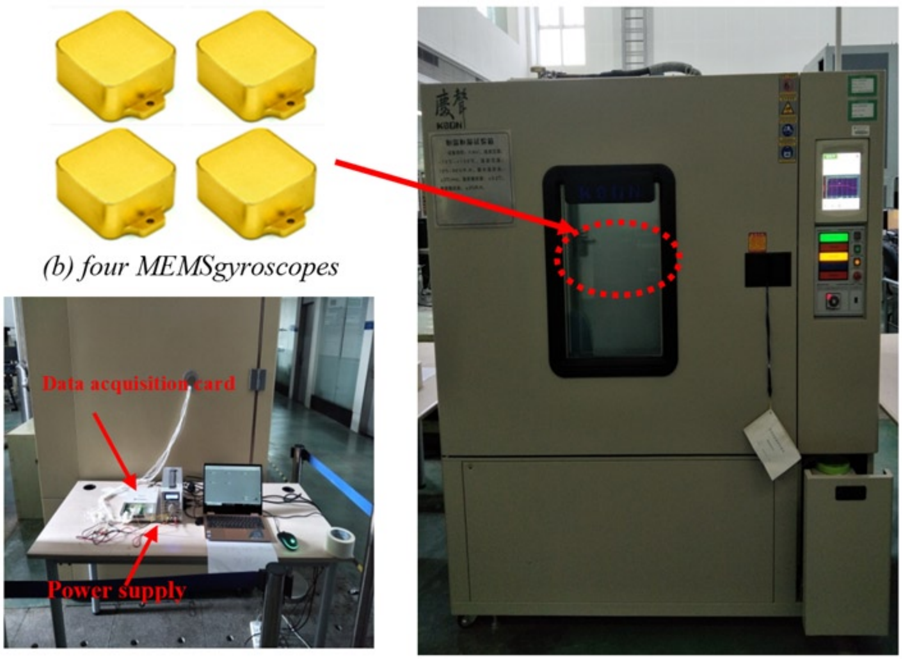


Fig. 4. Test platform of MEMS gyroscope under constant temperature stress

4.2. Analysis of test data

The output voltage of 12 MEMS gyroscopes at three constant stress levels are shown in Figure 5 (a) (b) (c). Allan variance curves at different stress level are shown in Figure 5 (d), all MEMS gyroscopes have line with slopes of −1/2 and 0 at lower correlation time, and have line with slope of 1/2 at different correlation time after the line with slope of 0. It indicates that the angular random walk, bias instability and rate random walk are the main characteristics in the output of MEMS gyroscopes.

In addition, as shown in Figure 5(d), no line with slope of −1 and +1 appear in the Allan variance curve, which indicates that the quantization noise and rate slope are not the main characteristics in the output of MEMS gyroscope. This is because the MEMS gyroscopes have no digital-to-analog conversion process from the output of capacitance to output of voltage, other scholars have same conclusions [12, 21, 26].

4.3. Degradation model

The degradation model of MEMS gyroscope can be established based on the test data and the new method. Step 1. Convert the output voltage of the 12 MEMS gyroscopes to angular rate, and then calculate the root mean square value of the angular rate every day based on the equation (2). Step 2. The bias instability of MEMS gyroscope in every day can be calculated by the least squares method, i.e. equation (6). Step 3. The degradation trend of bias instability in 23 days can be obtained. Step 4. Fit the degradation trend of bias instability with linear function. As shown in Figure 6, bias instability of all MEMS gyroscope increase approximately linearly with time at different temperature stress level. The degradation rates of each gyroscope under different temperature stress levels are shown in Table 1.

So the degradation model of the MEMS gyroscope based on Figure 6 and Table 1 can be expressed as:

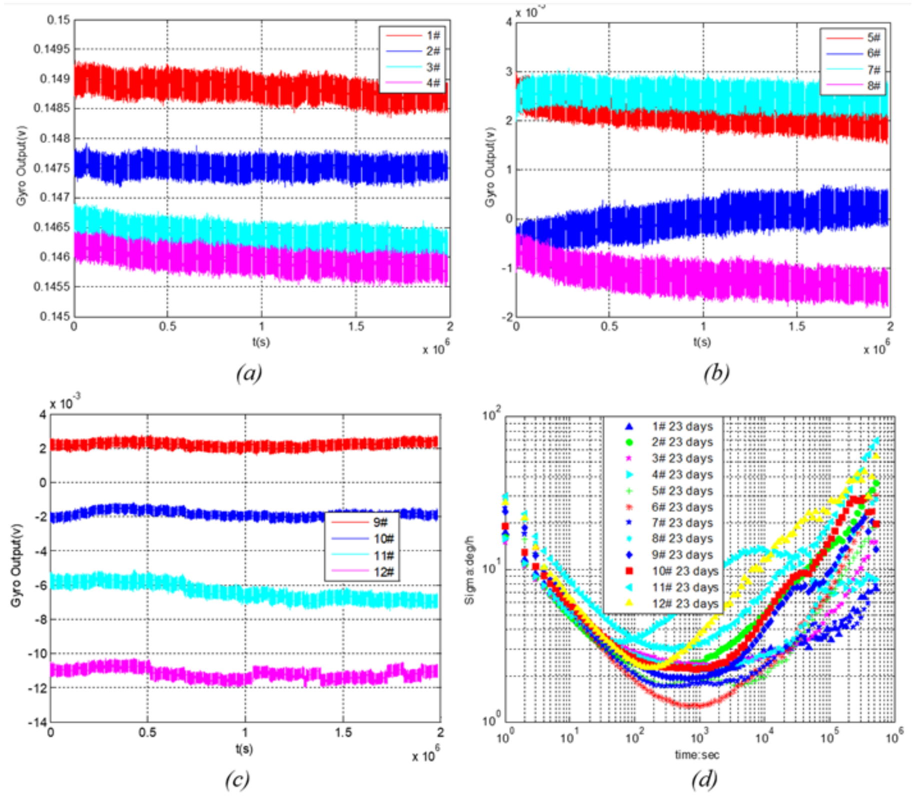


Fig. 5. Voltage output and Allan variance curves of 12 MEMS gyroscopes under different temperature stresses

Table 1. Degradation rate of MEMS gyroscopes' bias instability at different temperature stress levels

Stress level		1#-12#MEMS				Mean
S_0	interception	4.2401	4.2174	4.4450	4.5752	4.3694
	degradation rate	0.0066	0.0132	0.0424	0.0178	0.0200
S_1	interception	1.6059	1.4122	0.8630	1.4004	1.3204
	degradation rate	0.0521	0.0137	0.0222	0.0316	0.0299
S_2	interception	1.5055	2.1970	1.4755	3.2460	2.1060
	degradation rate	0.0406	0.0241	0.0840	0.0507	0.0499

$$B_s(t) = Kt + b \tag{34}$$

where b is the interception of the fitting line.

In addition, Table 1 shows that the degradation rate of bias instability is larger in higher temperature stress. This is consistent with the theoretical analysis results in Section 2.3. The degradation model of the MEMS gyroscope under different stress levels can be established by the mean values of degradation rate and intercept, which can be expressed as respectively:

$$B_s(t) = 0.0200t + 4.3694 \tag{35}$$

$$B_s(t) = 0.0299t + 1.3204 \tag{36}$$

$$B_s(t) = 0.0499t + 2.1060 \tag{37}$$

4.4. Lifetime prediction

The literature shows that the failure threshold B_s^{cir} of the MEMS gyroscope is 25°/h [1-3]. The lifetime of 12 MEMS gyroscopes can be calculated based on equation (34) and Table 1, as shown in Table 2.

So the interception and degradation of the Arrhenius equation can be expressed as:

Table 2. The lifetime of 1#-12# MEMS gyroscope

S_0	MEMS gyroscope	1#	2#	3#	4#	Mean
	lifetime	3145.4	1574.4	484.8	1147.5	1588.0
S_1	MEMS gyroscope	5#	6#	7#	8#	Mean
	lifetime	449	1721.7	1087.3	746.8	1001.2
S_2	MEMS gyroscope	9#	10#	11#	12#	Mean
	lifetime	578.7	946.2	280.1	429.1	558.5

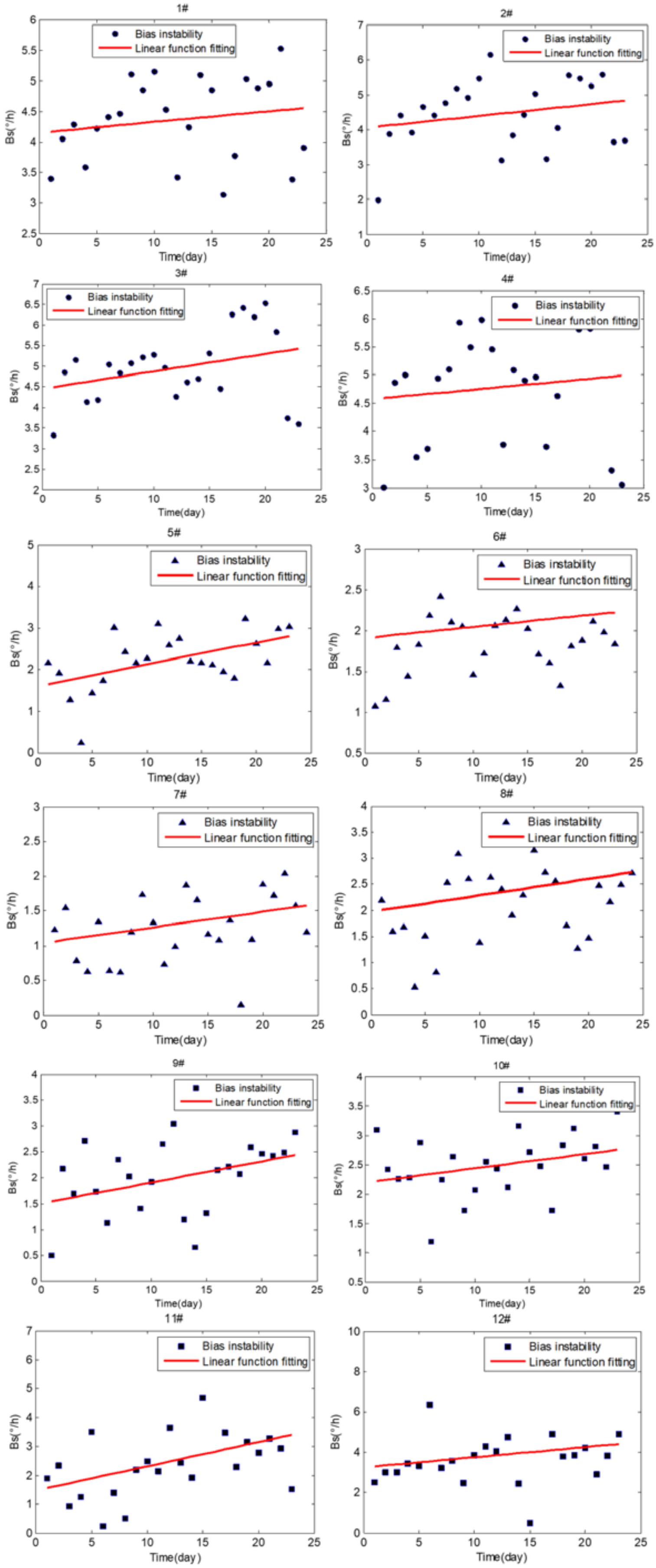


Fig. 6. Bias instability of 12 MEMS gyroscopes changes with time under different temperature stresses

$$\begin{aligned} &\ln A = 4.7868 \\ &\frac{E}{k} = 2592.2 \end{aligned} \tag{38}$$

The parameter A and the activation energy E can be expressed as:

$$\begin{aligned} &A = 119.9148 \\ &E = 0.2234 \end{aligned} \tag{39}$$

So the acceleration factor model between the stress level S_0 and S_2 can be expressed as:

$$\begin{aligned} &AF_{02} = \frac{TF_0}{TF_2} = \frac{\exp(-E/kT_2)}{\exp(-E/kT_0)} = \exp\left(\frac{E}{k}\left(\frac{1}{T_0} - \frac{1}{T_2}\right)\right) \\ &= \exp\left(\frac{0.2234}{8.6171\times 10^{-5}} \times \left(\frac{1}{25+273} - \frac{1}{60+273}\right)\right) = 2.4950 \end{aligned} \tag{40}$$

The lifetime of MEMS gyroscope under normal stress level can be calculated based on the mean lifetime at stress level S_2 and equation (40):

$$TF_{S_0} = TF_{S_2} \bullet AF_{02} = 558.5 \times 2.4950 = 1393.5 \tag{41}$$

The lifetime calculated by acceleration factor model is close to the mean pseudo lifetime (1588 days) under different stress levels, the relative error is 12.25%.

Similarly, Substituting the mean degradation rate of MEMS gyroscope under stress level S_0 and S_1 into equation (34), the Arrhenius equation can be expressed as:

$$\begin{aligned} &\ln(0.0200) = \ln A - \frac{E}{k}\frac{1}{25+273} \\ &\ln(0.0299) = \ln A - \frac{E}{k}\frac{1}{45+273} \end{aligned} \tag{42}$$

So the interception and degradation of the Arrhenius equation can be expressed as:

$$\begin{aligned} &\ln A = 2.4818 \\ &\frac{E}{k} = 1905.4 \end{aligned} \tag{43}$$

The parameter A and the activation energy E can be expressed as:

$$\begin{aligned} &A = 11.9626 \\ &E = 0.1642 \end{aligned} \tag{44}$$

So the acceleration factor model between the stress level S_0 and S_1 can be expressed as:

$$\begin{aligned} &AF_{01} = \frac{TF_0}{TF_1} = \frac{\exp(-E/kT_1)}{\exp(-E/kT_0)} = \exp\left(\frac{E}{k}\left(\frac{1}{T_0} - \frac{1}{T_1}\right)\right) \\ &= \exp\left(\frac{0.1642}{8.6171\times 10^{-5}} \times \left(\frac{1}{25+273} - \frac{1}{45+273}\right)\right) = 1.4950 \end{aligned} \tag{45}$$

The lifetime of MEMS gyroscope under normal stress level can be calculated based on the mean lifetime at stress level S_1 and equation (40):

$$TF_{S_0} = TF_{S_1} \bullet AF = 1001.2 \times 1.4950 = 1496.8 \tag{46}$$

The lifetime calculated by acceleration factor model is close to the mean pseudo lifetime (1588 days) under different stress levels, the relative error is 5.74%.

5. Conclusions

This paper establishes the degradation model and acceleration factor model of MEMS gyroscope based on Allan variance and ADTs, and proposes a universal lifetime prediction method for MEMS gyroscope. The main conclusions are as follows:

The degradation characteristic (bias instability) of MEMS gyroscope is extracted based on Allan variance, and the effect of temperature stress on the degradation rate of bias instability is analyzed. The theoretical analysis shows that the degradation rate of bias instability would increase with the increase of the temperature.

ADTs of MEMS gyroscope are designed and conducted, the experimental results indicate that bias instability of MEMS gyroscope increase approximately linearly with time at different temperature stress level, and the degradation rate of bias instability is larger in higher temperature stress.

Based on the theoretical analysis and experimental results above, the acceleration factor model of MEMS gyroscope under temperature stress is established, and the lifetime of the MEMS gyroscope is predicted. The presented methodology for lifetime prediction of MEMS gyroscope could reduce the testing duration and expense prominently. It may also be used as a reference for lifetime prediction for other MEMS devices.

Acknowledgement

The authors would like to acknowledge the financial supports provided by the National Natural Science Foundation of China under Grant 51875570, the National Key Research and Development Program of China under Grant No.2017YFC0806302, and the National Natural Science Foundation of China under Grant 51705526.

References

1. Analog devices, ADXRS300 Data sheet, 2004. http://www.analog.com/static/imported -files/data_sheets/ADXRS300.pdf.
2. Analog devices, ADXRS646 Data sheet, 2011. https://www.analog.com/cn/products /adxrs 450.html.
3. Analog devices, ADXRS646 Data sheet, 2012. https://www.analog.com/cn/products/ adxrs646.html#product-overview.
4. Chateauneuf, and Alaa. Accelerated life testing and degradation modeling. Reliability Engineering System Safety 2014; 131: 228, https://doi.org/10.1016/j.ress.2014.05.004.
5. Chen Wenhua, Liu Juan, Gao Liang, et al. Accelerated degradation reliability modeling and test data statistical analysis of aerospace electrical connector. Chinese Journal of Mechanical Engineering 2011; 24(6): 957-962, https://doi.10.3901/ CJME2011.06.957.
6. Chien-Yu Peng and Sheng-Tsaing Tseng. Progressive-stress accelerated degradation test for highly-reliable products. IEEE Transactions on Reliability 2010; 59(1):30-37, https://doi. 10.1109/TR.2010.2040769.

7. Chikovani V V, Kyiv, Ukraine. Performance parameters comparison of ring laser, coriolis vibratory and fiber-optic gyros based on Allan variance analysis. IEEE 2nd International Conference Actual Problems of Unmanned Air Vehicles Developments. Proceedings 2013; 153-156, https://doi.org/10.1109/APUAVD.2013.6705312.
8. Chun-Lin Lu, Meng-Kao Yeh. Thermal stress analysis for a CMOS-MEMS microphone with various metallization and materials. Microelectronic Engineering 2019; 213: 47–54, https://doi.org/10.1016/j.mee.2019.04.013.
9. El-Sheimy N, Hou H, Niu X, et al. Analysis and modeling of inertial sensors using allan variance. IEEE Transactions on instrumentation and measurement 2007; 57(1): 140-149, https://doi.org/10.1109/tim.2007.908635.
10. Fu-Kwun Wang, Tao-Peng Chu. Lifetime predictions of LED-based light bars by accelerated degradation test. Microelectronics Reliability 2012; 52(7):1332–1336, https://doi. 10.1016/j.microrel.2012.02.019.
11. Grantham B E, Bailey M A. A least-squares normalized error regression algorithm with application to the Allan variance noise analysis method, In: Proceedings of IEEE/ION PLANS 2006; 750-755, https://doi.org/10.1109/PLANS.2006.1650671.
12. Heera Mallik M, Divya Jyothi K, Mithun S Varma, Divya Rao A. Agrawal. Minimum Variance Optimal Filter Design for a 3x3 MEMS Gyroscope Cluster Configuration. IFAC 2016; 49.1: 639-645, https://doi.org/10.1016/j.ifacol.2016.03.128.
13. Jacopo Iannacci. Reliability of MEMS: A perspective on failure mechanisms, improvement solutions and best practices at development level. Displays 2015; 37: 62-71, https://doi.org/10.1016/j.displa.2014.08.003.
14. Jianbin Su, Dingbang Xiao, Xiong Wang, Zhihua Chen, Xuezhong Wu. Vibration sensitivity analysis of the 'Butterfly-gyro' structure. Microsystem Technologies 2014; 20: 1281-1290, https://doi.org/10.1007/s00542-013-1913-x.
15. Jin-Won Joo and Sung-Hoon Choa. Deformation behavior of MEMS gyroscope sensor package subjected to temperature change. IEEE Transactions on Components and Packaging Technologies 2007; 30.2: 346-354, https://doi.org/10.1109/TCAP T.2007.897948.
16. Kim D, Thomas R, and Closkey M. Spectral analysis of vibratory gyro noise. IEEE tranc. Sensors 2013; 13(11): 4361-4374, https://doi.org/10.1109/JSEN.2013.2269797.
17. Kunsong Lin, Yunxia Chen, Dan Xu. Reliability assessment model considering heterogeneous population in a multiple stresses accelerated test. Reliability Engineering System Safety 2017; 165: 134-143, https://doi.org/10.1016/j. ress.2017.03.013.
18. Lawrence C N, Darryll J P. Characterization of ring laser gyro performance using the Allan variance method. Journal of Guidance, Control, and Dynamics 1997; 20 (1): 211-214, https://doi.org/10.2514/2.4026.
19. Liang Xue, Cheng-Yu Jiang, Hong-Long Chang, Yong Yang, Wei Qin, Wei-Zheng Yuan. A novel Kalman filter for combining outputs of MEMS gyroscope array. Measurement 2012; 45: 745-754, https://doi.org/10.1016/j.measurement.2011.12.016.
20. Lu X, Chen X, Wang Y, Tan Y. Consistency analysis of degradation mechanism in step-stress accelerated degradation testing. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (2): 302-309, http://dx.doi.org/10.17531 /ein.2017.2.19.
21. Miroslav Matejček, Mikulaš Šostronek. New Experience with Allan Variance Noise analysis of accelerometers. October 2017; https://doi.org/10.23919/KIT.2017.8109457.
22. Mulloni V, Lorenzelli L, Margesin B, Barbato M, Meneghesso G. Temperature as an accelerating factor for lifetime estimation of RF-MEMS switches. Microelectronic Engineering, 2016; 160: 63-67, https://doi.org/ 10.1016/j.mee.2016.03.023.
23. Nelson W, Accelerated testing: statistical models, test plans, and data analysis. John Wiley & Sons: New York, 1990; NY. 10.1002/9780470316795.
24. Ningfang S, Yuan R, Jin J. Autonomous estimation of Allan variance coefficients of onboard fiber optic gyro. Journal of Instrumentation 2011; 6.09: P09005-P09005, https://doi.org/10.1088/1748-0221/6/09/P09005.
25. Oliveira V R B, Colosimo E A. Comparison of methods to estimate the time-to-failure distribution in degradation tests. Quality Reliability Engineering International 2004; 20(4): 363-73, https://doi.org/10.1002/qre.567.
26. Pin Lv, Jianye Liu, Jizhou Lai, Kai Huang. Allan variance method for gyro noise analysis using weighted least square algorithm. Optik 2015; 126: 2529-2534, https://doi.org/10.1016/j.ijleo.2015.06.044.
27. ReliaSoft Co. Accelerated Life Testing Reference. Tucson, AZ: ReliaSoft Publishing, 2014, https://www.ReliaSoft.com.
28. Richard J V, Ahmed S K. Statistical modeling of rate gyros. IEEE Transactions on Instrumentation and Measurement 2012; 61: 673-684, https://doi.org/10.1109/tim.2011.2171609.
29. Saeedivahdat A, Abdolkarimzadeh F, Feyzi A, Rezazadeh G, Tarverdilo S. Effect of thermal stresses on stability and frequency response of a capacitive microphone. Microelectronics Journal 2010; 41: 865-873.
30. Si X S, Wang W B, Hu C H, Zhou D H, Pecht M G. 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.
31. Songhua Hao, Jun Yang, Christophe Berenguer. Nonlinear step-stress accelerated degradation modeling considering three sources of variability. Reliability Engineering System Safety 2018; 172: 207-215, https://doi.org/10.1016/j.ress.2017.12.012.
32. Tabata O, Tsuchiya T. Reliability of MEMS: testing of materials and devices. Weinheim, Wiley, 2006, https://doi.org/10.1007/s10800-008-9482-x.
33. Wang YS, Fang X, Zhang CH, Chen X, Lu JZ. Lifetime prediction of self-lubricating spherical plain bearings based on physics-of-failure model and accelerated degradation test. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18(4): 528-538, https://doi.org/10.17531/ein.2016.4.7.
34. Wang Y, Zhang C, Chen X, Tan Y. Lifetime prediction method for electron multiplier based on accelerated degradation test. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 484-490.
35. Yang Z, Chen Y X, Li X F, Zio E, and Kang R. Smart electricity meter reliability prediction based on accelerated degradation testing and modeling. Electrical Power and Energy Systems 2014; 56: 209-219, https://doi.org/10.1016/j.ijepes.2013.11.023.
36. Zhiqiang X and Gebre-Egziabher D. Modeling and bounding low inertial sensor errors. In proc. IEEE/ION Position, Location and Navigation Symposium 2008: 1122-1132, https://doi.org/10.1109/PLANS.2008.4569999.

Yao LIU
Yashun WANG
Zhengwei FAN
Laboratory of Science and Technology on Integrated Logistics Support
College of Intelligence Science and Technology
National University of Defense Technology
Yanwachi str., Changsha, 410073, Hunan, China

Zhanqiang HOU
College of Intelligence Science and Technology
National University of Defense Technology
Yanwachi str., Changsha, 410073, Hunan, China

Shufeng ZHANG
Xun CHEN
Laboratory of Science and Technology on Integrated Logistics Support
College of Intelligence Science and Technology
National University of Defense Technology
Yanwachi str., Changsha, 410073, Hunan, China

E-mails: yaoliu133@126.com, wangyashun@nudt.edu.cn,
fanzhengwei15@nudt.edu.cn, houzhanqiang@nudt.edu.cn,
sfzhang@nudt.edu.cn, chenxun@nudt.edu.cnj237

Article citation info:
CZAPP S, SZULTKA S, RATKOWSKI F, TOMASZEWSKI A. Risk of power cables insulation failure due to the thermal effect of solar radiation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 232–240, http://dx.doi.org/10.17531/ein.2020.2.6.

Stanisław CZAPP
Seweryn SZULTKA
Filip RATKOWSKI
Adam TOMASZEWSKI

RISK OF POWER CABLES INSULATION FAILURE DUE TO THE THERMAL EFFECT OF SOLAR RADIATION

RYZYKO USZKODZENIA CIEPLNEGO IZOLACJI KABLI ELEKTROENERGETYCZNYCH Z POWODU ODDZIAŁYWANIA PROMIENIOWANIA SŁONECZNEGO

Low-voltage, as well as high-voltage power cable lines, are usually buried in the ground. The ampacity of the power cables in the ground mainly depends on the thermal resistivity of the soil, which may vary in a wide range. A common practice in power cable systems performance is to supply them from a pole of an overhead line. If so, a section of the line is located in free air and can be directly exposed to solar radiation. In some cases, the ampacity of power cables placed in free air is lower than in the ground. Differences in ampacities can be very high if thermal resistivity of the soil is very low, and simultaneously solar irradiation of cables in air occurs. This paper presents the risk of power cables overheating and in consequence the risk of their failure, when part of the underground power cable line is placed in free air. Temperature distribution of cables in the air (with and without solar radiation) for various load currents is presented. Thermal endurance of power cables insulation, operating with the overheating, is estimated.

Keywords: *failure risk, heat transfer, numerical modelling, power cables, solar radiation.*

Linie kablowe zarówno niskiego, jak i wysokiego napięcia zwykle buduje się jako podziemne. Obciążalność kabli układanych w ziemi w znacznym stopniu zależy od rezystywności cieplej gruntu, a może się ona zmieniać w bardzo szerokim zakresie. Obecnie powszechną praktyką jest zasilanie linii kablowych z linii napowietrznych, co sprawia, że pewien odcinek linii kablowej znajduje się w powietrzu i może być poddany bezpośredniemu oddziaływaniu promieniowania słonecznego. W pewnych przypadkach obciążalność prądowa długotrwała kabli w powietrzu jest niższa niż w ziemi – różnice w tej obciążalności mogą być bardzo duże, jeżeli grunt ma niską rezystywność cieplną, a na odcinek linii w powietrzu oddziałuje promieniowanie słoneczne. W artykule przedstawiono problem przegrzania kabli elektroenergetycznych, gdy przyjęta obciążalność linii kablowej wynika z warunków dla ułożenia w ziemi, a na pewnym odcinku linia jest umieszczona w powietrzu. Przedstawiono rozkłady temperatury kabli w powietrzu (z uwzględnieniem i bez uwzględnienia promieniowania słonecznego) dla różnych prądów obciążenia kabli. Oszacowano trwałość termiczną izolacji kabli, mających przez znaczny przedział czasu temperaturę wyższą niż dopuszczalna długotrwale.

Słowa kluczowe: *ryzyko uszkodzenia, wymiana ciepła, modelowanie numeryczne, kable elektroenergetyczne, promieniowanie słoneczne.*

1. Introduction

Distribution of power in power networks is performed with the use of overhead power lines as well as underground power cables. The investment cost of the underground power cable distribution systems is higher compared to the use of overhead lines but gives higher reliability of supply, especially reflected in improved SAIDI and SAIFI indicators [1, 19, 24].

Power cables are usually buried in the ground, but in many cases, their ending sections are placed in air, to be connected with conductors of overhead lines, as it is presented in Fig. 1. Depending on the height of the pole, length of the power cables in air can be from a few to several meters. Given that the cable section in the air is connected in series with a section buried in the ground, the ampacity of the whole power cable line depends on the section for which thermal condition for heat transfer from the cables is the worst. The worst thermal condition is expected for the section in air, during sunny weather and without any wind.

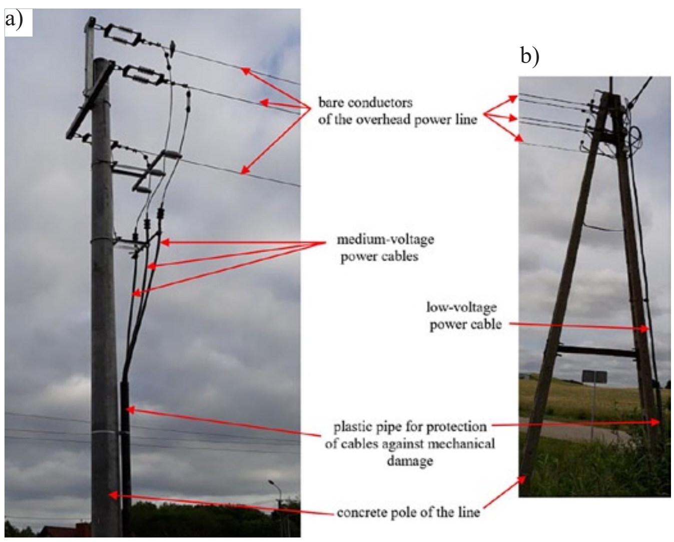


Fig. 1. Poles for: a) medium-voltage overhead power line, b) low-voltage overhead power line; and their connections with power cable lines

The problem of power cables heating and calculation of their ampacity are the subject of many papers and standards, especially [12-

16]. Knowledge of the actual operating conditions of power cables helps to avoid design errors and, as a result, it may increase the reliability and safety of power installations [2, 17, 26].

The effect of sunlight on the heating of power cables is not fully studied, and the provisions of the standards do not fully describe this effect [4, 6-8, 18, 25-26]. Paper [26] clearly indicates that cables exposed to solar radiation may be damaged very fast. In the described installation (Fig. 2), power cables were put into operation during autumn. In spite of a very low value of the load current, their first damage occurred during the nearest summer. After this summer, power cables operated without problems, but their thermal damage returned during the next summer.

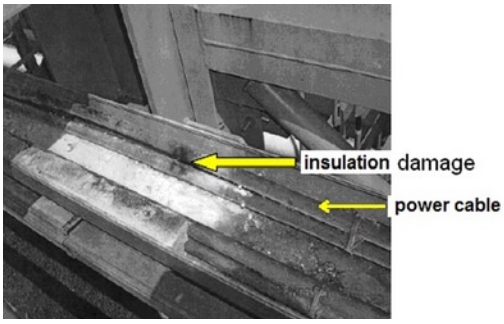


Fig. 2. Thermal damage of the power cables exposed to solar radiation [26]

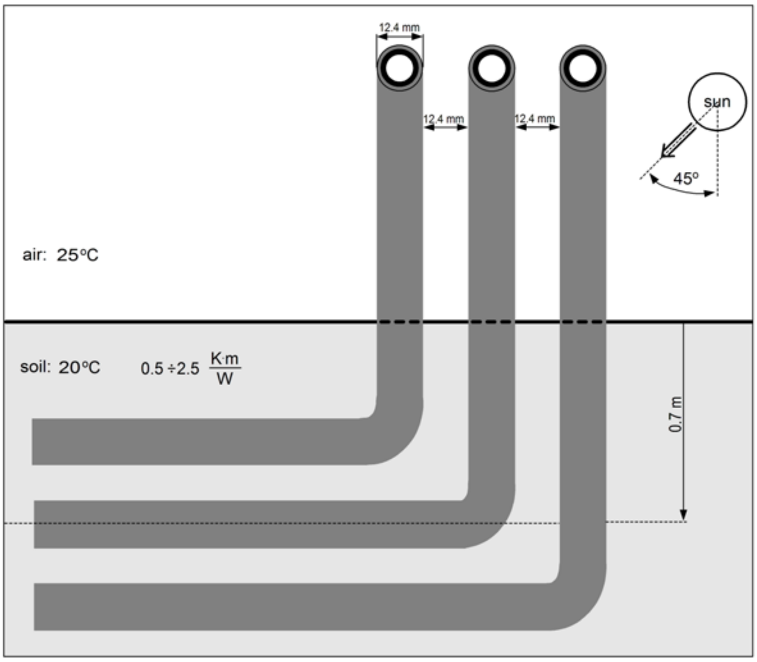


Fig. 3. Analyzed arrangement of the power cable system composed of three single-core cables

Accurate calculation of power cables ampacity and temperature of the insulation for various ambient conditions are possible only with the use of the numerical approach [8-10, 13, 27]. This paper presents the problem of evaluation of the power cables ampacity and their insulation temperature when cables are placed in changing ambient condition “ground-air”, and especially solar radiation may occur. The common practice is to calculate the ampacity taking into account thermal conditions occurring in the ground. Unfortunately, such an assumption may be dangerous for the part of the cable line in free air. In this paper, the authors prove that thermal condition in the air can give a negative effect on ampacity and endurance of the whole power cable line, in particular when solar radiation occurs. It is very important in terms of the reliability of power supply, because, due to significant overheating of cables in air, fast destruction of their insulation may appear. For calculation of the ampacity and insulation temperature, advanced computer modelling is employed.

A power cable line presented in Fig. 3 is considered in this paper. This cable line is placed partially in the ground and partially in free air (section in the air goes from the ground to a pole of the overhead line, as it is presented in Fig. 1). The power cable line is composed of three single-core PVC-insulated cables (maximum permissible continuous operating temperature is equal to 70°C). The power cables nominal cross-sectional area of the copper conductor is equal to 35 mm^2. They are laid in flat formation (0.7 m from the ground surface), the spacing between cables is equal to their external diameter. Ambient air temperature is 25°C and soil temperature is 20°C (reference ambient conditions for Poland according to IEC 60287-3-1 [16]). Thermal resistivity of the soil is considered to be within the range $\rho_s = (0.5–2.5)$ (K·m)/W.

The main purpose of the investigation is to calculate ampacity of power cable line for various thermal resistivities of the soil, and after assuming the calculated ampacity as a power cable load, temperature of the cable’s insulation in free air is evaluated – with and without solar radiation. On the base of the calculated temperature of the cables in air, a decrease in their thermal endurance is evaluated.

2. Calculation of power cables ampacity

For a given value of current flowing in a core of a power cable, heat balance in a steady-state can be described by the following equation:

$$q_c + q_r = q_s + q_J \tag{1}$$

where:

- q_c – heat flux density dissipated to the surroundings of power cables; by convection (cables in the air) or by conduction (cables in the ground), W/m^2,
- q_r – heat flux density dissipated to the surroundings of power cables by radiation, W/m^2,
- q_s – heat flux density delivered to the power cables by solar radiation, W/m^2,
- q_J – heat flux density generated in the conductor due to the flow of electric current (Joule’s heat), W/m^2.

In the case of cables laid in the ground, heat flux density q_c mainly depends on the value of thermal resistivity of the soil. The lower value of thermal resistivity of the soil, the higher value of the ampacity of power cables [5]. In the case of cables laid in free air, heat flux density q_c mainly depends on the value of convective heat transfer coefficient α. This coefficient is directly related to the speed and direction of the wind flowing around the cables as well as the temperature difference between the cable and air. Therefore, in the case of sunny, windless weather, convective heat transfer coefficient reaches small values. Additional heat flux may occur for cables in free air – heat flux density q_s generated by solar radiation. It all makes that in the case of a cable line placed in series in the ground of high thermal conductivity (low thermal resistivity) and in air, where there is slight heat exchange by convection and there is sunlight, risk of overheating of the power cables may occur.

For the purpose of power cable systems projects, calculation of power cables ampacity is usually performed with the use of IEC 60287 [14, 15] provisions. These IEC standards allow to evaluate ampacity of power cables placed in various configurations as well as ambient conditions, especially placed in the ground and in air.

The ampacity of an underground power cable, when drying out of the soil is excluded, can be calculated as follows [14, 15]:

$$I_{\max}=\sqrt{\frac{\Delta\theta-W_{\mathrm{d}}\cdot\left[0.5\cdot T_1+n_{\mathrm{c}}\cdot(T_2+T_3+T_4)\right]-\Delta\theta_{\mathrm{add}}}{R\cdot T_1+n_{\mathrm{c}}\cdot R\cdot(1+\lambda_1)\cdot T_2+n_{\mathrm{c}}\cdot R\cdot(1+\lambda_1+\lambda_2)\cdot(T_3+T_4)}} \quad (2)$$

where:

$I_{\max}$ – ampacity of a power cable, A,

$\Delta\theta$ – maximum permissible temperature rise of the conductor above ambient temperature, K,

R – AC current resistance of a conductor at its maximum permissible temperature, Ω/m,

W_{d} – dielectric losses per phase, W/m,

T_1 – thermal resistance (per core/phase) between the conductor and sheath/insulation, (K·m)/W,

T_2 – thermal resistance between the sheath/insulation and armour, (K·m)/W,

T_3 – thermal resistance of external serving of the cable, (K·m)/W,

T_4 – external thermal resistance of surrounding medium (soil/backfill), (K·m)/W,

n_{c} – number of conductors in a multicore power cable, -,

λ_1 – ratio of the total losses in metallic sheaths (if any) to the total conductor losses, -,

λ_2 – ratio of the total losses in metallic armour (if any) to the total conductor losses, -.

$\Delta\theta_{\mathrm{add}}$ – additional reducing factor of the maximum permissible temperature rise of the conductor above ambient temperature, K, (for a cable directly buried in the ground $\Delta\theta_{\mathrm{add}}=0$).

Calculation of ampacity of power cables according to IEC 60287 [14, 15] provisions can be performed with the use of CYMCAP software [3]. Table 1 presents the result of this calculation, and Fig. 4 depicts the distribution of the temperature in the ground for three selected thermal resistivities of the soil ρ_s: 0.5, 1.0 and 2.5 (K·m)/W.

One can see that the ampacity of the analyzed power cable line (part in the ground) strictly depends on the thermal resistivity of the soil. It may vary almost twice if this resistivity changes from 2.5 (K·m)/W to 0.5 (K·m)/W, and it is very high for the latter value. Thus, for the safe operation of the underground power cable line with a section placed in air, it is important to evaluate the temperature of the power cable in this section, especially if direct solar radiation may occur. If the temperature exceeds permissible 70°C, it is necessary to reduce permissible load of the part of the cable line in the ground.

Including solar radiation in the calculation of power cables temperature and ampacity is not easy. Methods of power cables ampacity calculation, included in standards IEC 60287 [14, 15], utilize Neher-McGrath assumptions but are characterized by simplifications. For more complicated cases of cables arrangement, e.g. in case of strong solar radiation, and especially in case of the mixed effect of solar radiation and wind, it is no possible to calculate the ampacity with sufficient accuracy. Therefore, the ampacity and temperature of the

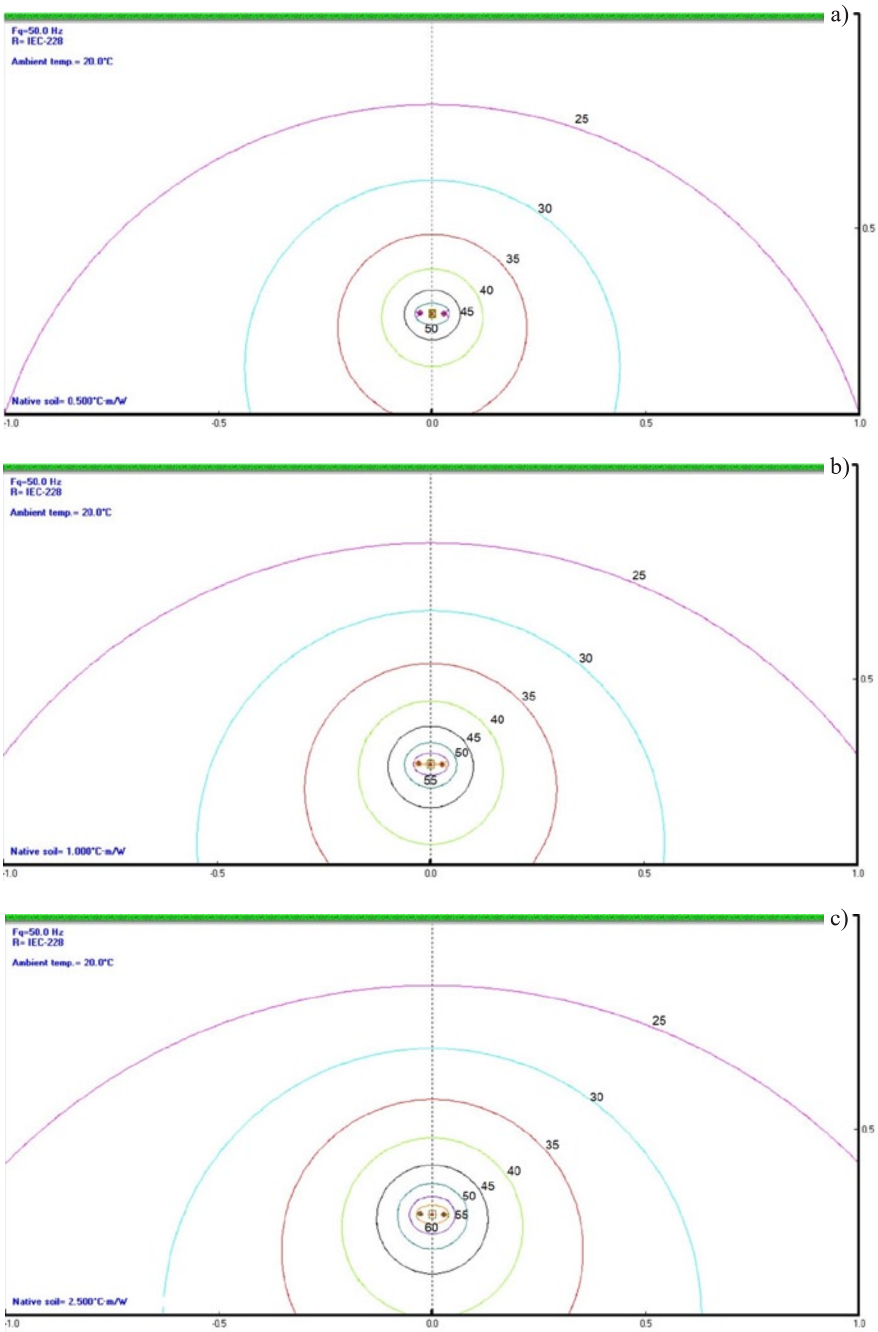


Fig. 4. Temperature distribution (°C) around the analyzed three power cables in flat formation, directly buried in the ground at the depth 0.7 m, for the following load current I_{load} and thermal resistivity of the soil ρ_s:

a) I_{load} = 230 A, ρ_s = 0.5 (K·m)/W, b) I_{load} = 176 A, ρ_s = 1.0 (K·m)/W, c) I_{load} = 118 A, ρ_s = 2.5 (K·m)/W

Table 1. Ampacity of the analyzed power cable line calculated with the use of CYMCAP software

soil, (K·m)/W	0.5	1.0	1.5	2.0	2.5
ampacity, A	230	176	148	130	118

analyzed power cable line are evaluated with the use of the advanced numerical modelling, what is presented in the next section.

3. Modelling of the thermal condition of power cables in free air

In order to investigate the thermal processes in the cable line placed in free air, a computational fluid dynamics implemented in Ansys software has been used. Steady-state fluid flow has been modelled and then heat exchange calculations have been performed. The 3D computational domain is presented in Fig. 5. It consists of three power cables surrounded by air. The total number of the finite elements in the numerical grid exceeds 16×10^6.

a) b)

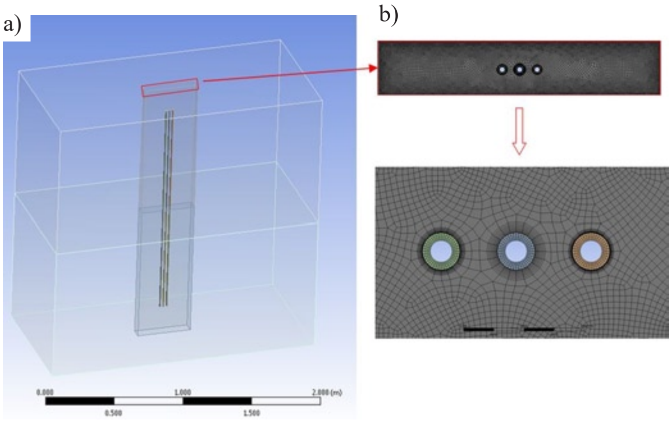
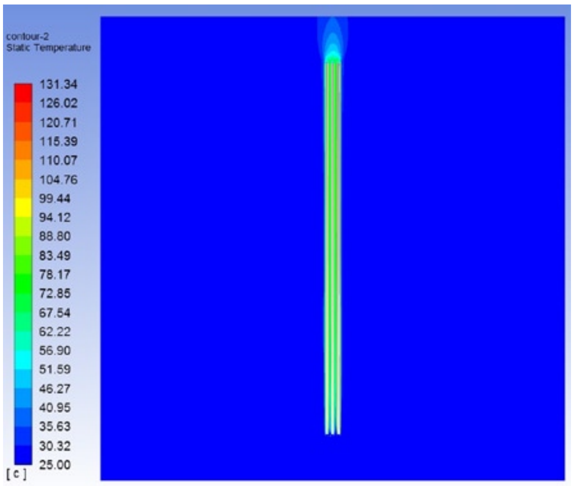
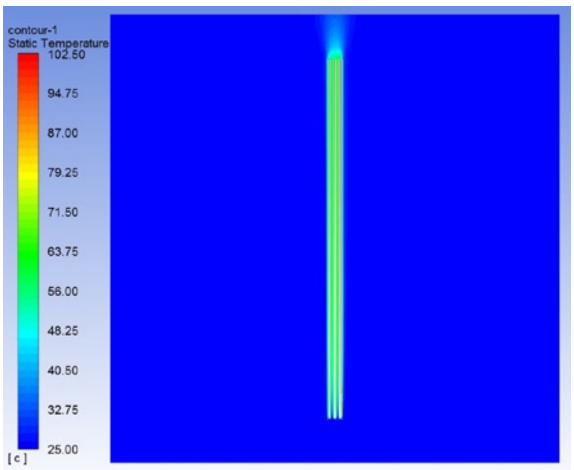


Fig. 5. Computational domain of the analyzed cable system (a) and density of the numerical grid (b)

a)

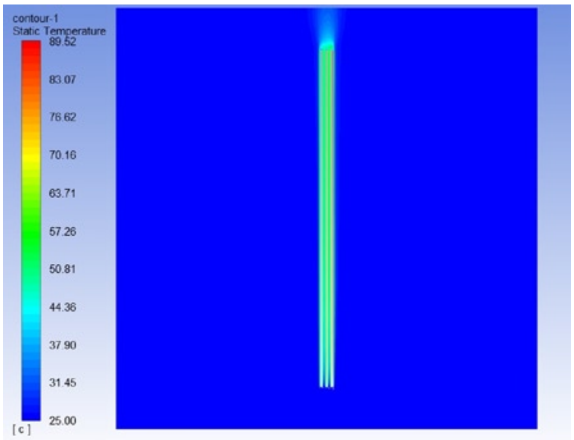


b)

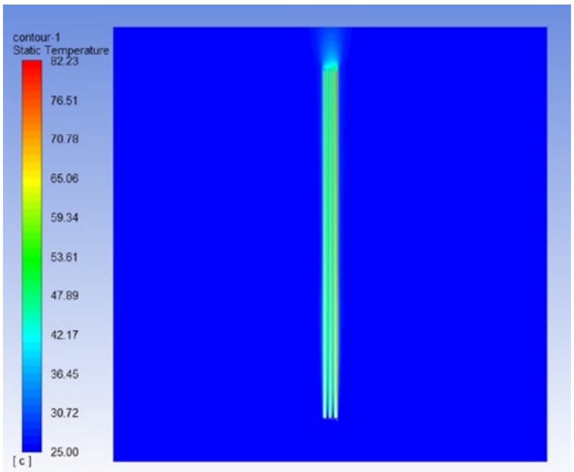


Heat is introduced to the cables from two different sources: from the electric current to internal surface of the cable insulation and also from solar radiation to half of the external surface of the cable (at the right side – see Fig. 3). The lack of accurate measurement data regarding solar radiation incident on the surface of the earth [23] caused that the heat flux density supplied from sun radiation is calculated for the sun's altitude of 45°, which is consistent with conditions in the

c)



d)



e)

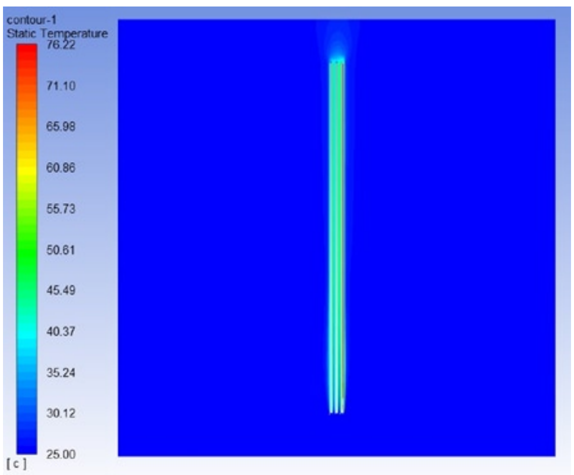
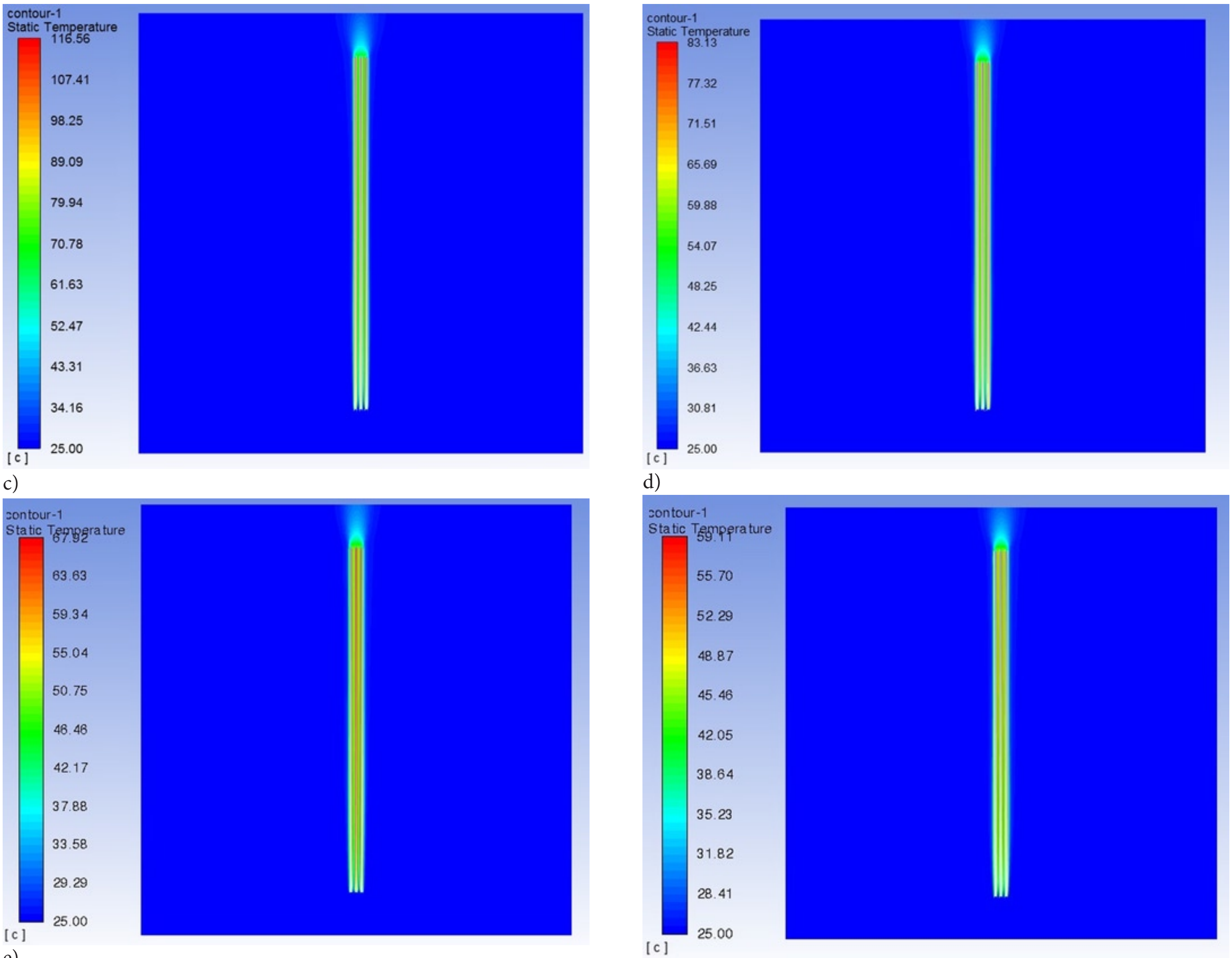


Fig. 6. Temperature distribution around the analyzed three power cables (with solar radiation) for the following load current I_{load}:
a) I_{load} = 230 A, max insulation temp. 131.34 °C,
b) I_{load} = 176 A, max insulation temp. 102.50 °C,
c) I_{load} = 148 A, max insulation temp. 89.52 °C,
d) I_{load} = 130 A, max insulation temp. 82.23 °C,
e) I_{load} = 118 A, max insulation temp. 76.22 °C

a) b) c) d)



e)

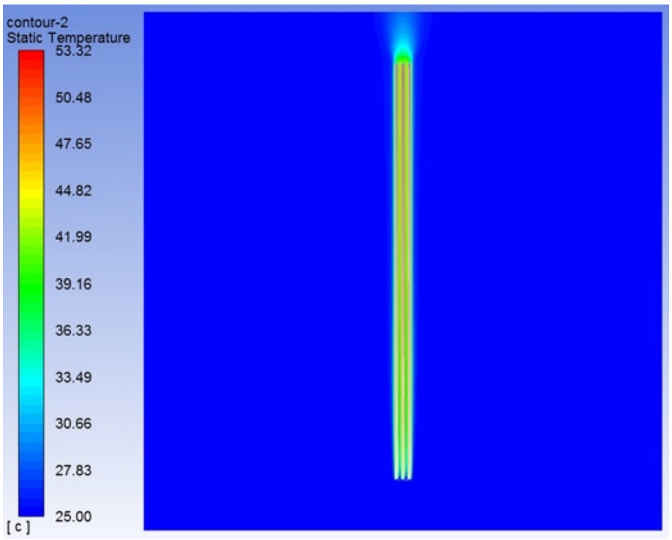


Fig. 7. Temperature distribution around the analyzed three power cables (without solar radiation) for the following load current I_{load}:

a) I_{load} = 230 A, max insulation temp. 116.56°C
b) I_{load} = 176 A, max insulation temp. 83.13°C
c) I_{load} = 148 A, max insulation temp. 67.92°C
d) I_{load} = 130 A, max insulation temp. 59.11°C
e) I_{load} = 118 A, max insulation temp. 53.32°C

European region during summer period. The external (ambient) air temperature is 25°C. In this case, heat exchanges between power cables and air by natural convection and by thermal radiation. In calculations, the case with no wind is analyzed, in order to show the worst thermal conditions, which may appear during the year. Therefore, the presence of gravity force has to be included, and air density is described by ideal gas law. The turbulence model chosen in calculations is *k-epsilon* standard with the option of full buoyancy effects enabled. The thermal radiation of cables external surface was modelled by Discrete Ordinates (DO) model.

Figures 6 and 7 summarizes the temperature distributions for power cables placed in air (with solar radiation – Fig. 6; without solar radiation – Fig. 7). For these cables, the current in the conductor was assumed as resulting from the ampacity of the cables placed in the ground (see Tab. 1). In the upper left corner of each subfigure in Figs 6 and 7, the maximum temperature of the cable insulation is displayed.

When solar radiation is taken into account, the maximum temperature of the insulation varies from 76.22°C (Fig. 6e) to 131.34°C (Fig. 6a), for cables load from 118 A to 230 A respectively. One can see that in each case the temperature exceeds the permissible level of 70°C. For cases without solar radiation and the mentioned range of the cables load, the maximum temperature of the insulation varies from 53.32°C (Fig. 7e) to 116.56°C (Fig. 7a).

Figure 8 contains examples of air velocity distributions around the power cables. The whole velocity field is the effect of natural convec-

Table 2. Overheating of the section of the power cable line placed in the air, for load current equal to the ampacity of the power cables buried in the ground

Cables							Air/ground percentage	
in the ground			in air					
Thermal resistivity of the soil	Ampacity	Max cables temp.	Load current	Solar radiation	Max cables temp.	Load current giving max temp. 70 °C	temp.	ampacity
(K·m)/W	A	°C	A	-	°C	A	%	%
0.5	230	70	230	no	117	152	167	66
				yes	131	96	187	42
1.0	176	70	176	no	83	152	118	86
				yes	102	96	145	54
1.5	148	70	148	no	68	152	97	103
				yes	90	96	129	65
2.0	130	70	130	no	59	152	84	117
				yes	83	96	119	74
2.5	118	70	118	no	53	152	76	129
				yes	76	96	108	81

tion. When the load is 230 A and solar radiation is taken into account, the maximum air velocity around the cables is 0.98 m/s. For the load equal to 118 A and when solar radiation is not taken into account, the maximum air velocity around the cables is only 0.60 m/s. Thus, it is very important to leave the free air movement along the cables. Otherwise, the temperature of the cable's insulation, especially in the presence of solar radiation, could rise even more.

When comparing load current giving max temp. 70°C of power cables in free air with ampacities of power cables in the ground, it may conclude that the latter is higher – in some cases significantly higher (Tab. 2). Aggregated results included in Tab. 2 show that for thermal resistivities of the soil ρ_s = 0.5 and 1.0 (K·m)/W the ampacity of the power cables in the ground (230 A and 176 A respectively) exceeds the ampacity of the power cables in free air even if no solar radiation occurs (152 A without solar radiation and 96 A with solar radiation).

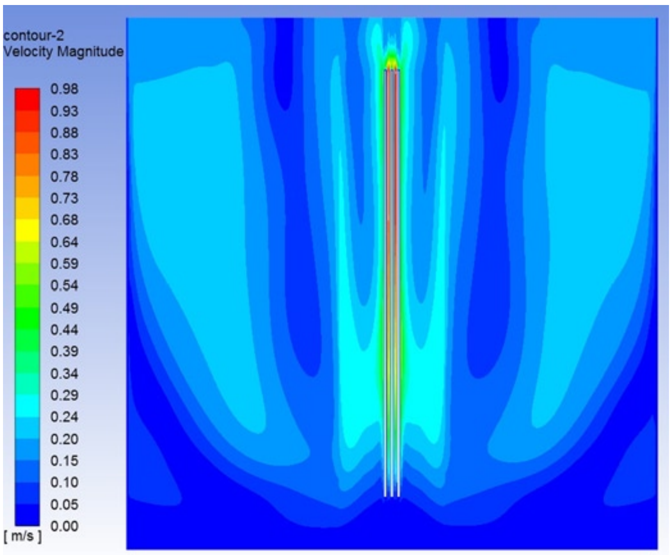
The worst case is for thermal resistivity of the soil equal to 0.5 (K·m)/W. In the case of load current equal to 230 A (ampacity in the ground) temperature of the cables insulation in the air is 117 °C without solar radiation and 131°C with solar radiation, what exceeds the permissible level (70°C) by 67% and by 87% respectively. When solar radiation occurs, PVC insulation of power cables in free air may be overheated (76°C) even if thermal resistivity of the soil is relatively high (e.g. 2.5 (K·m)/W). While the ampacity in the ground is equal to 118 A, the ampacity in the air is equal to the aforementioned only 96 A.

High cables ampacity in the ground (due to the low resistivity of the soil) may lead to strong overheating of power cable line situated in free air (during insolation), and this may lead to noticeable decrease its thermal endurance.

4. Thermal endurance of the power cables insulation

Exceeding the permissible temperature specified for a particular type of a power cable insulation for a long time causes a decrease in its designed endurance, according to the exponential relationship described by the Arrhenius curve [11, 22, 28].

a)



b)

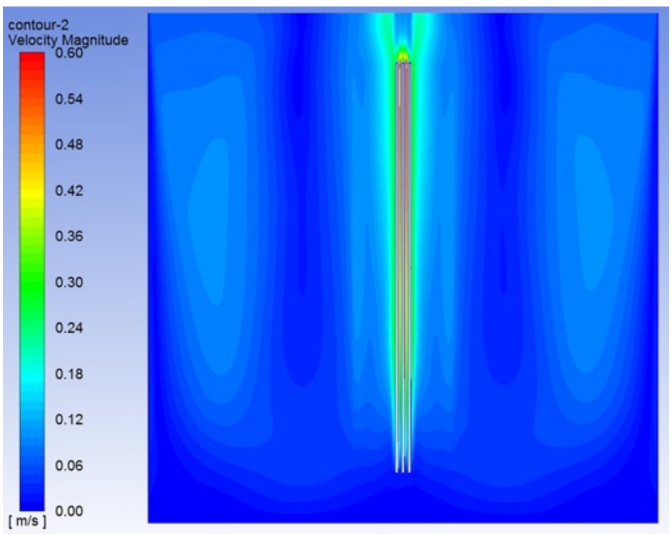


Fig. 8. Air velocity distribution around the analyzed three power cables for the following load current I_{load}: a) I_{load} = 230 A, with sun, max velocity 0.98 m/s, b) I_{load} = 118 A, without sun, max velocity 0.60 m/s

Table 3. Thermal endurance (at given overload and resultant endurance) of the power cables of PVC insulation placed in free air

Cables							
in the ground		in air					
Thermal resistivity of the soil	Ampacity	Solar radiation	Load current giving max temp. 70°C	Ground/air percentage ampacity	Cable overload	Thermal endurance	
						at given overload*	resultant
(K.m)/W	A	-	A	%	%	years	
0.5	230	no	152	151	51	0.24	2.09
		yes	96	240	140	0.00011	0.00061
1.0	176	no	152	116	16	5.3	15.5
		yes	96	183	83	0.02	0.08
1.5	148	no	152	97	0	20	20
		yes	96	154	54	0.19	0.99
2.0	130	no	152	86	0	20	20
		yes	96	135	35	0.96	4.34
2.5	118	no	152	78	0	20	20
		yes	96	123	23	2.86	9.54

** "at given overload" – it is assumed that cables insulation temperature results (all the year) from the ampacity given for the cables buried in the ground*

Usually, the continuous operating temperature of the power cables is selected so that the rated thermal endurance of the insulation is around 20–30 years. In this paper, it is assumed that this endurance is E_{rat} = 20 years (PVC insulation, maximum permissible continuous operating temperature 70°C). For cables with PVC insulation it was estimated that 1 hour of the operation at the 20% overload ($1.2I_{max}$) corresponds to 5 hours of the operation at maximum permissible continuous operating temperature, and 1 hour of the operation at the 45% overload ($1.45I_{max}$) reflects 50 hours of the operation at the aforementioned temperature [22].

Taking the above into account, thermal endurance of the cable insulation for the overloaded cable can be determined by the relationship:

$$E_{ins} = 20.976 \cdot \exp(-0.087 \cdot ol) \quad (3)$$

where:

E_{ins} – thermal endurance of the cable insulation, years,
ol – power cable overload, %.

Figure 9 presents a variation of the function E_{ins} described in (3) for cable overload within the range (0–45)%. One can see that thermal endurance of the cable insulation decreases five times (from 20 to 4 years) when the cable overload in equal to 20%. When the overload is equal to 45% or more, the thermal endurance is below 1 year.

Aggregated results of the thermal endurance calculation according to (3) are presented in Tab. 3. Consecutive calculations of the endurance named "at given overload" are performed with the assumption that power cables (their insulation) are operating all year with insulation temperature resulting from the load current being equal to the ampacity of cables buried in the ground. For example, thermal resistivity of the soil 0.5 (K·m)/W gives ampacity in the ground 230 A. Such a current gives (in the air of ambient temperature 25°C and presence of solar radiation) an overload equal to 140% (insulation temperature 131°C). In effect, thermal endurance of the power cables operating all the time in this temperature is equal to 0.24 years (around 3 months).

However, it is obvious that ambient temperature varies within the day and within the year seasons. It is important especially for cables placed in the air. In Polish climate conditions, the average sunshine

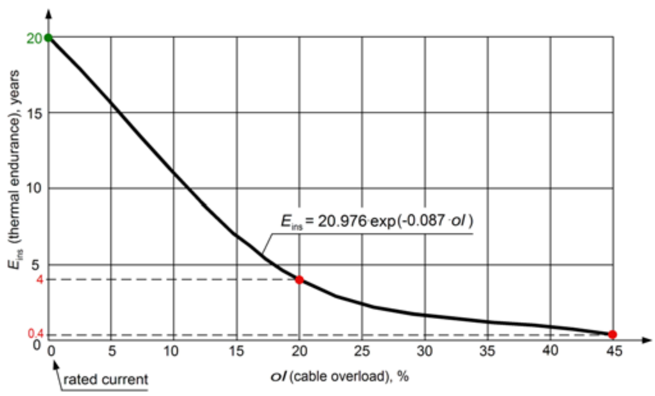


Fig. 9. Variation of the thermal endurance E_{ins} of the PVC insulation as a function of the cable overload

duration is assumed to be 1600 hours a year [21]. During this time, the direct influence of solar radiation is considered. The average number of days per year with a temperature above 25°C (without direct effect of solar radiation) in Polish conditions is 38 [20]. The information provided above was used to calculate the resultant thermal endurance of the power cables insulation (last column in Tab. 3), taking into account variation of ambient conditions within the year. Figure 10 presents graphical comparison of the resultant thermal endurance for all analyzed cases.

Example calculations are provided below for the case when thermal resistivity of the soil for cables buried in the ground is equal to 0.5 (K·m)/W (Tab. 3). It was assumed the insulation degradation proceeds linearly during the time for a constant overload. The resultant thermal endurance $E_{ins\text{-}res}$ is then as follows:

• without solar radiation:

$$E_{ins\text{-}res} = \frac{1}{\frac{t_a}{E_{rat}} + \frac{t_b}{E_{ins\text{-}b}}} = \frac{1}{\frac{\frac{365-38}{365}}{20} + \frac{1-\frac{365-38}{365}}{0.24}} = 2.09 \text{ years} \quad (4)$$

where:
t_a – period of the year for which there is no overload,
t_b – period of the year for which the overload occurs (1- t_a),
E_{rat} – rated thermal endurance (20 years for insulation temp. 70°C),
$E_{ins\text{-}b}$ – thermal endurance (years) at given overload (see Tab. 3),
365 – total days per year,
38 – days per year with temperature above 25°C [20],
0.24 – thermal endurance (years) for cable load 230 A and without solar radiation (see Tab. 3),

• with solar radiation:

$$E_{ins\text{-}res} = \frac{1}{\frac{t_a}{E_{rat}} + \frac{t_b}{E_{ins\text{-}b}}} = \frac{1}{\frac{\frac{8760-1600}{8760}}{20} + \frac{1-\frac{8760-1600}{8760}}{0.00011}} = 0.00061 \text{ years} \quad (5)$$

where:
8760 – total hours per year,
1600 – hours per year with solar radiation at least 1000 W/m^2 [21],
0.00011 – thermal endurance (years) for cable load 230 A and solar radiation (see Tab. 3).

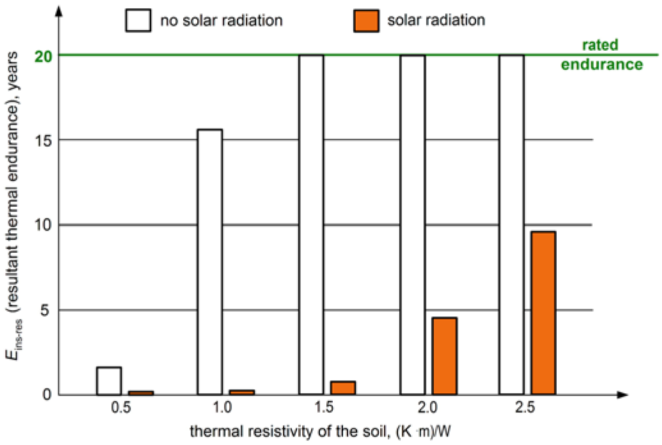


Fig. 10. Resultant thermal endurance of the insulation of power cables partially installed in the air (for detailed values see last column in Tab. 3)

The above-presented examples of calculation reflect the most unfavourable case of the line "ground-air". For this case, power cables are allowed to be loaded relatively high (230 A), due to good parameters of soil for heat dissipation from cables to the ground. Overheating and risk of thermal failure of the section in air occur because this section is allowed to be loaded significantly lower (152 A without solar radiation and 96 A with solar radiation – see Tab. 3). In consequence, resultant thermal endurance of the power cables insulation can be even around 90% lower than the rated thermal endurance (around 2 years instead of 20).

If ampacity of the power cable line located in the arrangement "ground-air" is adopted on the base of the thermal condition in the ground, for the cases with thermal resistivity of the soil equal to 0.5 and 1.0 (K·m)/W decrease in the thermal endurance and thermal failure of the power cables in air may occur even without solar radiation. For cases with thermal resistivity of the soil equal to 1.5, 2.0 and 2.5 (K·m)/W resultant thermal endurance is approx. equal to the rated endurance (20 years), but only when no solar radiation occurs. When solar radiation affects the cables, the endurance decrease is at least 50%. It gives a negative impact on power network maintenance costs as well as the reliability of supply.

During the design process of the power cable lines placed partially in the ground and in air, the ampacity of the cables in the ground should be evaluated taking into account ambient conditions in air, including solar radiation, especially when the ampacity is obtained for the soil of very low thermal resistivity.

5. Conclusions

Power cable lines are usually directly buried in the ground, which gives relatively favourable thermal conditions for heat transfer from the cables to the surrounding environment. However, in practice, these lines are very often supplied from poles of overhead lines and part of the cable line in free air cannot be loaded at the same level as the part in the ground. The real risk of the cables overheating in air occurs, especially in the presence of direct solar radiation. Results of the investigation conducted in this paper have shown that in case of the most unfavourable conditions (low thermal resistivity of the soil and strong solar radiation occur) power cables insulation in air may have temperature almost two times higher than the permissible value (131°C instead of 70°C). Overheating of the insulation leads to a decrease of its thermal endurance, which in real operating conditions may be over 10 times lower than the rated value, assumed during the project stage. All these aspects may lead to worsened reliability of supply and higher cost of the power network maintenance. Thus, in such cases, advanced modelling of the power cables thermal condition in the air is strongly recommended to be applied, in addition to the typical simple calculation according to the commonly used IEC standards.

References

1. Benato R, Paolucci A. EHV AC Undergrounding Electrical Power: Performance and Planning. Springer, 2010, https://doi.org/10.1007/978-1-84882-867-4.
2. Brender D, Lindsey T L. Effect of rooftop exposure in direct sunlight on conduit ambient temperatures. IEEE Transactions on Industry Applications 2008; 44 (6): 1872-1878, https://doi.org/10.1109/TIA.2008.2006301.
3. CYMCAP - software for power cable ampacity rating.
4. Czapp S, Czapp M, Szultka S, Tomaszewski A. Ampacity of power cables exposed to solar radiation - recommendations of standards vs. CFD simulations. 17th International Conference Heat Transfer and Renewable Sources of Energy (HTRSE-2018), Międzyzdroje, Poland, 02-05.09.2018, E3S Web of Conferences 2018; 70 (03004): 1-5, https://doi.org/10.1051/e3sconf/20187003004.
5. Czapp S, Ratkowski F. Effect of soil moisture on current-carrying capacity of low-voltage power cables. Przeglad Elektrotechniczny 2019; 95 (6): 154-159, https://doi.org/10.15199/48.2019.06.29.
6. Czapp S, Ratkowski F, Szultka S, Tomaszewski A. Overheating of underground power cable line due to its partial exposition to solar radiation. 24th International Conference on Methods and Models in Automation and Robotics (MMAR) 2019, Międzyzdroje, Poland, 26-29.08.2019: 396-400, https://doi.org/10.1109/MMAR.2019.8864691.

7. Czapp S, Szultka S, Tomaszewski A. CFD-based evaluation of current-carrying capacity of power cables installed in free air. 18th International Scientific Conference on Electric Power Engineering (EPE) 2017, Kouty nad Desnou, Czech Republic, 17-19.05.2017: 692-697, https://doi.org/10.1109/EPE.2017.7967271.
8. Czapp S, Szultka S, Tomaszewski A, Szultka A. Effect of solar radiation on current-carrying capacity of PVC-insulated power cables - the numerical point of view. Tehnicki Vjesnik 2019; 26 (6): 1821-1826, https://doi.org/10.17559/TV-20181029214825.
9. De Leon F. Calculation of underground cable ampacity. CYME Int. TD 2005: 1-6.
10. De Leon F. Major factors affecting cable ampacity. IEEE Power Engineering Society General Meeting 2006: 1-6, https://doi.org/10.1109/PES.2006.1708875.
11. Gebura A, Kowalska D, Tokarski T. Badania przyśpieszonego starzenia przewodów elektrycznych. Research Works of Air Force Institute of Technology, 2003.
12. HD 60364-5-52: Low-voltage electrical installations - Part 5-52: Selection and erection of electrical equipment - Wiring systems, 2011.
13. Holyk C, Anders G J. Power cable rating calculations-A historical perspective. IEEE Industry Applications Magazine 2015; 21 (4): 6-64, https://doi.org/10.1109/MIAS.2015.2417094.
14. IEC 60287-1-1: Electric cables - Calculation of the current rating - Part 1-1: Current rating equations (100% load factor) and calculation of losses - General, 2006.
15. IEC 60287-2-1: Electric cables - Calculation of the current rating - Part 2-1: Thermal resistance - Calculation of the thermal resistance, 2015.
16. IEC 60287-3-1: Electric cables - Calculation of the current rating - Part 3-1: Sections on operating conditions - Reference operating conditions and selection of cable type, 1999.
17. Kacejko P, Kmak J, Nowak W, Pijarski P, Szpyra W, Tarko R, Wydra M. Dynamic management of transmission capacity in power systems. Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej 2017; 53: 107-110.
18. Klimenta D, Perović B, Klimenta J, Jevtić M, Milovanović M, Krstić I. Modelling the thermal effect of solar radiation on the ampacity of a low voltage underground cable. International Journal of Thermal Sciences 2018; 134: 507-516, https://doi.org/10.1016/j.ijthermalsci.2018.08.012.
19. Kornatka M. Analysis of the exploitation failure rate in Polish MV networks. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (3): 413-419, https://doi.org/10.17531/ein.2018.3.9.
20. Kossowska-Cezak U. Zmiany wieloletnie liczby termicznych dni charakterystycznych w Warszawie (1951 - 2010). Prace Geograficzne 2014; 136: 9-30.
21. Kuczmarski M. Usłonecznienie Polski i jego przydatność dla helioterapii, Dokumentacja Geograficzna 1990; 4.
22. Instalacje elektryczne i teletechniczne. Poradnik montera i inżyniera elektryka (red. J. Strzałka). Rozdział: Zabezpieczenia w instalacjach elektrycznych. VERLAG DASHÖFER, Warszawa 2001-2019, ISBN 978-83-88285-11-0.
23. Notton G, Voyant C, Fouilloy A, Duchaud J L, Nivet M L. Some applications of ANN to solar radiation estimation and forecasting for energy applications. Applied Sciences 2019; 209 (9): 1-20, https://doi.org/10.3390/app9010209.
24. Olejnik B, Łowczowski K. Techniczne metody poprawy współczynników SAIDI oraz SAIFI stosowane w sieci dystrybucyjnej, Computer Applications in Electrical Engineering 2016, Poznan, Poland, 18-19.04.2016.
25. Shabani H, Vahidi B. A probabilistic approach for optimal power cable ampacity computation by considering uncertainty of parameters and economic constraints, International Journal of Electrical Power and Energy Systems 2019; 106: 432-443, https://doi.org/10.1016/j.ijepes.2018.10.030.
26. Spyra F. Wpływ czynników zewnętrznych na obciążalność prądową kabli w elektroenergetycznej linii kablowej. Energetyka 2007; 6-7: 451-454.
27. Yang L, Qiu W, Huang J, Hao Y, Fu M, Hou S, Li L. Comparison of conductor-temperature calculations based on different radial-position-temperature detections for high-voltage power cable. Energies 2018; 11 (1): 1-17, https://doi.org/10.3390/en11010117.
28. Zawodniak J. Ageing processes in insulation of cable line and overhead cover conductor line. Automatyka, Elektryka, Zakłócenia 2018; 31 (1): 34-40, https://doi.org/10.17274/AEZ.2018.31.03.

Stanisław CZAPP
Seweryn SZULTKA
Filip RATKOWSKI
Gdańsk University of Technology
Faculty of Electrical and Control Engineering
Narutowicza 11/12 str., 80-233 Gdańsk, Poland

Adam TOMASZEWSKI
Polish Academy of Sciences
Institute of Fluid Flow Machinery
Fiszera 14 str., 80-231 Gdansk, Poland

E-mails: stanislaw.czapp@pg.edu.pl, seweryn.szultka@pg.edu.pl, filip.ratkowski@pg.edu.pl, atomaszewski@imp.gda.pl

Article citation info:
YU G, DU Y, YAN L, REN F. Stress-strength interference-based importance for series systems considering common cause failure. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 241–252, http://dx.doi.org/10.17531/ein.2020.2.7.

Guangwei YU
Yanwei DU
Li YAN
Fangyu REN

STRESS-STRENGTH INTERFERENCE-BASED IMPORTANCE FOR SERIES SYSTEMS CONSIDERING COMMON CAUSE FAILURE

OCENA OPARTEJ NA MODELU OBCIĄŻENIOWO-WYTRZYMAŁOŚCIOWYM WAŻNOŚCI ELEMENTÓW SYSTEMU SZEREGOWEGO Z UWZGLĘDNIENIEM USZKODZEŃ WYWOŁANYCH WSPÓLNĄ PRZYCZYNĄ

Series systems, whose structures are simple, are widely discovered in practical engineering, but the interdependency between the components is complex, such as common cause failure. With the consideration of the components' strength, this paper focuses on ranking the importance measure of components considering the common cause failure based on the stress-strength interference (SSI) model. The weakest component can be identified by integrating the SSI model with the importance measure when the strength mean and variance of the component under the load stress is known. Firstly, the analytic methods are proposed to calculate the SSI-based importance of components in the series systems. Then, the monotonicity of SSI-based importance is analyzed by changing the strength mean or strength variance of one component. The results show that the SSI-based importance of components, whose parameters are changed, will reduce monotonically with the increase of strength mean or increase monotonically with the increase of strength variance. Finally, a component replacement method is developed based on the rules that both the importance of replaced component and the importance ranks should be unchanged after the replacement. SSI-based importance can help engineers to make maintenance decisions, and the component replacement method can increase the diversity of spare parts by finding the equivalent components.

Keywords: *importance measure; common cause failure; stress-strength interference; monotonicity analysis; component replacement.*

Systemy szeregowe, które są szeroko stosowane w praktyce inżynieryjnej, charakteryzują się prostą strukturą, jednak współzależności między ich elementami są złożone, czego przykładem są uszkodzenia wywołane wspólną przyczyną. Rozważając wytrzymałości składowych systemu, opracowano metodę szeregowania miar ważności składowych z uwzględnieniem uszkodzeń wywołanych wspólną przyczyną. Metoda ta pozwala zidentyfikować najsłabsze ogniwo systemu. Miarę istotności zintegrowano z modelem obciążeniowo-wytrzymałościowym (SSI), biorąc pod uwagę średnią i wariancję wytrzymałości elementu pod obciążeniem. W pierwszym kroku opracowano metody analityczne pozwalające na obliczanie opartej na SSI ważności elementów w systemach szeregowych. Następnie analizowano monotoniczność opartej na SSI ważności zmieniając średnią lub wariancję wytrzymałości jednego z elementów. Wyniki pokazują, że mierzona w oparciu o SSI ważność elementów, których parametry są zmieniane, maleje monotonicznie wraz ze wzrostem średniej wytrzymałości lub rośnie monotonicznie wraz ze wzrostem wariancji wytrzymałości. Na podstawie przeprowadzonych badań, opracowano metodę wymiany części, opartą na zasadzie polegającej na tym, że zarówno ważność zastąpionego elementu, jak i rangi ważności powinny pozostać niezmienione po wymianie. Możliwość określania ważności opartej na modelu SSI może pomóc inżynierom w podejmowaniu decyzji dotyczących konserwacji, zaś proponowana metoda wymiany elementów systemu pozwala zwiększyć różnorodność części zamiennych poprzez znalezienie równoważnych elementów.

Słowa kluczowe: *miara ważności; uszkodzenia wywołane wspólną przyczyną; model obciążeniowo-wytrzymałościowy; analiza monotoniczności; wymiana części.*

1. Introduction

The protection and security of components should be considered in the implementation of risk management, and the interdependencies within the components are a significant challenge for risk management. The common cause failure (CCF), which can cause the failure of multiple components with the common reason, is a typical reason of the interdependence between components in series systems. The importance analysis of components plays a vital role in the risk management of series systems. Importance measure is one of the significant branches of reliability theory and has a significant advance with the development of reliability engineering. It can evaluate the impact of the individual component on the system reliability when the component is failure. The results of importance measures could facilitate the reliability design, component assignment problem, redundancy allocation, system upgrading, fault diagnosis, and maintenance. Nowadays, importance measures have been widely applied in the fields of engineering, such as oil and gas transmission, railway systems, nuclear power production, manufacturing systems, and computer systems, and so on [17].

In 1968, Birnbaum first put forward the calculation method of importance measures for binary systems [4], and Birnbaum importance is classified into three categories as follows. The first category is the structure importance measure, which represents the role of the positions in the system that the components occupy; The second category is the reliability importance measure, which considered the component reliability to evaluate the effect of a component on the system reliability; The third category is the lifetime importance measure, which considered the component reliability in the life cycle to evaluate the influence of components on the system reliability. Based on the Birnbaum importance, researchers have evaluated the influence of component's reliability on the system reliability from different perspectives. So many new kinds of importance measures are proposed, such as F-V importance measure [13, 31], BP structure importance measure [1], critical importance measure [19], risk achievement worth (RAW), risk reduction worth (RRW) [32], improvement potential importance [38], and differential importance measure [5]. In recent years, some researchers have been extended importance measures to by considering the maintenance policy [9, 10], the transition rate increases over time [8], or the constraints on cost for improving the system reliability [24].

A binary system assumes that the state of the system and component only has two states: functioning and failure. However, there are a large number of multi-state systems (MSS) with more than two states in practice [22]. Barlow and Wu [2] summarized the analysis methods of MSS, where the system state was defined to be the worst state when the component is in the best minimal path set, or equivalently, the best state when the component is in the worst minimal cut set. Many of the results for the binary cases can be computed for MSS by using the binary structure and reliability function. Griffith [16] proposed the MSS performance, which described that the performance level is corresponding to different system states, and studied the effect of component improvement on the system performance. Wu and Chan [35] defined a new unity function based on the component state of MSS for measuring which component has the maximum contribution to improving the system performance. Zio and Podofillini [46] proposed the approach to evaluate the importance of all the components concerning a given performance level and expanded some binary importance measures to MSS, such as RAW, RRW, F-V, and Birnbaum importance. Ramirez-Marquez and Coit [25, 26] put forward the composite importance measures to evaluate the effect of all the states of components on the system reliability, which could break through the limitations of only considering the effect of a single state of the component on the system reliability. Levitin et al. [20] evaluated the importance measures for MSS based on the universal generating function technique and verified the effectiveness of the approach. Shrestha et al. [28] presented an analytical method based on multi-state multi-valued decision diagrams for multistate component importance analysis. Zhao et al. [42, 43] presented the mission success importance for multi-state repairable k-out-of-n systems. Do Van et al. [6, 7] put forward the multi-directional sensitivity measure within the framework of Markovian systems, which calculated the differential importance measure of risk-informed decision-making in the context of Markov reliability models. Natvig [23] raised the dynamic and stationary importance measures in repairable and nonreparable multistate coherent systems. Zhao et al. [45] introduced the redundancy importance measure into the multi-objective optimization of reliability-redundancy allocation problems for serial parallel-series systems. Si et al. [29] proposed the concept of integrated importance measure, which concerned the probability distributions and transition intensities of the component states simultaneously. Wang et al. [33] considered the improvement of system reliability based on Birnbaum importance by increasing the maintenance cost. Zhang et al. [41] proposed the Birnbaum importance-based quantum genetic algorithm for solving the component assignment problems.

However, no single type of importance measure can fit for all systems and conditions. Various importance measures for the same system may get the different ranks of components and lead to making different decisions. With the development of science and technology, engineering systems become more complicated, such as higher order systems, multi-loop control systems, nonlinear systems, hierarchy systems, and uncertain systems. For the design and optimization of such complex systems, some new importance measures are needed to judge the relative strength of a component in a system for different criteria [18]. At present, the importance measures are always calculated based on the independent reliabilities by assuming the component failure in a system is statistical independent, such as references [44] and [30]. This assumption is accordant with the possible working conditions for electronic systems, while it could not apply to the mechanical systems with complicated failure modes and failure mechanisms.

CCF was firstly proposed by Fleming in 1975 to represent the multiple components failures caused by the common reason [12]. CCF exists widely in most kinds of complex industrial systems, especially for nuclear facilities, weapon systems, and aerospace systems. Since the 1970s, researchers have put forward many analysis models for the CCF problems, where explicit analysis and implicit substitution were two typical modelling methods [40]. The precise analysis method denotes the system reliability based on the component state directly with the large scale of computation, which is generally applicable to all kinds of CCF.

The stress-strength interference (SSI) model has been commonly used in the reliability modelling of mechanical systems [11, 47]. Bhattacharyya and Johnson [3] established the interference reliability model for k out of n system by assuming that the system component was independent and identically distributed. For the 1 out of 2: G system, Lewis [21] disposed the Poisson distributed load (stress) by Markov model and evaluated the reliability of the system with CCF. Moreover, Xie et al. [36] introduced the concept of order statistics into the SSI model and calculated the equivalent strength of series, parallel, and k out of n systems for evaluating the system reliability. With the assumption of strength degradation, Xue and Yang [39] put forward the deterministic strength degradation model and random strength degradation model based on the interference analysis methods. Wang and Xie [34] structured the equivalent load according to the probability of order statistics when the load was applied at multiple times under a Poisson process. Shen et al. [27] evaluated the structure reliability when the load was under a Poisson process, and the degradation of structural strength was under a Gamma process. Gao et al. [15] established the dynamic reliability model based on the equivalent strength degradation paths to analyze the mechanical components with uncertain strength caused by material parameters. Furthermore, Gao and Xie [14] extended the dynamic reliability models for mechanical load-sharing parallel systems with strength degradation path dependence. Generally, current importance measures cannot assist effective decision making for the complex systems with CCF or dependent components [37]. So, taking advantage of the SSI model, this paper will propose the importance measure for systems with CCF to fit for the features of complex systems.

In practice, the component reliability is hard to observe and record, but the strength information of components can be observed easily. Therefore, the SSI model can be used to simplify the analysis of CCF in the system. Considering the advantages of the importance measure and the SSI, SSI-based importance measure is developed to evaluate the importance ranking of components. The significance of SSI-based importance can be summarized as follows. (1) The evaluation method of SSI-based importance can identify the weakest link in the system, which can help engineers to make decisions for maintenance activities. (2) The component replacement method can find more equivalent components, which can increase the diversity of spare parts.

The remaining of this paper is organized as follows. Section 2 describes the ideas of SSI-based importance considering CCF in general and gives the analytic expression of component importance in series systems. Section 3 analyzes the monotonicity of SSI-based importance for series systems. Section 4 introduces two numerical experiments to verify the monotonicity of SSI-based importance. Section 5 introduces a new component replacement method based on the ideas that both the SSI-based importance of the replaced component and the importance ranks of all components should be unchanged after replacing. Section 6 concludes the research work.

2. SSI-based importance considering CCF

The performance of components is always related to the load stress and the strength parameters (mean and variance) of components. The traditional reliability importance measure considers the effect of inherent component reliability on the system reliability instead of considering the effect of the load stress and the strength parameters of the components. For series systems, each component has independent strength. If the system is under one shock, this shock will act on all the components simultaneously, as the same load stress. Supposing the component will not bear any other individual stress, the component's reliability is the interference of the same load stress and component strength, and the failure of component occurs when the stress exceeds strength. Sometimes, the components failed because of the occurrence of CCF, and the CCF can be equitant to the same stress load acting on all the components. Therefore, SSI-based importance is developed to analyze the importance of components in the system under the same load stress.

2.1. SSI-based importance in series systems considering CCF

For a system with n components, the reliability of component j $(1 \le j \le n)$ can be written as $R_j = R_j(l, s_j)$, which is the function of the same load stress l and component strength s_j. The system reliability is the function of system structure and components' reliabilities, which can be noted as $R_s = \phi(R_1,...,R_n) = \phi(l, s_1, s_2, \cdots, s_n)$, in which ϕ is the system structure function; l is common load stress and $s_j (j = 1, 2, \cdots, n)$ is the strength of component *j*.

According to the idea of Birnbaum importance, the SSI-based importance of component *j* can also be expressed by the system structure function, load stress, and component strength, as shown in Equation 1:

$$I_j = \frac{\partial\phi(l, s_1, s_2, \cdots, s_n)}{\partial R_j(l, s_j)} = \frac{\partial\phi(l; s_1, s_2, \cdots, s_n) / \partial s_j}{\partial R_j(l, s_j) / \partial s_j} \tag{1}$$

Since it is difficult to directly obtain the derivative of system reliability on component reliability in the form of SSI model, we will explore the relationship between importance measure and component strength. The component reliability will be improved with the improvement of component strength when the load stress is fixed. That is to say, and component reliability $R_j(l, s_j)$ is monotonic for the variable s_j when stress *l* remains unchanged. The inverse function $R_j(l, s_j)$ can be obtained as $s_j = s_j(R_j, l)$, so the SSI-based importance can be expressed as the following formula in Equation 1. Then the partial derivative of s_j on $R_j(l, s_j)$ can be written as reciprocal of the partial derivative of $R_j(l, s_j)$ on s_j. Therefore, the ultimate expression of SSI-based importance is shown as the ratio of two derivatives on s_j in Equation 1. The numerator of the final expression represents the influence of component strength on the system reliability, and the denominator represents the influence of component strength on the reliability of itself. The real significance of SSI-based importance represents the relative change rate caused by the change of the component strength.

SSI-based importance extends the connotations of Birnbaum importance from component reliability to component strength. It is easy to obtain the specific value of SSI-based importance, but it is hard to determine the derivative of the system reliability. Therefore, Equation 1 can be transformed as the limit format based on the definition of the derivative as

$$I_j = \lim_{\Delta s_j \to 0} \frac{[R_s(t_1) - R_s(t_2)] / \Delta s_j}{[R_j(t_1) - R_j(t_2)] / \Delta s_j} \tag{2}$$

where $R_s(t_1)$ and $R_s(t_2)$ denote the system reliability before (at the time t_1) and after (at the time t_2) the change of component strength, respectively; $R_j(t_1)$ and $R_j(t_2)$ denote the component reliability before and after the change of component strength; Δs_j is the change of the strength for component *i*.

2.2. Analytic evaluation of SSI-based importance in series systems considering CCF

The structures of some systems are the series system, which will fail if any component fails, and the reliability block diagram of the series system is shown in Fig. 1.

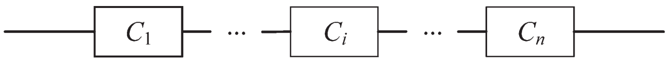


Fig. 1. Reliability block diagram of the series system

Assume the probability density function (PDF) of the strength for component *i* is $f_{si}(s)$, $i = 1, \cdots, n$. All components in the system bear the same load stress, and the corresponding PDF is $f_l(l)$. The reliability of series system can be expressed based on SSI in Equation 3:

$$R_{seri}(t) = \int_0^\infty f_l(l) \prod_{i=1}^{n} [1 - \int_0^l f_{si}(s) ds] dl \tag{3}$$

Assume that only the strength of component *j* changes. According to Equations 2 and 3, the SSI-based importance of component *j* can be described as follows:

$$\begin{aligned} I_j^{seri} &= \lim_{\Delta sj \to 0} \frac{\left\{ \int_0^\infty f_l(l) \prod_{i=1, i \ne j}^{n} [\int_l^\infty f_{si}(s) ds][\int_l^\infty (f_{sj1}(s) - f_{sj2}(s)) ds] dl \right\} / \Delta s_j}{\left\{ \int_0^\infty f_l(l) [\int_l^\infty \Delta_j(s) ds] dl \right\} / \Delta s_j} \\ &= \frac{\int_0^\infty f_l(l) \prod_{i=1, i \ne j}^{n} [\int_l^\infty f_{si}(s) ds][\int_l^\infty \lim_{\Delta sj \to 0} (\Delta_j(s) / \Delta s_j) ds] dl}{\int_0^\infty f_l(l) [\int_l^\infty \lim_{\Delta sj \to 0} (\Delta_j(s) / \Delta s_j) ds] dl} \end{aligned} \tag{4}$$

where the initial strength PDF of component *j* is $f_{sj1}(s)$, and the after-change PDF is $f_{sj2}(s)$, $\Delta_j(s) = f_{sj1}(s) - f_{sj2}(s)$ denotes the strength change of component *j*.

In order to analyze the SSI-based importance clearly, we have discussed the three forms of SSI-based importance of component *j* in series systems as follows. If the strength of component *j* follows a continuous univariate distribution, Equation 4 can be simplified as follows:

$$\begin{aligned} I_j^{seri} &= \frac{\int_0^\infty f_l(l) \prod_{i=1, i \ne j}^{n} [\int_l^\infty f_{si}(s) ds][\int_l^\infty (df_{sj}(s) / ds) ds] dl}{\int_0^\infty f_l(l) [\int_l^\infty (df_{sj}(s) / ds) ds] dl} \\ &= \frac{\int_0^\infty f_l(l) \prod_{i=1, i \ne j}^{n} [\int_l^\infty f_{si}(s) ds][f_{sj}(\infty) - f_{sj}(l)] dl}{\int_0^\infty f_l(l) [f_{sj}(\infty) - f_{sj}(l)] dl} \end{aligned} \tag{5}$$

Generally, the variable PDF tends to be '0' when the variable tends to be endless. So the SSI-based importance for the series system can be written as:

$$I_j^{seri} = \frac{\int_0^\infty f_l(l) f_{sj}(l) \prod_{i=1, i\neq j}^{n} [\int_l^\infty f_{si}(s)ds]dl}{\int_0^\infty f(l) f_{sj}(l)dl} . \tag{6}$$

Supposing that the failure of components in the system is independent, the importance measure for independent components can be expressed as:

$$I_{j\_indep}^{seri} = \prod_{i=1, i\neq j}^{n} \int_0^\infty f_l(l)[\int_l^\infty f_{si}(s)ds]dl = \prod_{i=1, i\neq j}^{n} R_i , \tag{7}$$

where R_i is the reliability of component C_i.

According to Equation 6, it is clear that the strength distribution of all components is taken into consideration for calculating the SSI-based importance. However, the importance of component is independent with its strength in Equation 7. In production practice, the distribution of component strength is the statistical result of test data, which may be different because of different quality levels and different processes. Therefore, the proposed importance measure can describe the influence of strength change on system reliability well for engineering applications. The SSI-based importance can be evaluated by solving the complex analytic geometry integration for series system.

3. Monotonicity analysis of SSI-based importance for series system

The degradation of components is related to the inherent reliability, the strength mean, and the strength variance. For the traditional Birnbaum importance, the component importance has no concern with its inherent reliability, which is related to the reliabilities of other components. For the SSI-based importance of component *i*, its importance measure also has nothing to do with the inherent reliability of component *i*, but its importance measure has a close relationship with the strength mean and the strength variance. The reduction of strength mean or the increment of strength variance appears in component, which means the component begins to degrade. Therefore, the monotonicity analysis of SSI-based importance is discussed based on the changes of strength mean and the strength variance.

3.1. Monotonicity of SSI-based importance about strength mean

When component strength and stress distributions are a normal distribution, Equation 6 can be simplified base on the PDF of the normal distribution as Equation 8:

$$I_j^{seri} = \frac{\int_0^\infty \exp(-\frac{(l-\mu_l)^2}{2\sigma_l^2})\exp(-\frac{(l-\mu_j)^2}{2\sigma_j^2}) \prod_{i=1, i\neq j}^{n} [\int_l^\infty \exp(-\frac{(s-\mu_i)^2}{2\sigma_i^2})ds]dl}{(2\pi)^{\frac{n-1}{2}} \prod_{i=1, i\neq j}^{n} \sigma_i \int_0^\infty \exp(-\frac{(l-\mu_l)^2}{2\sigma_l^2})\exp(-\frac{(l-\mu_j)^2}{2\sigma_j^2})dl}$$
$$= \frac{\int_0^\infty f(l) f_j(l;\mu_j,\sigma_j) \prod_{i=1, i\neq j}^{n} H_i(l)dl}{p\int_0^\infty f(l) f_j(l;\mu_j,\sigma_j)dl} \tag{8}$$

where $f(l) = \exp(-\frac{(l-\mu_l)^2}{2\sigma_l^2})$ is the function of variables μ_l and σ_l, $f_j(l;\mu_j,\sigma_j) = \exp(-\frac{(l-\mu_j)^2}{2\sigma_j^2})$ is the function of variables μ_j and σ_j, $H_i(l) = \int_l^\infty \exp(-\frac{(s-\mu_i)^2}{2\sigma_i^2})ds$ is the function of parameters of all components except component j, $p = (2\pi)^{\frac{n-1}{2}} \prod_{i=1, i\neq j}^{n} \sigma_i$.

In order to analyze the monotonicity of SSI-based importance, we can conduct the partial derivation of Equation 8 about μ_j when σ_j is constant. The result is represented by Equation 9:

$$\frac{\partial I_j^{seri}}{\partial \mu_j} = \frac{\partial I_j^{seri}}{\partial f_j(l,\mu_j)} \times \frac{\partial f_j(l,\mu_j)}{\partial \mu_j}$$
$$= \frac{\int_0^\infty f(l) \prod_{i=1, i\neq j}^{n} H_i(l)g(l)dl \int_0^\infty f(l) f_j(l,\mu_j)dl}{p[\int_0^\infty f(l) f_j(l,\mu_j)dl]^2}$$
$$- \frac{\int_0^\infty f(l) f_j(l,\mu_j) \prod_{i=1, i\neq j}^{n} H_i(l)dl \int_0^\infty f(l)g(l)dl}{p[\int_0^\infty f(l) f_j(l,\mu_j)dl]^2} \tag{9}$$

where $g(l) = \frac{\partial f_j(l,\mu_j)}{\partial \mu_j}$, $f_j(l,\mu_j)$ represents $f_j(l;\mu_j,\sigma_j)$ when σ_j is a constant.

Since the integral operation in Equation 9 is complex, a unique series system with four components is introduced to illustrate the monotonicity of SSI-based importance. The strength of components follows the normal distribution of $s_1 \sim N(\mu_1, 50)$, $s_2 \sim N(420, 50)$, $s_3 \sim N(440, 50)$, $s_4 \sim N(460, 50)$, and the same load stress on these four components follows $l \sim N(300, 60)$. When μ_1 varies in the interval [350, 700], the SSI-based importance and the partial derivation of importance measure can be evaluated by Equations 8 and 9.

In order to analyze the changes of the SSI-based importance and its rate, the strength variances of components in the series systems are the same. The evaluation results are shown in Fig. 2. From the top figure, the importance value decreases when the strength mean increases, which indicates that the better the component quality is, the lower its importance value is. The bottom figure presents the partial derivative of importance, which means the change rate of importance. It is noteworthy that when the strength mean of component 1 reaches 462, the change rate is the highest at the point. The parameters of components 1 are almost the same as that of component 4 at this moment. Both of them have similar reliability because they have the same strength mean and strength variance.

3.2. Monotonicity of SSI-based importance about strength variance

In order to analyze the monotonicity of SSI-based importance, we can assume μ_j as a constant and conduct the partial derivation of Equation 8 about σ_j. The result is represented by Equation 10.

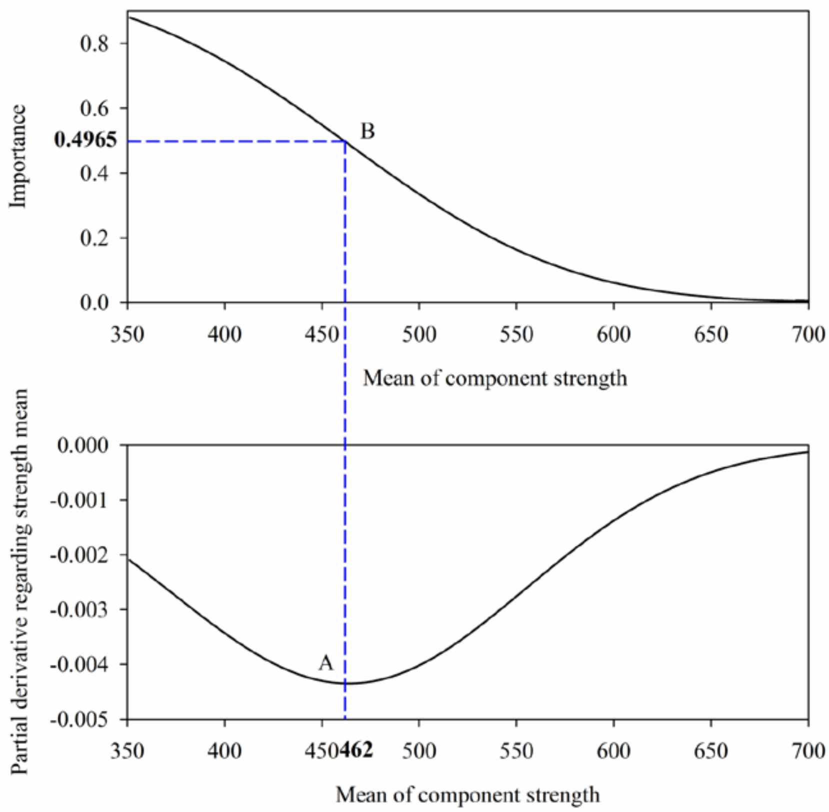


Fig. 2. The tendency of importance on the strength mean of component

tance of component will increases, which is the top one in Fig. 3. In the bottom of Fig. 3, the change rate reaches the highest value when the variance of component 1 is 22. The corresponding reliabilities of 4 components are R_{C1}=0.9698, R_{C2}=0.9635, R_{C3}=0.9797, R_{C4}=0.9894, respectively. This phenomenon illustrates that SSI-based importance increases fast when its inherent reliability is equal to the lowest reliable component.

4. Numerical experiments

In this section, we applied two numerical experiments to illustrate the methods in Sections 2 and 3 clearly. Experiment I shows the SSI-based importance ranking of 5 series systems with different component strength distributions, whose parameters are unchanged. Experiment II illustrates the changes of SSI-based importance of each component in series systems when the parameters of component 1 change, but the parameters of other components remain unchanged.

4.1. Experimental design

Experiment I: There are 5 series systems with different numbers of components, and the parameters of components are shown in Table 1. The experiment is established to illustrate the evaluation of SSI-based importance according to Equation 6, which also can determine the importance ranking of components in the series systems. The same load stress follows $l \sim N(200,30)$.

$$\frac{\partial I_j^{seri}}{\partial \sigma_j} = \frac{\partial I_j^{seri}}{\partial f_j(l,\sigma_j)} \times \frac{\partial f_j(l,\sigma_j)}{\partial \sigma_j}$$
$$= \frac{\int_0^\infty f(l) \prod_{i=1,i\neq j}^{n} H_i(l)h(l)dl \int_0^\infty f(l) f_j(l,\sigma_j)dl}{p[\int_0^\infty f(l) f_j(l,\sigma_j)dl]^2}$$
$$- \frac{\int_0^\infty f(l) f_j(l,\sigma_j) \prod_{i=1,i\neq j}^{n} H_i(l)dl \int_0^\infty f(l)h(l)dl}{p[\int_0^\infty f(l) f_j(l,\sigma_j)dl]^2} \quad (10)$$

where $h(l) = \frac{\partial f_j(l,\sigma_j)}{\partial \sigma_j}$, $f_j(l,\sigma_j)$ represents $f_j(l;\mu_j,\sigma_j)$ when μ_j is constant.

Similarly, a unique series system with four components is adopted to analyze the monotonicity of SSI-based importance with strength variance. The component follows the normal distributions as $s_1 \sim N(400,\sigma_1)$, $s_2 \sim N(440,50)$, $s_3 \sim N(460,50)$, $s_4 \sim N(480,50)$. The same load stress follows $l \sim N(300,60)$. In order to analyze the changes in the SSI-based importance and its rate with the increase of component variance, the strength mean of components in the series systems is unchanged. If σ_1 varies in the interval [10, 110], the SSI-based importance and its rate can be evaluated by Equations 8 and 9, which are shown in Fig. 3. Since the increase of strength variance indicates the component quality decreases, the SSI-based impor-

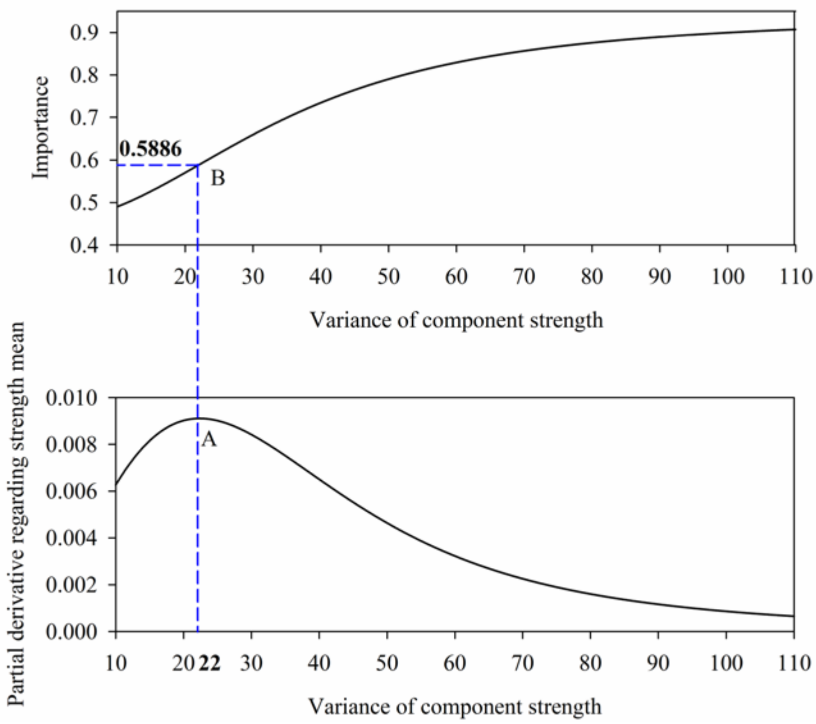


Fig. 3. The tendency of importance on strength variance of component

Experiment II: A four-component series system is introduced to illustrate the changes of SSI-based importance if the parameters of component 1 change and the parameters of other components are un-

Table 1. The parameters of components in Experiment I

Component #	System 1	System 2	System 3	System 4	System 5
1	$N(270,35)$	$N(290,30)$	$N(400,35)$	$N(315,45)$	$N(350,45)$
2	$N(370,60)$	$N(350,15)$	$N(300,45)$	$N(310,50)$	$N(255,10)$
3	/	$N(350,20)$	$N(335,55)$	$N(365,25)$	$N(380,25)$
4	/	$N(360,55)$	$N(280,60)$	$N(370,45)$	$N(390,10)$
5	/	/	$N(360,40)$	$N(280,40)$	$N(350,15)$
6	/	/	$N(290,15)$	$N(320,20)$	$N(360,50)$
7	/	/	/	$N(310,15)$	$N(360,45)$
8	/	/	/	$N(350,35)$	$N(300,25)$
9	/	/	/	/	$N(350,60)$
10	/	/	/	/	$N(275,10)$

Table 2. The parameters of components in Experiment II

Case #	Component 2	Component 3	Component 4
1	$N(200,30)$	$N(250,30)$	$N(300,30)$
2	$N(200,30)$	$N(250,30)$	$N(300,50)$
3	$N(200,30)$	$N(250,50)$	$N(300,30)$
4	$N(200,50)$	$N(250,30)$	$N(300,30)$
5	$N(200,30)$	$N(250,50)$	$N(300,50)$
6	$N(200,50)$	$N(250,50)$	$N(300,30)$
7	$N(200,50)$	$N(250,30)$	$N(300,50)$
8	$N(200,50)$	$N(250,50)$	$N(300,50)$

changed. The purposes of this experiment are to illustrate the changes of SSI-based importance for component 1 and analyze the changes of importance for other components except for the component 1. There are 6 cases that depend on the variance of components 2-4 comparing with the variance of stress. All the parameters of components are listed in Table 2 because the variance of component is 30 or 50, which means the strength variance of component is lower or higher than that of load stress. This experiment is introduced to analyze the changes of importance ranking for this specific series system. The same load stress follows $l \sim N(150,40)$, and the strength of component 1 follows $s_1 \sim N(\mu_1,\sigma_1)$. Let σ_1 vary in the interval [20, 50], and let μ_1 vary in the interval [180, 400].

4.2. Results analysis of experiments

4.2.1 Results of Experiment I

The SSI-based importance of all components in the series systems can be evaluated based on Equation 6, and the importance ranks of all components also can be obtained by comparing the importance values of all components. The results of Experiment I are shown in Table 3, which lists the importance value and importance ranks of all components in 5 series systems.

From Table 3, the strength of components with lower mean and higher variance has a higher importance measure, such as component 1 is more important than components 2 and 3 in System 2. If the strength mean of component is the same, the strength of component with higher variance has higher importance, such as component 2 ($s_2 \sim N(310,50)$) is more important than component 7 ($s_7 \sim N(310,15)$) in System 4. Similarly, if the strength variance of component is the same, the higher the strength mean of component is, the less important the component is, such as the component 10 ($s_{10} \sim N(275,10)$) is less important than component 2 ($s_2 \sim N(255,10)$) in System 5. If the variance and mean of component strength has a similar percentage increase (or decrease), the importance will decrease (or increase), which means the mean of component strength has more effect on the changes of component importance. For example, the strength mean of component 6 decreases 27.5% compared with that of component 1 in System 3, and the strength variance increases 57.1% from component 1 to component 7; but the SSI-importance of component increases with the decrease of strength mean. Therefore, the SSI-based importance of components is

Table 3. The SSI-based importance ranks of 5 series systems in Experiment I

System #	SSI-based importance value	SSI-based importance ranks
1	[0.9856, 0.7928]	$I_1 > I_2$
2	[0.9743, 0.0001, 0.2981, 0.9091]	$I_1 > I_4 > I_3 > I_2$
3	[0.2067, 0.7378, 0.7033, 0.9178, 0.4603, 0.3545]	$I_4 > I_2 > I_3 > I_5 > I_6 > I_1$
4	[0.7586, 0.8192, 0.1328, 0.5754, 0.8847, 0.2520, 0.1996, 0.4634]	$I_5 > I_2 > I_1 > I_4 > I_8 > I_6 > I_7 > I_3$
5	[0.5918, 0.8648, 0.0063, 2.36E-40, 1.04E-07, 0.6412, 0.5460, 0.3893, 0.7914, 0.1527]	$I_2 > I_9 > I_6 > I_1 > I_7 > I_8 > I_{10} > I_3 > I_5 > I_4$

higher when the strength mean of component is lower, or the strength variance of component is higher, and the mean has a higher effect on the SSI-based importance than that of variance.

4.2.2 Results of Experiment II

If the strength mean or strength variance of component 1 changes, then the changing tendency of SSI-based importance for all components can be recorded by Experiment II. From the results of Experiment II, we can find the SSI-based importance of any components in 8 cases has a similar change tendency. For component 1, the SSI-based importance decreases with the increase of strength mean of component 1, while the importance increases with the increase of strength variance. However, for components 2, 3, and 4, the SSI-based importance increases with the increase of strength mean of component 1, and the importance almost remains unchanged with the increase of variance when the strength mean is determined. In order to illustrate the detailed changes of importance, the results of cases 1 and 8 are shown in Fig. 4 and Fig. 5, respectively.

From Fig. 4, the SSI-based importance of component is close to 0 when strength mean is 25 and strength variance is 400. When the strength mean is fixed, the importance decreases with the decrease of the strength variance. Such as the importance is almost 0.2 when strength mean is 400 and strength variance is 56. However, the importance of the other three components increases with the increase of the strength mean for component 1, such as the importance of component 2 is 0.96 when strength mean is 390 while the importance becomes 0.47 when strength

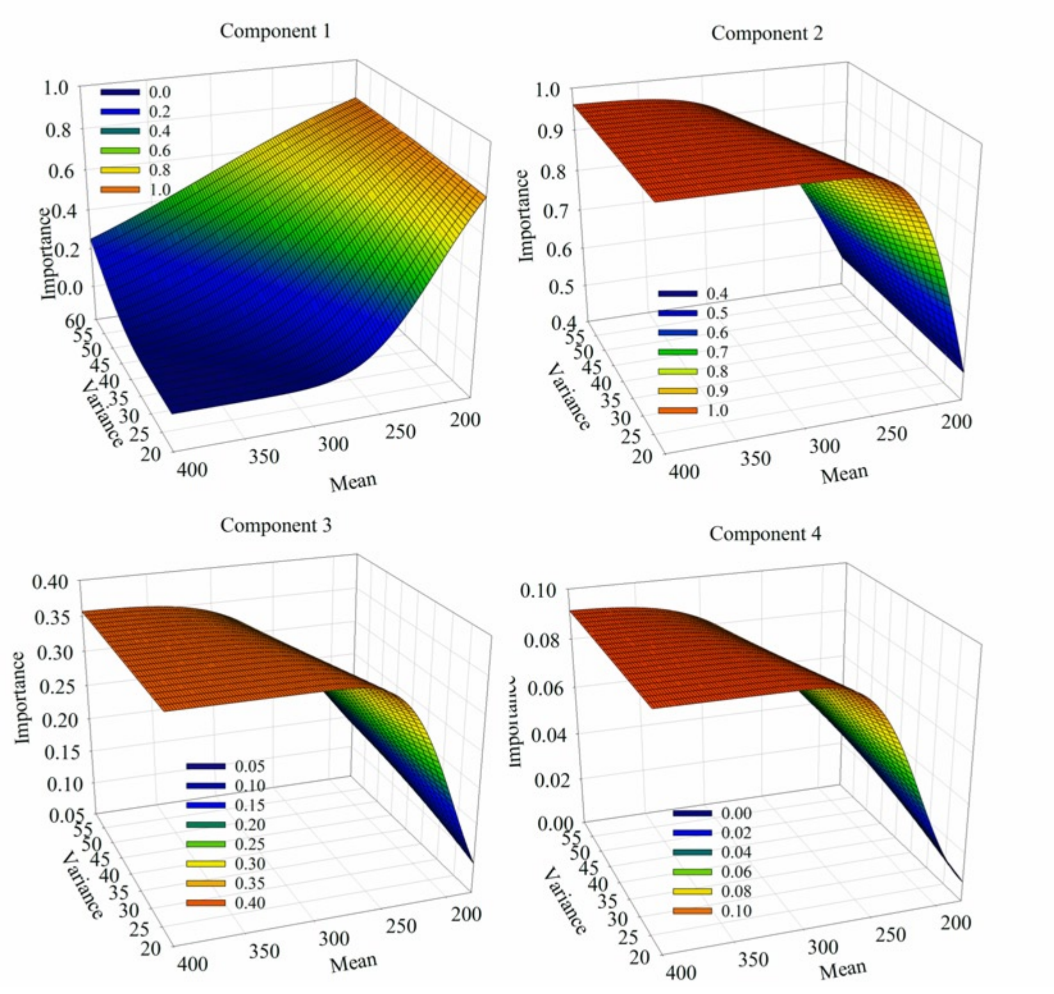


Fig. 4. The changes of SSI-based importance for case 1 in Experiment II

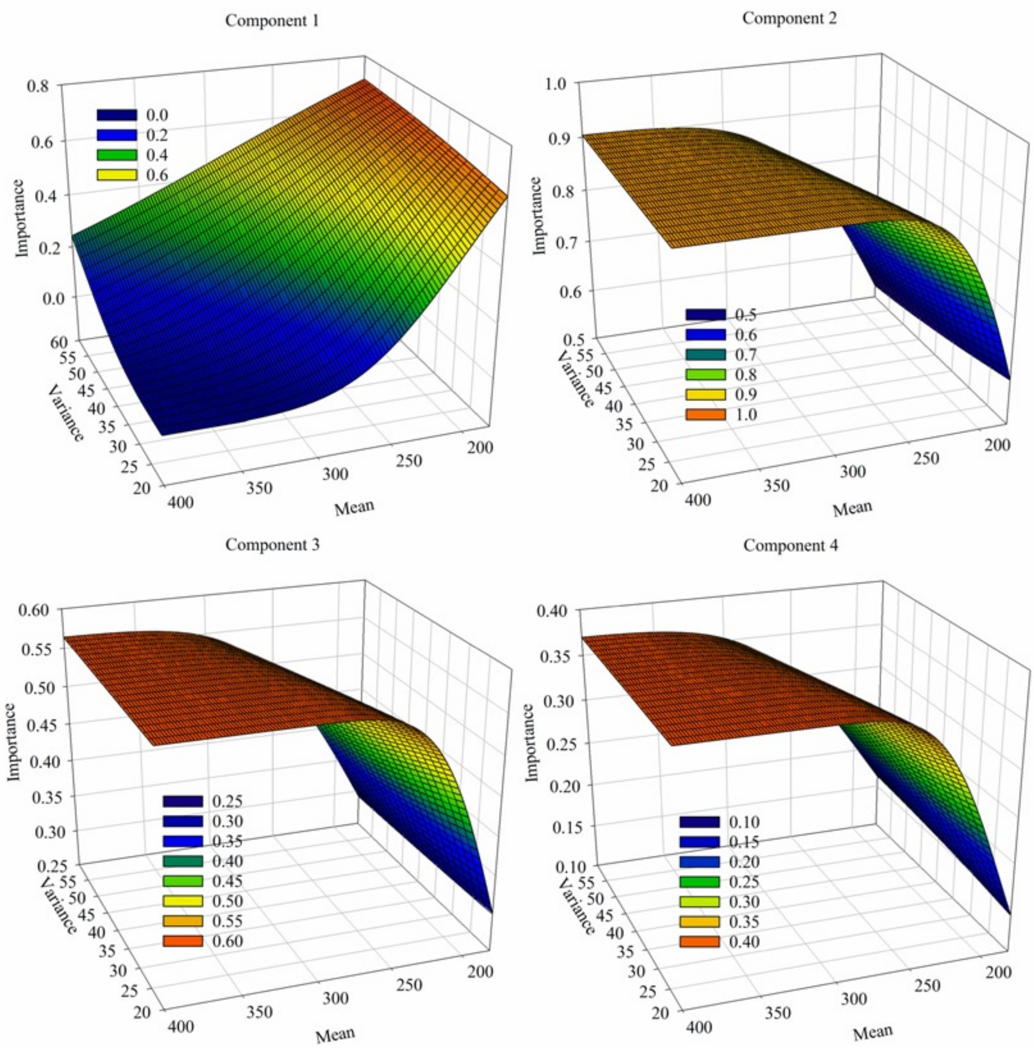


Fig. 5. The changes of SSI-based importance for case 8 in Experiment II

mean is 180. When the strength mean is known, the importance of these three components is almost unchanged. Component 1 has the lowest importance value, and component 2 has the highest importance when the strength mean of component 1 is higher.

From Fig. 5, we can find that the results of case 8 are similar to that of case 1 while the importance values are different. Because the strength variance of components 2, 3, and 4 becomes higher, the importance value of these three components is higher than that of case 1, but the change tendency is also the same with case 1.

Because the strength variance has little effect on the importance ranks, the importance ranks should be discussed with the increase of strength mean when the strength variance is determined. For each case, the strength variance is determined randomly, and the change of importance ranks is shown in Fig. 6. From Fig. 6, the SSI-based importance of component 1 decreases with the increase of the strength mean of component 1, but the importance of other components increases with the increase of the strength mean of component 1. The importance ranks of components 2, 3, and 4 are fixed no matter the parameters of component 1 are, and the importance of component 1 decreases from the highest to the lowest with the increase of strength mean. From example, the SSI-based importance of component ranks as the highest one when the strength mean is less than 205 in case 3, and the importance ranks of components is $I_1 > I_2 > I_3 > I_4$; the importance ranks of components is $I_2 > I_1 > I_3 > I_4$ when the strength mean of component is in the interval [205, 225]; the importance ranks of components is $I_2 > I_3 > I_1 > I_4$ when the strength mean of component is in the interval [225, 331]; the importance ranks of components is $I_2 > I_3 > I_4 > I_1$ when the strength mean of component is larger than 331.

Therefore, these two experiments illustrated the contributions of the proposed importance. Experiment I described how to determine the importance ranks once the parameters of components are known; Experiments II verified the monotonicity of SSI-based importance about strength mean or strength variance, and the changes of importance ranks in series systems are illustrated.

5. Component replacement method considering SSI-based importance

From the results in Section 4, the components with different combinations of strength mean and strength variance may have different SSI-based importance. For the component whose parameters can be adjusted, the SSI-based importance of this component increases with the increment of the strength mean but decreases with the increment of strength variance, There are different combinations of strength mean and variance for the component to remain the importance unchanged. However, sometimes although the importance of components with different combinations of strength mean and variance are the same while the combination may change the importance ranks of components. The SSI-based importance and the importance ranks of this component after replacing should be the same as before replacing. Therefore, a new component replacement method is developed where

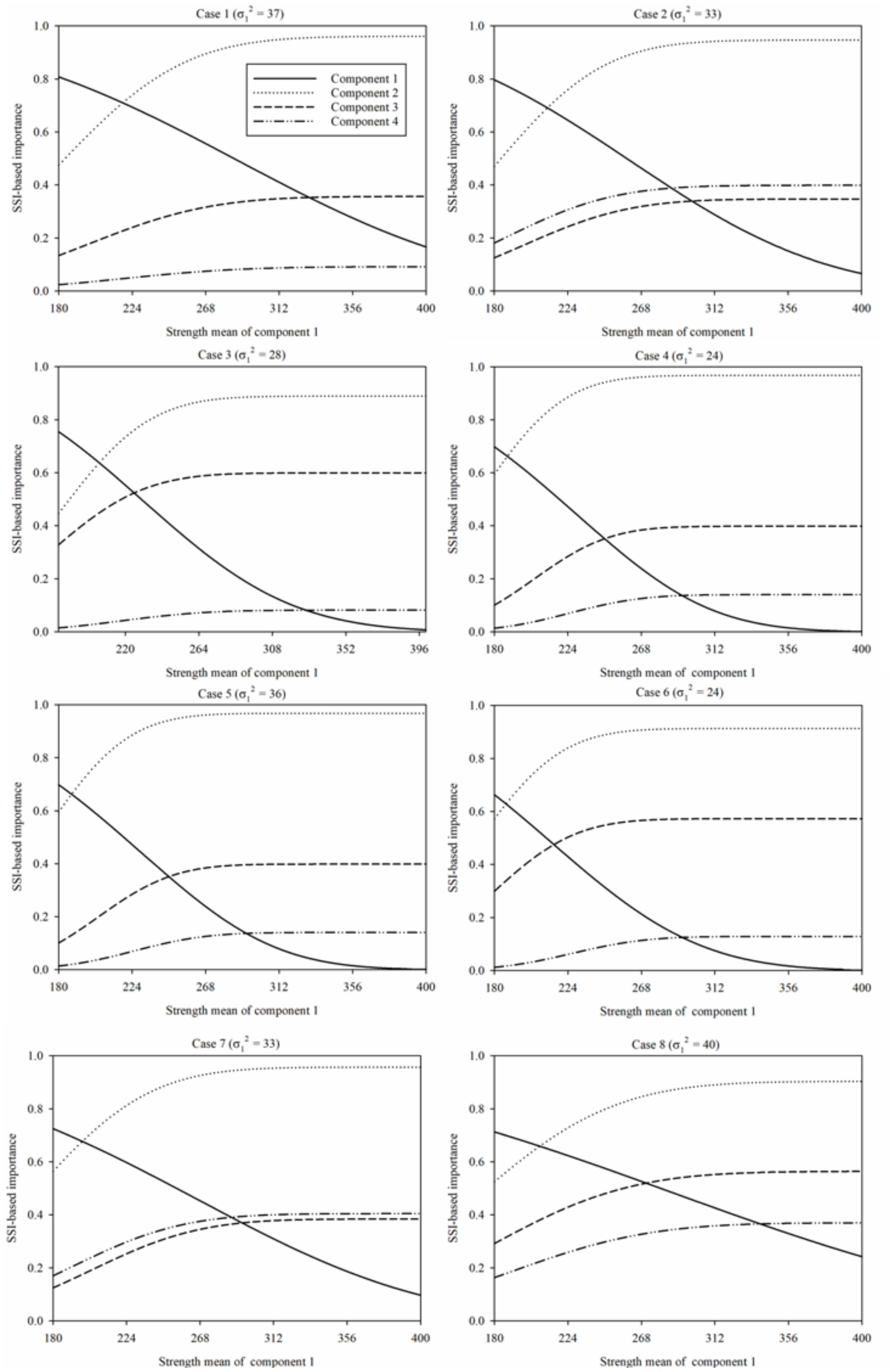


Fig. 6. The changes of SSI-based importance for case 8 in Experiment II

both the importance value and the importance ranks should remain unchanged after replacing the component.

5.1. The component replacement search algorithm

In order to find the general solutions, assuming component strength and stress distributions are normal distribution in a series system, and all parameters of distributions are known. If the component *k* can be replaced by the component *j*, the SSI-based importance of these two components should be the same. If I_k is equal to I_j, the relationship of strength mean and strength variance between components C_k and C_j based on Equation 11, which is shown as follows:

$$\int_0^\infty \exp(-\frac{(l-\mu_l)^2}{2\sigma_l^2})\exp(-\frac{(l-\mu_j)^2}{2\sigma_j^2})\left\{\prod_{i=1,i\neq j}^{n}[\int_l^\infty \exp(-\frac{(s-\mu_i)^2}{2\sigma_i^2})ds]-p\right\}dl=0 \tag{11}$$

where $p = I_k(2\pi)^{\frac{n-1}{2}}\prod_{i=1,i\neq j}^{n}\sigma_i$ is a constant, μ_j and σ_j are unknown parameters.

Considering that the difficulty of solving μ_j with σ_j explicitly, a component replacement search algorithm is proposed to determine the parameters of component after replacing, the process of component replacement considering SSI-based importance can be summarized in Fig. 7.

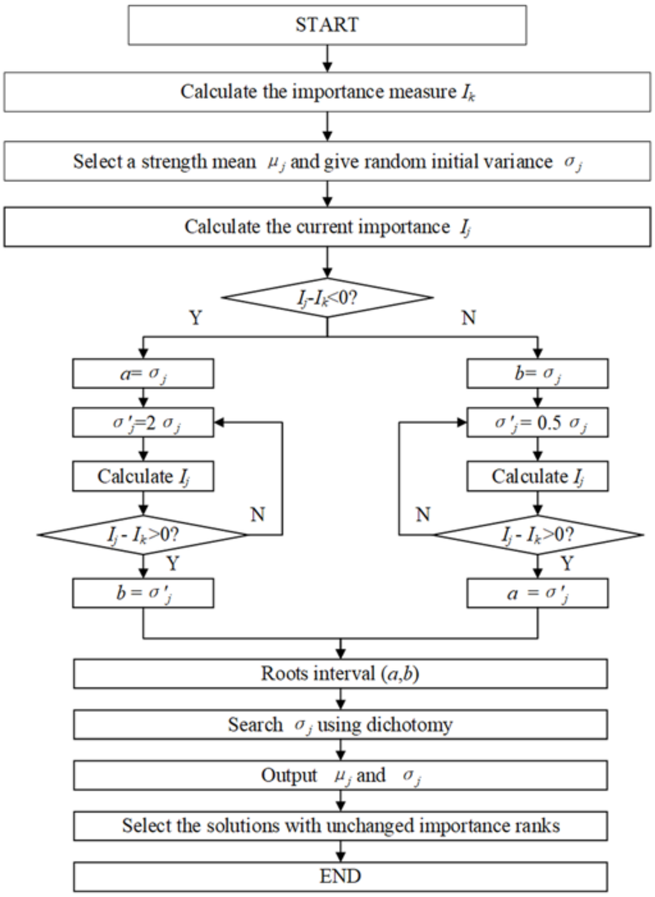


Fig. 7. The component replacement search algorithm

5.2. An example to illustrate the search process

A simple series system contains four infrastructures {C_1, C_2, C_3, C_4} with $s_1 \sim N(800,45)$, $s_2 \sim N(850,60)$, $s_3 \sim N(850,50)$, $s_4 \sim N(870,55)$, and the stress follows $l \sim N(600,40)$. The initial importance measure of component 1 is 0.9877, which can be calculated by Equation 6.

If we select

μ_1=[801,803,805,807,809,811,813,815,817,819,821,823,825,827,829] respectively, the strength variance σ_1 of component 1 can be obtained based on the component replacement search algorithm. The strength variance of component 1 and the importance ranks of components are listed in Table 4.

From Table 4, the solutions are listed with the increase of the strength mean of component 1, and the importance ranks of components are changed when the strength mean is larger than 817. According to the previous analysis, the solution that changes the importance ranking should be excluded from the available set, because the importance ranks of all components in the system should be unchanged after replacement. Sometimes, the available solution may not exist; we need to narrow the interval of strength mean. Actually, for engineering practice, other factors, such as the constraints of cost and resources should be considered to determine the available replacement solution.

Table 4. The result of the component replacement search algorithm

μ_1	σ_1	Importance ranks	Available solution
800	45.00	$I_1 > I_2 > I_4 > I_3$	✓
801	45.24	$I_1 > I_2 > I_4 > I_3$	✓
803	45.70	$I_1 > I_2 > I_4 > I_3$	✓
805	46.19	$I_1 > I_2 > I_4 > I_3$	✓
807	46.65	$I_1 > I_2 > I_4 > I_3$	✓
809	47.11	$I_1 > I_2 > I_4 > I_3$	✓
811	47.56	$I_1 > I_2 > I_4 > I_3$	✓
813	48.01	$I_1 > I_2 > I_4 > I_3$	✓
815	48.45	$I_1 > I_2 > I_4 > I_3$	✓
817	48.89	$I_1 > I_2 > I_4 > I_3$	✓
819	49.33	$I_2 > I_1 > I_4 > I_3$	×
821	49.77	$I_2 > I_1 > I_4 > I_3$	×
823	50.21	$I_2 > I_1 > I_4 > I_3$	×
825	50.63	$I_2 > I_1 > I_4 > I_3$	×
827	51.04	$I_2 > I_1 > I_4 > I_3$	×
829	51.46	$I_2 > I_1 > I_4 > I_3$	×

6. Conclusions

Some conclusions of SSI-based importance can be summarized as follows. (1) SSI-based importance of components, whose parameter changes, reduces monotonically with the increase of strength mean or increases monotonically with the increase of strength variance. (2) The strength mean has more impact on the SSI-based importance change, while the strength variance has less effect on the change of SSI-based importance. (3) The components with different combinations of strength mean and strength variance can be replaced when the importance ranks of components are unchanged after replacement. In future, the complex interdependency of components and systems should be considered, and the SSI model also can be applied to the system with more complicated structure.

Acknowledgment

The authors gratefully acknowledge the financial support for this research from the Xi'an Technological University Teaching Reform Research Project (No. 16JGY07).

References

1. Barlow R E, Proschan F. Importance of system components and fault tree events. Stochastic Processes and their applications 1975; 3(2): 153-173, http://dx.doi.org /10.1016/0304-4149(75)90013-7.
2. Barlow R E, Wu A S. Coherent systems with multi-state components. Mathematics of Operations Research 1978; 3(4): 275-281, http://dx.doi.org/10.1287/moor.3.4.275.
3. Bhattacharyya G, Johnson R A. Estimation of reliability in a multicomponent stress-strength model. Journal of the American Statistical Association 1974; 69(348): 966-970, http://dx.doi.org/10.1016/0026-2714(93)90362-3.
4. Birnbaum Z W. On the importance of different components in a multicomponent system. Washington Univ Seattle Lab of Statistical Research 1968, http://dx.doi.org/10.21236/ad0670563.
5. Borgonovo E, Apostolakis G E. A new importance measure for risk-informed decision making. Reliability Engineering & System Safety 2001; 72(2): 193-212, http://dx.doi.org/10.1016/s0951-8320(00)00108-3.
6. Do Van P, Barros A, Bérenguer C. Reliability importance analysis of Markovian systems at steady state using perturbation analysis. Reliability Engineering & System Safety 2008; 93(11): 1605-1615, http://dx.doi.org/10.1016/j.ress.2008.02.020.
7. Do Van P, Barros A, Bérenguer C. From differential to difference importance measures for Markov reliability models. European Journal of Operational Research 2010; 204(3): 513-521, http://dx.doi.org/10.1016/j.ejor.2009.11.025.
8. Dui H, Chen L, Wu S. Generalized integrated importance measure for system performance evaluation: application to a propeller plane system. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (2): 279–286, http://dx.doi.org/10.17531/ein.2017.2.16.
9. Dui H, Li S, Xing L, Liu H. System performance-based joint importance analysis guided maintenance for repairable systems. Reliability Engineering & System Safety 2019; 186: 162-175, http://dx.doi.org/10.1016/j.ress.2019.02.021.
10. Dui H, Si S, Yam R C. Importance measures for optimal structure in linear consecutive-k-out-of-n systems. Reliability Engineering & System Safety 2018; 169: 339-350, http://dx.doi.org/10.1016/j.ress.2017.09.015.
11. Fang Y, Tao W, Tee K. A new computational method for structural reliability with big data. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (1): 159–163, http://dx.doi.org/10.17531/ein.2019.1.18.
12. Fleming K. Reliability model for common mode failures in redundant safety systems. In Modeling and simulation. Volume 6, Part 1, 1975.
13. Fussell J. How to hand-calculate system reliability and safety characteristics. IEEE Transactions on Reliability 1975; 24 (3): 169-174, http://dx.doi.org/10.1109/tr.1975.5215142.
14. Gao P, Xie L. Dynamic reliability models of mechanical load-sharing parallel systems considering strength degradation of components. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 2015; 229(13): 2484-2495, http://dx.doi.org/10.1177/0954406214560000.
15. Gao P, Yan S, Xie L, Wu J. Dynamic reliability analysis of mechanical components based on equivalent strength degradation paths. Strojniški vestnik-Journal of Mechanical Engineering 2013; 59(6): 387-399, http://dx.doi.org /10.5545/sv-jme.2012.541.
16. Griffith W S. Multistate reliability models. Journal of Applied Probability 1980; 17(3): 735-744, http://dx.doi.org/10.1017/s0021900200033842.
17. Kuo W, Zhu X. Importance measures in reliability, risk, and optimization: principles and applications. John Wiley & Sons: 2012, http://dx.doi.org/10.1002/9781118314593.
18. Kuo W, Zhu X. Some recent advances on importance measures in reliability. IEEE Transactions on Reliability 2012; 61(2): 344-360, http://dx.doi.org/10.1109/tr.2011.2182394.
19. Lambert H E. Fault trees for decision making in systems analysis; Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States): 1975, http://dx.doi.org/10.2172/4169124.
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, http://dx.doi.org/10.1016/s0951-8320(03)00171-6.
21. Lewis E E. A load-capacity interference model for common-mode failures in 1-out-of-2: G systems. IEEE Transactions on Reliability 2001; 50(1): 47-51, http://dx.doi.org/10.1109/24.935017.
22. Lisnianski A, Levitin G. Multi-state system reliability: assessment, optimization and applications. World Scientific Publishing Company: 2003, http://dx.doi.org 10.1142/5221.
23. Natvig B. Measures of component importance in nonrepairable and repairable multistate strongly coherent systems. Methodology and Computing in Applied Probability 2011; 13(3): 523-547, http://dx.doi.org/10.1007/s11009-010-9170-2.
24. Pavlović P, Makajić-Nikolić D, Vujošević M. A new approach for determining the most important system components and the budget-constrained system reliability improvement. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (3): 413–419, http://dx.doi.org/10.17531/ein.2017.3.12.
25. Ramirez-Marquez J E, Coit D W. Composite importance measures for multi-state systems with multi-state components. IEEE transactions on Reliability 2005; 54(3): 517-529, http://dx.doi.org/10.1109/tr.2005.853444.
26. Ramirez-Marquez J E, Coit D W. Multi-state component criticality analysis for reliability improvement in multi-state systems. Reliability Engineering & System Safety 2007; 92(12): 1608-1619, http://dx.doi.org/10.1016/j.ress.2006.09.014.
27. Shen J C, Liu L M, Jiao A S, Yi X B. A stress-strength interference model with strength degradation following a Gamma process, Advanced Materials Research. Trans Tech Publications 2014; 945: 1196-1200, http://dx.doi.org/10.4028/www.scientific.net/amr.945-949.1196.
28. Shrestha A, Xing L, Coit D W. An efficient multistate multivalued decision diagram-based approach for multistate system sensitivity analysis. IEEE Transactions on Reliability 2010; 59(3): 581-592, http://dx.doi.org/10.1109/tr.2010.2055922.
29. Si S, Dui H, Zhao X, Zhang S, Sun S. Integrated importance measure of component states based on loss of system performance. IEEE Transactions on Reliability 2012; 61(1): 192-202, http://dx.doi.org/10.1109/TR.2011.2182394.
30. Si S, Liu M, Jiang Z, Jin T, Cai Z. System reliability allocation and optimization based on generalized Birnbaum importance measure. IEEE Transactions on Reliability 2019; 68(3):831-843, http://dx.doi.org/10.1109/tr.2019.2897026.

31. Vesely W. A time-dependent methodology for fault tree evaluation. Nuclear engineering and design 1970; 13(2): 337-360, http://dx.doi.org/10.1016/0029-5493(70)90167-6.
32. Vesely W, Davis T, Denning R, Saltos N. Measures of risk importance and their applications; Battelle Columbus Labs.: 1983, http://dx.doi.org/10.2172/5786790.
33. Wang N, Zhao J B, Jiang, Z Y, Zhang S. Reliability optimization of systems with component improvement cost based on importance measure. Advances in Mechanical Engineering 2018; 10(11): 1687814018809781, http://dx.doi.org/10.1177/1687814018809781.
34. Wang Z, Xie L. Dynamic reliability model of components under random load. IEEE Transactions on Reliability 2008; 57, (3), 474-479, http://dx.doi.org/10.1109/tr.2008.928184.
35. Wu S, Chan L Y. Performance utility-analysis of multi-state systems. IEEE Transactions on Reliability 2003; 52(1): 14-21, http://dx.doi.org/10.1109/tr.2002.805783.
36. Xie L, Zhou J, Hao C. System-level load-strength interference based reliability modeling of k-out-of-n system. Reliability Engineering & System Safety 2004; 84(3): 311-317., http://dx.doi.org/10.1016/j.ress.2003.12.003.
37. Xie L, Zhou J, Wang X. Data mapping and the prediction of common cause failure probability. IEEE Transactions on Reliability 2005; 54(2): 291-296, http://dx.doi.org/10.1109/tr.2005.847244.
38. Xie M, Shen K. On ranking of system components with respect to different improvement actions. Microelectronics Reliability 1989; 29(2): 159-164, http://dx.doi.org/10.1016/0026-2714(89)90564-7.
39. Xue J, Yang K. Upper and lower bounds of stress-strength interference reliability with random strength-degradation. IEEE Transactions on Reliability 1997; 46(1): 142-145, http://dx.doi.org/10.1109/24.589940.
40. Yang Y, Huang H, Liu Y, Zhu S P, Peng W. Reliability analysis of electrohydraulic servo valve suffering common cause failures. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16(3): 354-359.
41. Zhang S, Zhao J B, Zhu W J, Du L. Reliability optimization of linear consecutive k-out-of-n: F systems with Birnbaum importance-based quantum genetic algorithm. Advances in Mechanical Engineering 2019; 11(4): 1687814019842996, http://dx.doi.org/10.1177/1687814019842996.
42. Zhao J B, Hou P Y, Cai Z Q, Li Y, Guo P. Research of mission success importance for a multi-state repairable k-out-of-n system. Advances in Mechanical Engineering 2018; 10(2): 1687814018762208, http://dx.doi.org/10.1177/1687814018762208.
43. Zhao J, Cai Z, Si W, Zhang S. Mission success evaluation of repairable phased-mission systems with spare parts. Computers & Industrial Engineering 2019; 132: 248-259, http://dx.doi.org 10.1016/j.cie.2019.04.038.
44. Zhao J, Si S, Cai Z. A multi-objective reliability optimization for reconfigurable systems considering components degradation. Reliability Engineering & System Safety 2019; 183: 104-115, http://dx.doi.org/10.1016/j.ress.2018.11.001.
45. Zhao J, Si S, Cai Z, Su M, Wang W. Multiobjective optimization of reliability-redundancy allocation problems for serial parallel-series systems based on importance measure. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 2019;233(5): 881-897, http://dx.doi.org/10.1177/1748006X19844785.
46. Zio E, Podofillini L. Monte Carlo simulation analysis of the effects of different system performance levels on the importance of multi-state components. Reliability Engineering & System Safety 2003; 82(1): 63-73, http://dx.doi.org/10.1016/s0951-8320(03)00124-8.
47. Zuo F, Yu L, Mi J, Liu Z, Huang H. Reliability analysis of gear transmission with considering failure correlation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2015; 17 (4): 617–623, http://dx.doi.org/10.17531/ein.2015.4.19.

Guangwei YU
Yanwei DU
Li YAN
Department of Industrial Engineering
School of Mechatronic Engineering
Xi'an Technological University
38 Mailbox, No.2 Xuefuzhonglu Road, Weiyang District, Xi'an 710021, China

Fangyu REN
Luoyang Institute of Electro-optical Devices, Aviation Industry Corporation of China
No.613 Guanlin Road, Luolong District, Luoyang 471003, China

E-mails: yuguangwei@xatu.edu.cn, duyanwei@xatu.edu.cn, yanli@xatu.edu.cn, renfangyu@aliyun.com

Article citation info:
CHE H, ZENG S, GUO J. A reliability model for load-sharing k-out-of-n systems subject to soft and hard failures with dependent workload and shock effects. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 253–264, http://dx.doi.org/10.17531/ein.2020.2.8.

Haiyang CHE
Shengkui ZENG
Jianbin GUO

A RELIABILITY MODEL FOR LOAD-SHARING *K*-OUT-OF-*N* SYSTEMS SUBJECT TO SOFT AND HARD FAILURES WITH DEPENDENT WORKLOAD AND SHOCK EFFECTS

MODEL NIEZAWODNOŚCI DLA SYSTEMÓW TYPU *K*-Z-*N* Z PODZIAŁEM OBCIĄŻENIA PODLEGAJĄCYCH USZKODZENIOM PARAMETRYCZNYM I KATASTROFICZNYM, W KTÓRYCH ZACHODZI ZALEŻNOŚĆ MIĘDZY OBCIĄŻENIEM PRACĄ A SKUTKAMI OBCIĄŻEŃ LOSOWYCH

A component in a k-out-of-n system may experience soft and hard failures resulting from exposure to natural degradation and random shocks. Due to load-sharing characteristics, once a component fails, the surviving components share an increased workload, which increases their own degradation rates. Moreover, under the larger workload, random shocks may cause larger abrupt degradation increments and larger shock sizes. Therefore, the system experiences the dependent workload and shock effects (DWSEs). Such dependence will cause the load-sharing system to fail more easily, though it is often not considered in existing methods. In this paper, to evaluate the system reliability more accurately, we develop a novel reliability model for load-sharing k-out-of-n systems with DWSEs. In the model, the joint probability density function of shock effects to soft and hard failures is developed to describe the DWSEs on a component. To derive an analytical expression of system reliability with load-sharing characteristics and DWSEs, conditional probability density function is used to model the random component failure times. A load-sharing Micro-Electro-Mechanical System (MEMS) is then utilized to illustrate the effectiveness of the reliability model.

Keywords: *reliability modeling, load-sharing k-out-of-n systems, dependent workload and shock effects, degradation, random shocks.*

Element systemu k-z-n może ulegać uszkodzeniom parametrycznym i katastroficznym wynikającym z ekspozycji na naturalne procesy degradacji i obciążenia losowe. Ze względu na równomierny podział obciążenia między wszystkie elementy systemu, gdy jeden element ulega awarii, obciążenie pracą przypadające na pozostałe komponenty zwiększa się, podnosząc tempo degradacji każdego z nich. Ponadto, przy większym obciążeniu pracą, obciążenia losowe mogą powodować większe nagłe przyrosty degradacji i zwiększać rozmiary obciążeń. Mówi się wtedy o istnieniu zależności między obciążeniem pracą a skutkami obciążeń losowych (dependent workload and schock effects (DWSE). Taka zależność powoduje, że system z podziałem obciążeń łatwiej ulega uszkodzeniom. Fakt ten jest często pomijany w obecnie stosowanych metodach oceny niezawodności. W niniejszym artykule przedstawiamy nowatorski model oceny niezawodności systemów k-z-n z podziałem obciążenia i zależnością DWSE, który pozwala dokładniej ocenić niezawodność takich systemów. W modelu, opracowano wspólną funkcję gęstości prawdopodobieństwa skutków obciążeń losowych dla uszkodzeń parametrycznych i katastroficznych, która pozwala opisać zależność DWSE dla elementu systemu. Aby wyprowadzić analityczne wyrażenie niezawodności systemu z podziałem obciążenia i DWSE, do modelowania czasów losowych uszkodzeń elementów systemu wykorzystano funkcję warunkowej gęstości prawdopodobieństwa. Skuteczność modelu niezawodności zilustrowano na przykładzie układu mikroelektromechanicznego z podziałem obciążenia (MEMS).

Słowa kluczowe: *modelowanie niezawodności, systemy k-z-n z podziałem obciążenia, zależność między obciążeniem pracą a skutkami obciążeń losowych, degradacja, obciążenia losowe.*

1. Introduction

In reliability engineering, redundancy technique is widely applied to ensure a system remain functional over a long period of time. A k-out-of-n system is a typical redundant system with n components. At a minimum, it requires k operational components for the system to work normally [25, 30, 38]. Many reliability models of k-out-of-n systems have been developed, which assume that components work independently [6, 41]. However, many systems are load-sharing, such as micro-engines in a Micro-Electro-Mechanical System (MEMS) [4, 13], common buses in a common bus performance sharing system [40], and gear pair systems in a machines transmission system [37], which makes the assumption of independent components unrealistic [10, 24]. A common feature in a load-sharing system is that the workload is shared equally or unequally by the surviving components, and when a component fails, its load is distributed to the working components [32]. The increased workload on the component strongly affect its degradation rate and failure rate [11], which has been proved by many empirical studies of mechanical systems [7, 26], and battery systems [20]. Therefore, due to load-sharing characteristics, the components are stochastically dependent on each other.

Although numerous studies have explored the reliability of load-sharing systems considering the dependence among the components, they ignore the detrimental effects of random shocks on system reliability. Taghipour et al. [25] propose a periodic inspection optimization policy of a load-sharing system, where stochastic dependence among thecomponents is considered by sharing a certain amount of load. Zhang et al. [38] develop a reliability model of a load-sharing system with dependent components that equally share the system load before and after other components have failed. Ye et al. [36] develop a reliability model of a water filtering system with multiple filters, where the workload influences the filter degradation. In their model, the dominant failure type is degradation, while hard failure due to a shock will not occur. Kong et al. [12] investigate the dependence between component lifetime and load level through a link function. Although such methods successfully consider the effects of workloads on the degradation processes or failure rates, they do not consider random shocks which can accelerate the degradation process and cause sudden hard failure.

In fact, the components in a load-sharing k-out-of-n system are subject to soft and hard failures [1]. The soft failures are mainly due to degradation processes and the hard failures are due to random shocks, while the degradation processes and random shocks may be dependent [18]. For example, MEMS may be a load-sharing system where multiple micro-engines work together to perform more reliably [4, 13] and each micro-engine experiences dependent wear degradation and random shocks [18]. Based on the reliability testing experiments in [27], the dominant failure mechanism of micro-engine is determined as wear on rubbing surfaces which usually leads to either broken pin joints or seized micro-engines [29]. In addition, Tanner et al. [28] investigate shock effects on a micro-engine through shock tests, finding that random shocks will cause wear debris, which will accelerate the wear on rubbing surfaces. Moreover, the misalignment of the springs may occur and a large enough shock can result in a spring fracture. Therefore, the micro-engines will experience soft failure (i.e., wear) and hard failure (i.e., spring fracture) due to simultaneous exposure to degradation processes and random shocks.

To develop a reliability model of systems with degradation processes, random shocks, and their dependence, many literatures assume that random shocks can (a) cause abrupt degradation increases [8, 14, 15, 18, 21, 22, 39], (b) increase the degradation rate [2, 19, 31, 39], or (c) increase the hazard rate of sudden failure [3, 33]. On the other hand, random shocks may be influenced by the current degradation level. Yang et al. [34] and Che et al. [5] suggest that the occurrence of random shocks is affected by the degradation level of the system. Yang et al. [35] develop a reliability model where the magnitude of the damage caused by a shock load is correlated to the system degradation level. As reviewed above, the reliability modeling for systems with dependent degradation processes and random shocks has been thoroughly investigated, while such systems are usually series or parallel systems without load-sharing characteristics.

In literature, only a few authors analyze the reliability of a load-sharing system with dependent degradation process and random shocks, and the studies are limited in some respects. Random shocks commonly affect the components of a load-sharing system in two respects: (i) being transmitted to a shock size to components and then inducing a hard failure suddenly if the size is huge enough; (ii) creating a shock damage and then contributing to soft failure. Liu et al. [13] develop a reliability model of a load-sharing MEMS with three micro-engines subject to continuous degradation processes under a constant load or a cumulative load. In their model, degradation is the dominant failure type, while shocks only cause degradation increases and cannot lead to hard failure. In practice, a huge shock may lead to a common cause failure of the entire system [16]. Che et al. [4] develop a reliability model of a load-sharing system with dependent degradation process and random shocks. In their model, the shock effects are independent of workload, which may not be applicable in all situations.

In fact, components are subject to both workload and shock load, and both types of loads contribute to soft and hard failures. For a load-sharing system, overload is a typical shock load, such as a surge of workload for micro-engines [13] and the over discharge for battery packs [20]. When the arrival shock is an overload, its effects (i.e. the transmitted shock sizes and transmitted shock damages) depend on the resultant load of the workload and overload. After a component fails, the workload shared by each surviving component will increase, and under the high workload, the degradation rate of components will increase. In addition, the resultant load will also increase, causing the shock effects on the components to become more serious.Therefore, shock effects to soft and hard failures are dependent on the current components' workload. Load-sharing system experiences the dependent workload and shock effects (DWSEs), and the dependence is first studied to evaluate the reliability of load-sharing systems. The reliability may be overestimated without considering the dependence scenario.

Due to load-sharing characteristics and DWSEs, the degradation rate and shock effects to soft and hard failures are all dependent on the number of failed components. In addition, the failure times of the components and the arrival times of the random shocks are both stochastic. It is more practical but also presents new challenging issues to build a reliability model. In this paper, a reliability model of load-sharing systems subject to soft and hard failures with DWSEs is developed. In the model, the joint probability density function of shock effects to soft and hard failures given the number of failed components is developed to describe the DWSEs on a surviving component. In addition, the conditional probability density function of component failure time and conditional total probability formula are utilized to model the system reliability. An analytical expression is then developed to calculate system reliability, which can save much calculation time. Finally, a load-sharing MEMS is utilized as a realistic application to illustrate the effectiveness of the reliability model.

The rest of this article is organized as follows. In Section 2, we presents the system description and its assumptions. In Section 3, the model of DWSEs on a component of a load-sharing system is described in detail, and then the reliability model of a load-sharing system is proposed in Section 4. In Section 5, the reliability model is illustrated by load-sharing micro-engines in MEMS developed at Sandia. Finally, Section 6 concludes the paper and makes some suggestions for further work.

2. System specifications

In this paper, we focus on a load-sharing k-out-of-n system with n identical components sharing a certain amount of load. Each component is subject to competing soft and hard failure processes due to experiencing degradation process and random shocks simultaneously. The reliability model is built based on the following assumptions, which are adapted from recent literatures [4, 13, 18, 22, 38].

1. Random shocks arrive following a Poisson process.
2. The components in the load-sharing k-out-of-n system fail due to soft failure and hard failure. Soft failure will occur when the overall degradation is beyond the threshold value of the component. Hard failure will occur suddenly when the shock load exceeds the maximum strength of the component.
3. The system consists of n identical components. It requires at least k components being operational for the system to work properly.
4. The system load is fixed and it is shared by surviving components equally after a soft failure of a component occurs. This leads to an increased component workload and a higher degradation rate.
5. The shock load is shared by surviving components equally. Once a hard failure of a surviving component occurs when the shared shock load exceeds its maximum strength, all of the other surviving components will fail due to the same shock

load at the meanwhile, which leads to the sudden failure of the load-sharing system.

The first two assumptions are taken from [18, 22], and they are widely applied to the components and systems subject to degradation processes and random shocks simultaneously. Assumptions 3 and 4 are taken from [4, 13, 38] and are basic assumptions for equal load-sharing k-out-of-n systems. Based on Assumptions 3 and 4, Liu et al. [13]conduct the reliability analysis of a load-sharing MEMS with three micro-engines. Assumption 5 is effective when the shock is an overload such as the surge of workload for micro-engines [13] and the over discharge for battery packs [20]. Overload is a typical shock load for a load-sharing system, and the load is shared by surviving components equally. Therefore, when the shock load is large enough, the load equally shared by each surviving component exceeds its maximum strength, and the hard failures of all surviving components occur suddenly based on the assumptions that the components are identical. Consequently, such shock load will result in the sudden failure of the load-sharing system.

As shown in Fig. 1, Peng et al. [18] develop a component reliability model considering two dependent competing failure processes: soft failure due to total degradation, and hard failure due to the same shocks. For each shock j , W_j is the transmitted shock size and hard failure occurs when W_j exceeds the maximum strength D , and Y_j denotes the abrupt damage in degradation process and soft failure occurs when the overall degradation $X_S(t)$ is greater than the threshold value H .

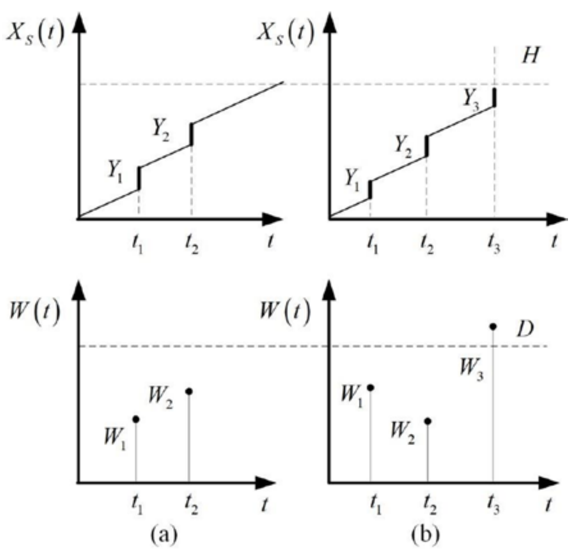


Fig. 1. Two dependent competing failure processes: (a) soft failure process, and (b) hard failure process [18]

As shown in Fig. 2, when a component in a load-sharing system fails, the system configuration changes, and the workload on each surviving component will increase, which will lead to a higher degradation rate (line a3) [17]. When a shock arrives, it can be transmitted to a shock size to the devices and induce a hard failure through line a1, and it can also create a shock damage to the devices and then contributes to soft failure through line a2. In addition, the shock load contributes to the failures together with the workload, and the workload will make shock effects more serious. Under an increased workload, the shock effects on each surviving component will be greater since the effects are caused by the resultant load of the workload and shock load. Thus the shock effects are dependent on the workload, and the load-sharing system experiences the DWSEs.

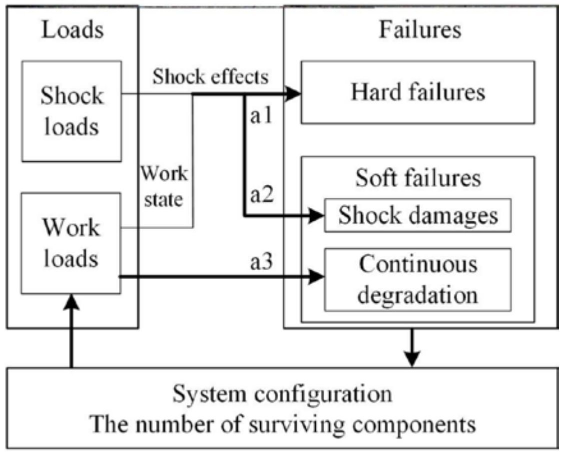


Fig. 2. The dependence analysis for the load-sharing systems

Due to the load-sharing characteristics and DWSEs, once a component fails, the workload shared on surviving components increases, resulting in that (i) the degradation process is accelerated, and (ii) the shock effects to soft and hard failures become worse. As shown in Fig. 3, for a load-sharing system with i failed components, the load of j th shock together with the shared workload will be transmitted to abrupt degradation damage Y_{ij} and shock size W_{ij} to each surviving-component in the system. When the $i+1$ th component has failed, the degradation rate increases significantly and the shock effects become more significant. As illustrated in Fig. 3, W_{i+1j} and Y_{i+1j} are greater than W_{ij} and Y_{ij} respectively.

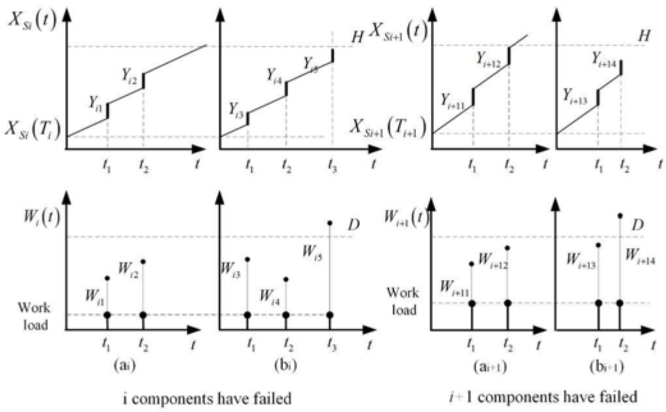


Fig. 3. Two dependent competing failure processes for a surviving component in a load-sharing system with different system configuration: (a) soft failure process, and (b) hard failure process, where $X_{Si}(t)$ is the total degradation of a component in the system with i failed components at time t .

3. Failure modeling for a component with DWSEs

In this section, we investigate the modeling for soft and hard failures of a surviving component with DWSEs. Firstly, the shock effects to soft and hard failures with DWSEs are modeled. Then, we develop the soft failure model and hard failure model of a surviving component with DWSEs.

3.1. Modeling of shocks considering DWSEs

The DWSEs on each surviving component are depicted in Fig. 4. When the j th system shock arrives with magnitude Z_j , it affects both the hard failure process and soft failure process for each C_i , where C_i is the surviving component in the load-sharing system with i failed components. Usually, the hard failure and soft failure may occur in different devices. For example, for a micro-engine, hard failure is mainly due to the spring fracture while soft failure is mainly due to wear on the rubbing surface. Therefore, Z_j can be transmitted to Z_{Hj}, which is the magnitude of the j th shock on the devices (e.g. the spring) where hard failures occur, and Z_{Sj} ,which is the magnitude of the j th shock on the devices (e.g. the rubbing surface) where soft failures occur. Z_{Hj} and Z_{Sj} are assumed to be independent, since they are applied to different devices. In addition, Z_{Hj} and Z_{Sj} apply to C_i together with the workload, and their resultant load can be transmitted to Z_{Hij} and Z_{Sij}, respectively. Then Z_{Hij} and Z_{Sij} are transmitted as shock sizes W_{ij} for the hard failure process and shock damage increments Y_{ij} for the soft failure process, respectively.

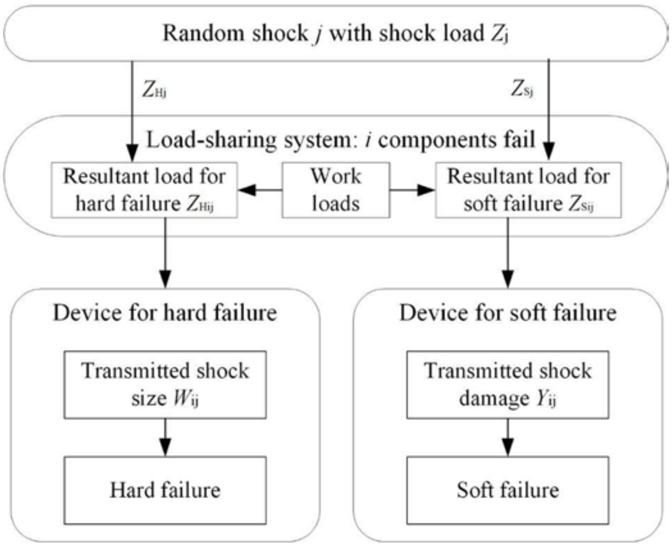


Fig. 4. The transmitted effects of system shock to the soft and hard failures

There are many ways to describe dependence characteristics in shock propagation, such as proportional correlated, additive dependent, and other more complicated models [23]. Song et al. [23] and Liu et al. [15] assume that the shock effects are linearly dependent on shock load. In this paper, a linear shock transmission model is also utilized to formulate the DWSEs. The shock size W_{ij} and the shock damage Y_{ij} to the component are transmitted linearly from Z_{Hij} and Z_{Sij} respectively, and Z_{Hij} and Z_{Sij} are also a linear function of Z_{Hj} and Z_{Sj} respectively. Then, W_{ij} and Y_{ij} can be simplified as a linear function of Z_{Hj} and Z_{Sj}, while to model the DWSEs, the transmission parameters are dependent on the workload. Based on Assumption 4, the workload shared by surviving components is only dependent on the number of failed components, i. Then, the transmission parameters are dependent on the value of i. Moreover, to consider purely random shock effects, two random terms, $\tilde{W}_{ij}$ and $\tilde{Y}_{ij}$, are present in response to a system shock, and they are not dependent on system shock loads. We assume:

$$W_{ij} = \alpha_i Z_{Hj} + \tilde{W}_{ij}, \tag{1}$$

$$Y_{ij} = \gamma_i Z_{Sj} + \tilde{Y}_{ij}, \tag{2}$$

where α_i is a transmission parameter between Z_{Hj} to the shock size for the hard failure process of C_i, and γ_i is the transmission parameter from Z_{Sj} to the shock damage for the soft failure process of C_i. The values of α_i and γ_i can be estimated from previous data, life testing, engineering judgment, and etc. As mentioned above, $\tilde{W}_{ij}$ is a random shock size contributing to C_i 's hard failure, and does not depend on Z_{Hj}. For some cases, $\tilde{W}_{ij}$ may be zero for all i or j. Similarly, $\tilde{Y}_{ij}$ is a random shock damage to soft failure process, which is not dependent on Z_{Sj}. In some special examples, $\tilde{Y}_{ij}$ may be zero, while in some other examples, the shock damage Y_{ij} is not exactly proportional to the shock magnitude Z_{Sj} and additional randomness can be introduced into Y_{ij} through $\tilde{Y}_{ij}$. Both $\tilde{W}_{ij}$ and $\tilde{Y}_{ij}$ are independent and identically distributed (i.i.d.) random variables.

The cumulative distribution function (CDF) for W_i, $F_{W_i}(w_i)$ can be derived as:

$$\begin{aligned} F_{W_i}(w_i) &= \Pr\{W_{ij} < w_i\} = \Pr\{\alpha_i Z_{Hj} + \tilde{W}_{ij} < w_i\} \\ &= \int_{z_{Hj}} \Pr\{\alpha_i z_{Hj} + \tilde{W}_{ij} < w_i\} f_{Z_{Hj}}(z_{Hj}) dz_{Hj} \end{aligned}. \tag{3}$$

Then, the probability distribution function (PDF) for W_i, $f_W(w)$ can be derived as:

$$f_{W_i}(w_i) = \int_{z_{Hj}} f_{\tilde{W}_i}(w_i - \alpha_i z_{Hj}) f_{Z_{Hj}}(z_{Hj}) dz_{Hj}. \tag{4}$$

Similarly, the CDF for Y_i, $F_{Y_i}(y_i)$ can be derived as:

$$\begin{aligned} F_{Y_i}(y_i) &= \Pr\{Y_{ij} < y_i\} = \Pr\{\gamma_i Z_{Sj} + \tilde{Y}_{ij} < y_i\} \\ &= \int_{z_{Sj}} \Pr\{\gamma_i z_{Sj} + \tilde{Y}_{ij} < y_i\} f_{Z_{Sj}}(z_{Sj}) dz_{Sj} \end{aligned}. \tag{5}$$

Then, the PDF for Y_i, $f_{Y_i}(y_i)$ can be derived as follows:

$$f_{Y_i}(y_i) = \int_{z_{Sj}} f_{\tilde{Y}_i}(y_i - \gamma_i z_{Sj}) f_{Z_{Sj}}(z_{Sj}) dz_{Sj}. \tag{6}$$

Based on Eqs. (4) and (6), the joint PDF for W and Y, $f_{W_i,Y_i}(w_i, y_i)$, is derived as:

$$f_{W_i,Y_i}(w_i, y_i) = \int_{z_{Hj}} f_{\tilde{W}_i}(w_i - \alpha_i z_{Hj}) f_{Z_{Hj}}(z_{Hj}) dz_{Hj} \times \int_{z_{Sj}} f_{\tilde{Y}_i}(y_i - \gamma_i z_{Sj}) f_{Z_{Sj}}(z_{Sj}) dz_{Sj}. \tag{7}$$

3.2. Modeling of soft and hard failures of a surviving component

Figure 1(b) shows an extreme shock model where a hard failure occurs when the shock size is beyond the maximal fracture strength D. In this paper, system shocks arrive following a Poisson process with rate λ. Based on the stress–strength model, the probability that C_i survives the applied stress from the j th system shock is:

$$P(W_{ij} < D) = F_{W_i}(D) \text{ for } j = 1, 2, ..., m_i. \tag{8}$$

Then, the probability that each C_i does not experience hard failure by time t, $P_i(NHF_t)$, is:

$$\begin{aligned}P_i\left(NHF_t\right) &= \Pr\left\{W_{01} < D, W_{02} < D, ..., W_{0m_0} < D, ..., W_{i1} < D, W_{i2} < D, ..., W_{im_i} < D\right\} \\ &= P\left\{W_{01} < D, W_{02} < D, ..., W_{0m_0} < D, ..., W_{i1} < D, W_{i2} < D, ..., W_{im_i} < D \mid m_0, m_1, ..., m_{i-1}, m_i\right\} \\ &= P\left(\bigcap_{i=0}^{N(t)} \bigcap_{j=0}^{m_i} W_{ij} < D\right) = \prod_{i=0}^{N(t)} F_{W_i}(D)^{m_i} = \prod_{i=0}^{N(t)} \left(\int_{z_H} f_{\tilde{W}_i}\left(D - \alpha_i Z_H\right) f_{Z_H}\left(z_H\right) dz_H\right)^{m_i}\end{aligned}, \quad (9)$$

where m_i is the number of shocks arrived in the time interval between T_i and T_{i+1}, as shown in Fig. 5, and T_i is the failure time of the i th component. $N(t)$ is the number of failed components by time t.

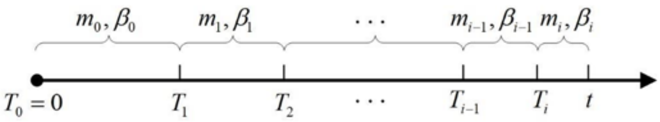


Fig. 5. Degradation process of C_i in a load-sharing system

As an example, if Z_{Hj} and $\tilde{W}_{ij}$ follow normal distributions, a more specific case for Eq. (9) can be derived as:

$$P_i\left(NHF_t\right) = \prod_{j=0}^{N(t)} \left(\int_{z_H} \Phi\left(\frac{D - \alpha_j z_H - \mu_{\tilde{W}}}{\sigma_{\tilde{W}}}\right) \varphi\left(\frac{z_H - \mu_{Z_H}}{\sigma_{Z_H}}\right) dz_H\right)^{m_j}, \quad (10)$$

where $\Phi(\bullet)$ and $\varphi(\bullet)$ are the CDF and PDF of a standard normally distributed variable, respectively.

As shown in Fig. 1 part (a), the soft failure of a component occurs when the overall degradation is greater than the threshold value H [18, 22, 23]. The overall degradation, $X_S(t)$, is affected by the load-sharing characteristics and DWSEs, and is accumulated by continual degradation and cumulative abrupt damage caused by shocks. According to the degradation models in many literatures [5, 18, 19], we also assume a linear degradation path to accumulate continual degradation, $X(t) = \mu + \beta t + \varepsilon$, where μ is constant and represents the initial component degradation, β is a random variable and represents degradation rate, and ε is a random error term and follows a normal distribution, $\varepsilon \sim N(0, \sigma^2)$.

Each C_i in a load-sharing system will experience the system configuration changing from no failed components to i failed components. Therefore, as shown in Fig. 5, its degradation rate will change from β_0 to β_i step by step, where β_i is the current degradation rate of each C_i. The degradation rate β is influenced by the workload on the component, and due to load-sharing characteristics, the following inequalities $\beta_i > \beta_{i-1} > ... > \beta_1 > \beta_0$ will exist. In addition, β_{i-1} will increase to β_i when the i th component fails. The value of $\boldsymbol{\beta}$ can be estimated through accelerated degradation test [17].

Therefore the total degradation of C_i by time t is denoted as:

$$X(t) = \begin{cases} \mu + \sum_{l=0}^{i-1} \beta_l\left(T_{l+1} - T_l\right) + \beta_i\left(t - T_i\right) + \varepsilon, & \text{if } i = N(t) \geq 1 \\ \mu + \beta_0 t + \varepsilon & \text{if } N(t) = 0 \end{cases}. \quad (11)$$

Moreover, a shock will cause a damage increment to the degradation process Y_{ij}, and a cumulative shock model is used to determine accumulated shock damage increments. The cumulative degradation damage increments $S(t)$ caused by shocks until time t can be derived as:

$$S(t) = \begin{cases} \sum_{i=0}^{N(t)} \sum_{j=1}^{m_i} Y_{ij} = \sum_{i=0}^{N(t)} \sum_{j=1}^{m_i} \left(\tilde{Y}_{ij} + \gamma_i Z_{Sj}\right), & \text{if } \sum_{i=0}^{N(t)} m_i > 0 \\ 0 & \text{if } \sum_{i=0}^{N(t)} m_i = 0 \end{cases}. \quad (12)$$

Therefore, the total degradation accumulated by both continual degradation and cumulative abrupt damages can be expressed as $X_S(t) = X(t) + S(t)$. Then the probability that the overalldegradation at time t is less than the threshold value H can be derivedas $P(X_S(t) < H) = P(X(t) + S(t) < H)$.

Conditioning on the times $T_1, T_2, ..., T_i$ and shock numbers $m_0, m_1, ..., m_{i-1}$, the probability that no soft failure will occur on C_i at time t can be derived as:

$$\begin{aligned}&P_i(NSF(t) \mid m_0, m_1, ..., m_{i-1}, T_1, T_2, ..., T_i) \\ &= \sum_{m_i=0}^{\infty} P_i(X_S(t) < H, m_i \mid m_0, m_1, ..., m_{i-1}, T_1, T_2, ..., T_i) \\ &= \sum_{m_i=0}^{\infty} P_i(X_S(t) < H \mid m_i, m_0, m_1, ..., m_{i-1}, T_1, T_2, ..., T_i) P(m_i \mid, m_0, m_1, ..., m_{i-1}, T_1, T_2, ..., T_i)\end{aligned}, \quad (13)$$

where $P(m_i \mid m_0, m_1, ..., m_{i-1}, T_1, T_2, ..., T_i)$ is the conditional probability that m_i shocks arrive in the time interval between T_i and t given the component failure times $T_1, T_2, ..., T_i$ and the shock numbers $m_0, m_1, ..., m_{i-1}$. The conditional probability is only dependent on T_i and t due to the characteristics of Poisson process and can be simplified as:

$$P(m_i \mid m_0, m_1, ..., m_{i-1}, T_1, T_2, ..., T_i) = P(m_i \mid T_i) = \frac{\exp(-\lambda\left(t - T_i\right))(\lambda\left(t - T_i\right))^{m_i}}{m_i!}. \quad (14)$$

Then Eq. (13) can be rewritten as:

$$\begin{aligned}&P_i(NSF(t) \mid m_0, m_1, ..., m_{i-1}, T_1, T_2, ..., T_i) \\ &= \sum_{m_i=0}^{\infty} P(\mu + \sum_{l=0}^{i-1} \beta_l\left(T_{l+1} - T_l\right) + \beta_i\left(t - T_i\right) + \sum_{l=0}^{i} \sum_{j=1}^{m_l} Y_{lj} + \varepsilon < H \mid m_i) \frac{\exp(-\lambda\left(t - T_i\right))(\lambda\left(t - T_i\right))^{m_i}}{m_i!}.\end{aligned} \quad (15)$$

Furthermore, if $f_{Z_{Sj}}^{\langle m \rangle}$ is considered to be the PDFof the sum of m i.i.d. Z_{Sj} variables,then $P_i\left(X_S(t) < H \mid m_i, m_0, m_1, ..., m_{i-1}, T_1, T_2, ..., T_i\right)$ in Eq. (13) can be derived to amore specific expression based on a convolution integral:

$$\begin{aligned}&P_i(X_S(t) < H \mid m_i, m_0, m_1, ..., m_{i-1}, T_1, T_2, ..., T_i) \\ &= \int_{u_1} ... \int_{u_i} P(\mu + \sum_{l=0}^{i-1} \beta_l\left(T_{l+1} - T_l\right) + \beta_i\left(t - T_i\right) + \sum_{l=0}^{i} \sum_{j=1}^{m_l} \tilde{Y}_{lj} + \sum_{l=0}^{i} \sum_{j=1}^{m_l} \gamma_l z_{Sj} + \varepsilon < H \mid \sum_{j=1}^{m_l} z_{Sj} = u_l) \\ &\times f_{Z_{Sj}}^{\langle m_1 \rangle}\left(u_1\right) ... f_{Z_{Sj}}^{\langle m_i \rangle}\left(u_i\right) du_1 ... du_i \\ &= \int_{u_1} ... \int_{u_i} P(\mu + \sum_{l=0}^{i-1} \beta_l\left(T_{l+1} - T_l\right) + \beta_i\left(t - T_i\right) + \sum_{l=0}^{i} \sum_{j=1}^{m_l} \tilde{Y}_{lj} + \sum_{l=0}^{i} \gamma_l u_l + \varepsilon < H) \\ &\times f_{Z_{Sj}}^{\langle m_1 \rangle}\left(u_1\right) ... f_{Z_{Sj}}^{\langle m_i \rangle}\left(u_i\right) du_1 ... du_i\end{aligned}. \quad (16)$$

Conditioning on that $\sum_{l=0}^{i}\sum_{j=1}^{m_l}\tilde{Y}_{lj}=y$, if the PDF of the purely random variables $\tilde{Y}_{ij}$ is $f_{\tilde{Y}}(y)$ for all i and j, Eq. (16) can be derived as:

$$\begin{aligned}
&P_i(X_S(t)<H\mid m_i,m_0,m_1,\dots,m_{i-1},T_1,T_2,\dots,T_i)\\
&=\int_{u_1}\dots\int_{u_i}P(\mu+\sum_{l=0}^{i-1}\beta_l\left(T_{l+1}-T_l\right)+\beta_i\left(t-T_i\right)+\sum_{l=0}^{i}\sum_{j=1}^{m_l}\tilde{Y}_{lj}+\sum_{l=0}^{i}\gamma_l u_l+\varepsilon<H)\\
&\times f_{Z_{Sj}}^{\langle m_1\rangle}(u_1)\dots f_{Z_{Sj}}^{\langle m_i\rangle}(u_i)du_1\dots du_i\\
&=\int_{u_1}\dots\int_{u_i}\left(\int_y P\left(\begin{array}{l}\mu+\sum_{l=0}^{i-1}\beta_l\left(T_{l+1}-T_l\right)+\beta_i\left(t-T_i\right)\\+\sum_{l=0}^{i}\sum_{j=1}^{m_l}\tilde{Y}_{lj}+\sum_{l=0}^{i}\gamma_l u_l+\varepsilon<H\mid\sum_{l=0}^{i}\sum_{j=1}^{m_l}\tilde{Y}_{lj}=y\end{array}\right)f_{\tilde{Y}}^{\left\langle\sum_{l=0}^{i}m_l\right\rangle}(y)dy\right)\\
&\times f_{Z_{Sj}}^{\langle m_1\rangle}(u_1)\dots f_{Z_{Sj}}^{\langle m_i\rangle}(u_i)du_1\dots du_i\\
&=\int_{u_1}\dots\int_{u_i}\left(\int_y P(\mu+\sum_{l=0}^{i-1}\beta_l\left(T_{l+1}-T_l\right)+\beta_i\left(t-T_i\right)+y+\sum_{l=0}^{i}\gamma_l u_l+\varepsilon<H)f_{\tilde{Y}}^{\left\langle\sum_{l=0}^{i}m_l\right\rangle}(y)dy\right)\\
&\times f_{Z_{Sj}}^{\langle m_1\rangle}(u_1)\dots f_{Z_{Sj}}^{\langle m_i\rangle}(u_i)du_1\dots du_i\quad.
\end{aligned}\tag{17}$$

As an example, if β_i, $\tilde{Y}_{ij}$, and Z_{Sj} all follow normal distributions, $\sum_{l=0}^{i}\gamma_l u_l$ is also a normal distribution $\sum_{l=0}^{i}\gamma_l u_l\sim N\left(\sum_{l=0}^{i}\gamma_l m_l\mu_{Z_S},\sum_{l=0}^{i}\gamma_l^2 m_l\sigma_{Z_S}^2\right)$. Then, a more specific case for Eq. (17) can be derived as:

$$\begin{aligned}
&P_i(X_S(t)<H\mid m_i,m_0,m_1,\dots,m_{i-1},T_1,T_2,\dots,T_i)\\
&=\int_u\Phi\left(\frac{H-\left(\mu+\sum_{l=0}^{i-2}\mu_{\beta_l}\left(T_{l+1}-T_l\right)+\mu_{\beta_{i-1}}\left(T_i-T_{i-1}\right)+\sum_{l=0}^{i}m_l\mu_{\tilde{Y}}+u\right)}{\sqrt{\sum_{j=0}^{i-2}\sigma_{\beta_j}^2\left(T_{j+1}-T_j\right)^2+\sigma_{\beta_{i-1}}^2\left(t-T_i\right)^2+\sum_{l=0}^{i}m_l\sigma_{\tilde{Y}}^2+\sigma^2}}\right)\varphi\left(\frac{u-\sum_{l=0}^{i}\gamma_l m_l\mu_{Z_S}}{\sqrt{\sum_{l=0}^{i}\gamma_l^2 m_l\sigma_{Z_S}}}\right)du\ .
\end{aligned}\tag{18}$$

Then, Eq. (15) can be expressed as:

$$\begin{aligned}
&P_i(NSF(t)\mid m_0,m_1,\dots,m_{i-1},T_1,T_2,\dots,T_i)\\
&=\sum_{m_i=0}^{\infty}\int_u\Phi\left(\frac{H-\left(\mu+\sum_{l=0}^{i-2}\mu_{\beta_l}\left(T_{l+1}-T_l\right)+\mu_{\beta_{i-1}}\left(T_i-T_{i-1}\right)+\sum_{l=0}^{i}m_l\mu_{\tilde{Y}}+u\right)}{\sqrt{\sum_{j=0}^{i-2}\sigma_{\beta_j}^2\left(T_{j+1}-T_j\right)^2+\sigma_{\beta_{i-1}}^2\left(t-T_i\right)^2+\sum_{l=0}^{i}m_l\sigma_{\tilde{Y}}^2+\sigma^2}}\right)\\
&\varphi\left(\frac{u-\sum_{l=0}^{i}\gamma_l m_l\mu_{Z_S}}{\sqrt{\sum_{l=0}^{i}\gamma_l^2 m_l\sigma_{Z_S}}}\right)du\times\frac{\exp(-\lambda\left(t-T_i\right))(\lambda\left(t-T_i\right))^{m_i}}{m_i!}\quad.
\end{aligned}\tag{19}$$

4. Reliability modeling for a load-sharing system

As pointed out in the previous section, the reliability model of a load-sharing system should consider DWSEs, load-sharing characteristics, and dependent competing failures (i.e., soft failures and hard failures). Based on Assumption 5, once a component hard failure occurs, the hard failures of the other surviving components will occur at the meanwhile, and the load-sharing system fails immediately. On the other hand, the soft failures of the surviving components will occur one by one due to the randomness of the degradation process. When a soft failure of component occurs, the component fails to function properly, and then the workload shared by each surviving component will increase, leading to a higher degradation rate of the surviving components and the more serious shock effects to soft and hard failures. Then, the failure of the load-sharing system will be accelerated.

To analyze the reliability of such load-sharing system, the difficulties lie in the stochastic nature of the soft failure times of components and arrival times of the random shocks and that the shock effects are dependent on the workload. In this section, the conditional probability density function and the joint probability density function are utilized to develop reliability models for a load-sharing system.

Given the times $T_1,T_2,\dots,T_{i-1}$ and the shocks numbers $m_0,m_1,\dots,m_{i-1}$, the conditional probability density function of the soft failure time of the i th component can be obtained as:

$$\begin{aligned}
&f_i\left(T_i\mid m_0,m_1,\dots,m_{i-1},T_1,T_2,\dots,T_{i-1}\right)=\frac{-dP(X_S(T_i)<H\mid m_0,m_1,\dots,m_{i-1},T_1,T_2,\dots,T_{i-1})}{dT_i}\\
&=\int_u-\varphi\left(\frac{H-\left(\mu+\sum_{l=0}^{i-2}\mu_{\beta_l}\left(T_{l+1}-T_l\right)+\mu_{\beta_{i-1}}\left(T_i-T_{i-1}\right)+\sum_{l=0}^{i-1}m_l\mu_{\tilde{Y}}+u\right)}{\sqrt{\sum_{j=0}^{i-2}\sigma_{\beta_j}^2\left(T_{j+1}-T_j\right)^2+\sigma_{\beta_{i-1}}^2\left(T_i-T_{i-1}\right)^2+\sum_{l=0}^{i-1}m_l\sigma_{\tilde{Y}}^2+\sigma^2}}\right)\\
&\times\left(\frac{\begin{array}{l}-\mu_{\beta_{i-1}}\left(\sum_{j=0}^{i-2}\sigma_{\beta_j}^2\left(T_{j+1}-T_j\right)^2+\sigma_{\beta_{i-1}}^2\left(T_i-T_{i-1}\right)^2+\sum_{l=0}^{i-1}m_l\sigma_{\tilde{Y}}^2+\sigma^2\right)\\-\sigma_{\beta_{i-1}}^2\left(T_i-T_{i-1}\right)\left(H-\left(\mu+\sum_{l=0}^{i-2}\mu_{\beta_l}\left(T_{l+1}-T_l\right)+\mu_{\beta_{i-1}}\left(T_i-T_{i-1}\right)+\sum_{l=0}^{i-1}m_l\mu_{\tilde{Y}}+u\right)\right)\end{array}}{\left(\sum_{j=0}^{i-2}\sigma_{\beta_j}^2\left(T_{j+1}-T_j\right)^2+\sigma_{\beta_{i-1}}^2\left(T_i-T_{i-1}\right)^2+\sum_{l=0}^{i-1}m_l\sigma_{\tilde{Y}}^2+\sigma^2\right)^{\frac{3}{2}}}\right)\\
&\times\varphi\left(\frac{u-\sum_{l=0}^{i-1}\gamma_l m_l\mu_{Z_S}}{\sqrt{\sum_{l=0}^{i-1}\gamma_l^2 m_l\sigma_{Z_S}}}\right)du\quad.
\end{aligned}\tag{20}$$

The conditional probability that C_i fails in an infinitesimal interval dT_i can be expressed as $f_i\left(T_i\mid m_0,m_1,\dots,m_{i-1},T_1,T_2,\dots,T_{i-1}\right)dT_i$. On that condition, for a load-sharing system, there are $n-i+1$ surviving components, so, the probability that the i th component fails at time T_i is:

$$P\left(T_i\mid m_0,m_1,\dots,m_{i-1},T_1,T_2,\dots,T_{i-1}\right)=\left(n-i+1\right)f_i\left(T_i\mid m_0,m_1,\dots,m_{i-1},T_1,T_2,\dots,T_{i-1}\right)dT_i\ .\tag{21}$$

The probability that $n-i$ components in the system work reliably with no soft failure, $P_{Si}\left(NSF_t\right)$, can be derived as:

$$\begin{aligned}
&P_{Si}\left(NSF_t\right)=\frac{n!}{\left(n-i\right)!}\sum_{m_i=0}^{\infty}\dots\sum_{m_0=0}^{\infty}\int_0^t f_1(T_1\mid m_0)P(m_0)\int_{T_1}^t f_1(T_2\mid m_1,m_0,T_1)P(m_1\mid m_0,T_1)\\
&\dots\int_{T_{i-1}}^t f_i(T_i\mid m_{i-1},m_0,m_1,\dots,m_{i-2},T_1,T_2,\dots,T_{i-1})P(m_{i-1}\mid m_0,m_0,m_1,\dots,m_{i-2},T_1,T_2,\dots,T_{i-1})\\
&\times P(X_S(t)<H\mid m_i,m_0,m_1,\dots,m_{i-1},T_1,T_2,\dots,T_i)^{n-i}\\
&\times P(m_i\mid m_0,m_1,\dots,m_{i-1},T_1,T_2,\dots,T_i)dT_idT_{i-1}\dots dT_1\quad.
\end{aligned}\tag{22}$$

The load-sharing system is subject to sudden failures due to random shocks. For a certain system configuration, each component is exposed to the same shock size. Then, once a hard failure occurs for a component, all of the other surviving components will fail due to the same shock at the meanwhile, which leads to a sudden failure of the load-sharing system. The probability that the load-sharing system with i failed components does not experience the sudden failure $P_{Si}\left(NHF_t\right)$ can be presented as:

$$P_{Si}\left(NHF_t\right)=\prod_{j=0}^{i}\left(\int_{z_H}\Phi\left(\frac{D-\alpha_j Z_H-\mu_W}{\sigma_W}\right)\varphi\left(\frac{z_H-\mu_{Z_H}}{\sigma_{Z_H}}\right)dz_H\right)^{m_j}. \quad (23)$$

Therefore,the probability that there are $n-i$ components working reliably R_{Si} can be denoted as:

$$\begin{aligned}R_{Si}&=P_{Si}\left(NSF_t\right)\times P_{Si}\left(NHF_t\right)\\&=\frac{n!}{(n-i)!}\sum_{m_i=0}^{\infty}\cdots\sum_{m_0=0}^{\infty}\int_0^t f_1(T_1\mid m_0)\int_{T_1}^t f_2(T_2\mid m_0,m_1,T_1)\ldots\int_{T_{i-1}}^t f_i(T_i\mid m_0,m_1,\ldots,m_{i-1},T_1,T_2,\ldots,T_{i-1})\\&\times\left(\int_u\Phi\left(\frac{H-\left(\mu+\sum_{l=0}^{i-1}\mu_{\beta_l}\left(T_{l+1}-T_l\right)+\mu_{\beta_i}\left(t-T_i\right)+\sum_{l=0}^{i}m_l\mu_{\tilde{Y}}+u\right)}{\sqrt{\sum_{j=0}^{i-1}\sigma_{\beta_j}^2\left(T_{j+1}-T_j\right)^2+\sigma_{\beta_i}^2\left(t-T_i\right)^2+\sum_{l=0}^{i}m_l\sigma_{\tilde{Y}}^2+\sigma^2}}\right)\times\varphi\left(\frac{u-\sum_{l=0}^{i}\gamma_l m_l\mu_{Z_S}}{\sqrt{\sum_{l=0}^{i}\gamma_l^2 m_l\sigma_{Z_S}}}\right)du\right)^{n-i}\\&\times\frac{\exp(-\lambda(t-T_i))(\lambda(t-T_i))^{m_i}}{m_i!}\times\prod_{l=2}^{i}\frac{\exp(-\lambda(T_l-T_{l-1}))(\lambda(T_l-T_{l-1}))^{m_{l-1}}}{m_{l-1}!}\\&\times\frac{\exp(-\lambda T_1)(\lambda T_1)^{m_0}}{m_0!}dT_i dT_{i-1}\ldots dT_1\times\prod_{j=0}^{i}\left(\int_{z_H}\Phi\left(\frac{D-\alpha_j Z_H-\mu_W}{\sigma_W}\right)\varphi\left(\frac{z_H-\mu_{Z_H}}{\sigma_{Z_H}}\right)dz_H\right)^{m_j}.\end{aligned} \quad (24)$$

A special case occurs when there is no soft fault by time t, and the probability of this occurrence can be obtained using

$$\begin{aligned}R_{S0}=&\sum_{m_0=0}^{\infty}\left(\int_u\Phi\left(\frac{H-\left(\mu+\mu_{\beta_0}t+m_0\mu_{\tilde{Y}}+u\right)}{\sqrt{\sigma_{\beta_0}^2t^2+m_0\sigma_{\tilde{Y}}^2+\sigma^2}}\right)\times\varphi\left(\frac{u-\gamma_0 m_0\mu_{Z_S}}{\sqrt{\gamma_0^2 m_0\sigma_{Z_S}}}\right)du\right)^n\\&\times\frac{\exp(-\lambda t)(\lambda t)^{m_0}}{m_0!}\times\left(\int_{z_H}\Phi\left(\frac{D-\alpha_j Z_H-\mu_W}{\sigma_W}\right)\varphi\left(\frac{z_H-\mu_{Z_H}}{\sigma_{Z_H}}\right)dz_H\right)^{m_0}.\end{aligned} \quad (25)$$

For a load-sharing k-out-of-n system, there must be at least k surviving components working for successful operation. Consequently, its reliability $R(t)$ can be determined according to:

$$R(t)=\sum_{i=0}^{n-k}R_{si}. \quad (26)$$

Moreover,the developed reliability model can be easily extend to a load-sharing parallel system through assigning the value k to 1.

5. Case study

In most cases, MEMS is a load-sharing system where multiple micro-engines work together to perform more reliably [4, 13]. The system has been widely applied to many intelligent mechatronic systems, and its reliability has been studied extensively [9, 27]. A 2-out-of-4 load-sharing MEMS with four identical micro-engines is utilized as a realistic application to illustrate the effectiveness and modeling capabilities of the proposed model in this paper. The micro-engine consists of multiple orthogonal linear comb drive actuators mechanically connected to a rotating gear. By applying voltages, the linear displacement of the comb drives is transformed into the circular motion of the gear via a pin joint.

Based on the results of the reliability tests from [28], each micro-engine is subject to two dependent competing failure processes: hard failure due to the spring fracture caused by the huge shock; and

Table 1. The corresponding parameters for the reliability analysis of micro-engines

Parameters	Value	Sources
H	0.00125 μm³	[26]
D	1.5 GPa	[26]
μ	0	[26]
β_i	$\beta_0\sim N\left(\mu_{\beta_0},\sigma_{\beta_0}^2\right)$, $\mu_{\beta_0}=8.4823\times10^{-9}\mu m^3$, $\sigma_{\beta_0}=6.0016\times10^{-10}\mu m^3$ $\beta_1\sim N\left(\mu_{\beta_1},\sigma_{\beta_1}^2\right)$, $\mu_{\beta_1}=1.2\times10^{-8}\mu m^3$, $\sigma_{\beta_1}=6.0\times10^{-9}\mu m^3$ $\beta_2\sim N\left(\mu_{\beta_2},\sigma_{\beta_2}^2\right)$, $\mu_{\beta_2}=2.0\times10^{-8}\mu m^3$, $\sigma_{\beta_2}=9.0\times10^{-9}\mu m^3$	β_0: [26]; β_1 and β_2 : Asumption
ε	$\varepsilon\sim N\left(0,\sigma^2\right)$, $\sigma=10^{-10}\mu m^3$	[19]
λ	2.5×10^{-5}	[19]
Z_H	$Z_H\sim N\left(\mu_{Z_H},\sigma_{Z_H}^2\right)$, $\mu_{Z_H}=1.2$GPa , $\sigma_{Z_H}=0.2$GPa	[19]
Z_S	$Z_S\sim N\left(\mu_{Z_S},\sigma_{Z_S}^2\right)$, $\mu_{Z_S}=1.2$GPa , $\sigma_{Z_S}=0.2$GPa	[19]
$\tilde{Y}$	$\tilde{Y}\sim N\left(\mu_{\tilde{Y}},\sigma_{\tilde{Y}}^2\right)$, $\mu_{\tilde{Y}}=1.0\times10^{-9}\mu m^3$, $\sigma_{\tilde{Y}}=1.0\times10^{-10}\mu m^3$	Asumption
$\tilde{W}$	$\tilde{W}\sim N\left(\mu_{\tilde{W}},\sigma_{\tilde{W}}^2\right)$, $\mu_{\tilde{W}}=0$, $\sigma_{\tilde{W}}=1.0\times10^{-3}$	Asumption
γ_i	$\gamma_0=2\times10^{-5}$, $\gamma_1=1\times10^{-4}$, $\gamma_2=5\times10^{-4}$	Asumption
α_i	$\alpha_0=0.9$, $\alpha_1=1.0$, $\alpha_2=1.2$	Asumption

soft failure due to the continual wear process on rubbing surfaces and substantial wear debris caused by shocks, which usually results in a seized micro-engine or a broken pin joint and then the micro-engine is deemed to have failed. Due to load-sharing characteristics, the degradation rate of a surviving micro-engine will increase after a micro-engine failure because of its increased shared workload. In addition, the shock effects to the soft failure and hard failure are dependent on the workload, and the load-sharing MEMS experiences DWSEs. In the application, both transmission parameters α_i and γ_i will increase after a component fails. Based on the data in [4, 18, 19, 26], along with some reasonable assumptions, the corresponding parameters for the reliability model are shown in Table 1.

5.1. Results analysis

Using Eq. (3), the probability that MEMS with zero failed micro-engine, one failed micro-engine, and two failed micro-engines will survive a shock with no sudden failure is calculated as 99.02%, 93.32%, and 59.87%, respectively. Based on Eqs. (24) and (25), for the MEMS, the probability that four micro-engines work reliably (i.e., R_{S0}), three micro-engines work reliably (i.e., R_{S1}) , and two micro-engines work reliably (i.e., R_{S2}) , is shown in Fig. 6, respectively. It can be seen that R_{S0} is very close to 1 and changes slowly at first, and then its declination increases sharply at approximately $t=8\times10^4$. In addition, R_{S1} and R_{S2} are very close to 0 before the time $t=8\times10^4$, and then increase to the peak point, and finally decrease to 0. This is mainly due to the following reasons:

(1) For each micro-engine, the arrival shocks will cause abrupt wear debris from the contact surface of the pin joint and the gear. At the beginning, the wear extent is far from the failure threshold *H*, and soft failure rarely occurs and hard failure is the main failure mode. Moreover, the arrival shocks are relatively few, and then the probability that hard failure occurs is also small. Therefore, before the time $t=8\times10^4$, R_{S0} is close to 1 and decreases slowly, and R_{S1} and R_{S2} are very close to 0.

(2) As time goes on, the wear extent increases and may approach the threshold. Therefore, for micro-engines, soft failure is more likely to occur when is large, which leads to the sharp decrease of R_{S0}. The micro-engines will fail in order due to wear. When the first micro-engine fails, R_{S1} will increase, and when the second micro-engine fails, R_{S2} will increase. Finally, they will decrease to 0 since all micro-engines will fail *t* when is large enough.

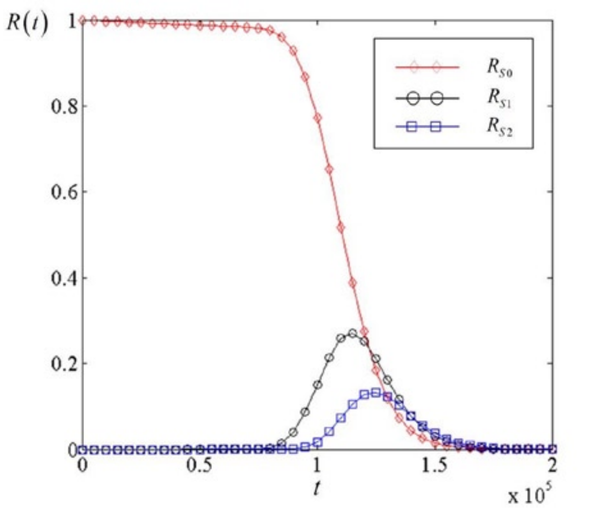


Fig. 6. The values of R_{S0}, R_{S1}, and R_{S2} for a 2-out-of-4 load-sharing system

By setting different values of the parameters in Eq. (24), we can calculate the system reliability without load-sharing characteristics and DWSEs (i.e., $\beta_0=\beta_1=\beta_2\sim N\left(\mu_{\beta_0},\sigma^2_{\beta_0}\right)$, $\gamma_0=\gamma_1=\gamma_2=2\times10^{-5}$ and $\alpha_0=\alpha_1=\alpha_2=0.9$), and only with load-sharing characteristics (i.e., $\gamma_0=\gamma_1=\gamma_2=2\times10^{-5}$ and $\alpha_0=\alpha_1=\alpha_2=0.9$). For a MEMS without load-sharing characteristics and DWSEs, the probabilities that four micro-engines work reliably (denoted by R'_{S0}), three micro-engines work reliably (denoted by R'_{S1}) , and two micro-engines work reliably (denoted by R'_{S2}), are plotted in Fig. 7(a). For a MEMS with only load-sharing characteristics, the probabilities that four micro-engines work reliably (denoted by R''_{S0}), three micro-engines work reliably (denoted by R''_{S1}), and two micro-engines work reliably (denoted by R''_{S2}), are plotted in Fig. 7(b). It can be seen that R'_{S0}, and R''_{S0} are all the same, R_{S1} and R_{S2} are lower than R''_{S1} and R''_{S2} respectively, and R''_{S1} and R''_{S2} are lower than R'_{S1} and R'_{S2} respectively. In addition, compared with the reduction from R'_{S1} to R''_{S1}, R''_{S2} decreases more significantly. This can be explained by the following facts:

(1) When considering the load-sharing characteristics, once a micro-engine fails, the degradation rate of the other surviving micro-engines will increase along with their shared workload. Thus, R''_{S1} and R''_{S2} are lower than R'_{S1} and R'_{S2} respectively. Moreover, when two micro-engines fail, the shared workload and degradation rate of the surviving micro-engines will increase further, and the third micro-engine will fail more easily. Therefore, R''_{S2} decreases more significantly.

(2) When considering the DWSEs, the transmitted shock damage and size are dependent on the workload. The shock damage and size transmitted to the surviving micro-engines will rise as an increasing number of micro-engines fail, and then both soft failure and hard failure will occur more easily. Therefore, R_{S1} and R_{S2} are lower than R''_{S1} and R''_{S2}, respectively.

(3) When no micro-engine has failed, the workload and shock effects on each micro-engine are the same when considering load-sharing characteristics and DWSEs or not, and thus R_{S0}, R'_{S0}, and R''_{S0} are all the same.

Based on Eq. (26), knowing R_{S0}, R_{S1}, and R_{S2}, we can get the reliability $R(t)$ considering load-sharing characteristics and DWSEs. Similarly, we can get the reliability $R'(t)$ without load-sharing characteristics and DWSEs, as well as the reliability $R''(t)$ only considering load-sharing characteristics. A comparison plot of $R(t)$, $R'(t)$, and $R''(t)$ is shown in Fig. 8. It is easy to find that $R(t)$ is lower than $R''(t)$ and $R''(t)$ is lower than $R'(t)$, which indicates that both the load-sharing characteristics and DWSEs decrease the system reliability.

5.2 Sensitive analysis

A sensitive analysis is conducted to study the effects of important parameters on system reliability. The transmission parameter from the system's shock magnitude to the transmitted shock size on C_i, α_i, and the transmission parameter from the system's shock magnitude to the transmitted shock damage on C_i, γ_i, are the parameters of interest. The results of the sensitive analysis of α_i and γ_i are shown in Fig. 9. The red line with rhombus shows the system reliability with α_i increasing from $[\alpha_0=0.9,\alpha_1=1.0,\alpha_2=1.2]$ to $[1.0,1.2,1.4]$. The system reliability decreases more sharply before $t=10^5$, and the system reliability decreases by 15.3% (from to 0.940 to 0.815) at time

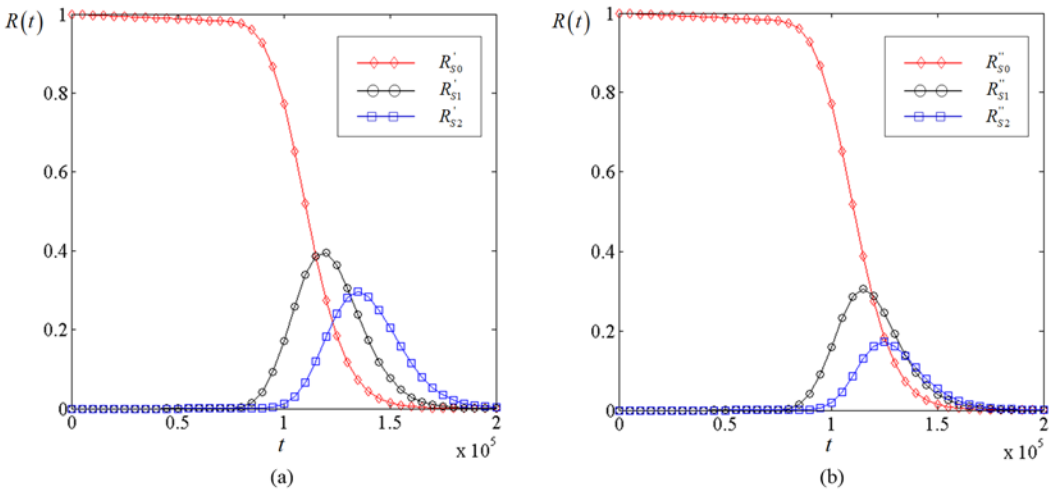


Fig. 7. The values of reliabilities for different system configurations: (a) without the load-sharing characteristic and the DWSEs and (b) only considering the load-sharing characteristic

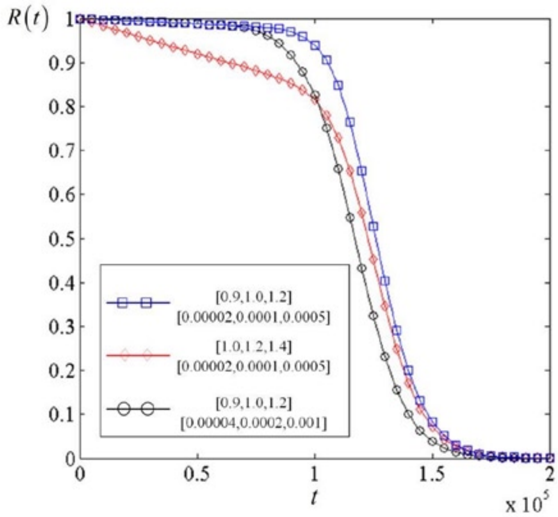


Fig. 8. The comparison plot of the reliabilities with different type of dependence

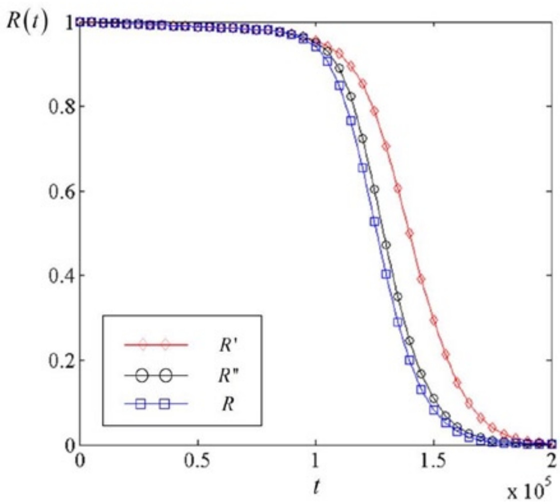


Fig. 9. The sensitivity analysis of reliability on α_i and γ_i

$t = 10^5$.This is mainly due to transmission parameters α_i having a significant affect on hard failures. A high α_i will therefore lead to a large shock size and more hard failures will occur.

The black line with circles shows the system reliability with γ_i increasing from $[\gamma_0 = 0.00002, \gamma_1 = 0.0001, \gamma_2 = 0.0005]$ to $[0.00004, 0.0002, 0.001]$. It can be seen that the two types of reliability curves are almost the same before $t = 7.5 \times 10^4$, and then the reliability with $[\gamma_0 = 0.00004, \gamma_1 = 0.0002, \gamma_2 = 0.001]$ decreases earlier. The system reliability curve moves to the left, which indicates that a larger value of γ_i results in a lower reliability performance.The main reasons for this phenomenon are the following:

(1) A higher γ_i will lead to a larger magnitude of wear debris caused by shocks, while the total wear debris is relatively small and is far from the threshold H at the beginning. Therefore, soft failure rarely occurs and the two types of reliability curves are almost the same.

(2) As t increases, for the system with higher γ_i, the wear debris increases more and approaches the threshold earlier. Thus, the reliability decreases earlier.

6. Conclusion

In this paper, a reliability model is developed for load-sharing k-out-of-n systems subject to the dependent competing soft and hard failures. A new dependence between workload and shock effects is investigated and the dependence is addressed in the model as a major extension from previous reliability models for load-sharing systems. The proposed reliability model is more realistic but difficult to develop due to the load-sharing characteristics and DWSEs. To derive an analytical reliability model, the joint probability density function of shock effects to the soft and hard failures and the conditional probability density function of random component failure times are proposed. A MEMS with four identical micro-engines is then utilized as a realistic application to demonstrate the proposed model. The results show that both the DWSEs and load-sharing characteristics lead to a lower reliability performance. Thus, the reliability evaluation of a load-sharing k-out-of-n system may be more accurate when considering the dependence between workload and shock effects.In future work, maintenance can be studied based on the proposed model and the optimal maintenance policies can be obtained to enhance the system reliability.

Acknowledgement

This paper was supported by the Academic Excellence Foundation of BUAA for PhD students.

Appendix

The notations used in formulating the reliability models are now listed.

DWSEs	Dependent workload and shock effects
MEMS	Micro-Electro-Mechanical System
CDF	Cumulative distribution function
PDF	Probability density function
n	Number of components in the load-sharing system
$N(t)$	Number of failed components by time t
C_i	Surviving components in the load-sharing system with i failed components
Z_j	Magnitude of the j th system shock
Z_{Hj}	Magnitude of the j th shock on the devices for hard failures
Z_{Sj}	Magnitude of the j th shock on the devices for soft failures
W_{ij}	Total transmitted shock size to C_i for the hard failure process from Z_{Hj} and workload
Y_{ij}	Total transmitted shock damage to C_i for the soft failure process from Z_{Sj} and workload
$\tilde{W}_{ij}$	Purely random shock effect for the j th system shock to C_i of the hard failure process
$\tilde{Y}_{ij}$	Purely random shock effect for the j th system shock to C_i of the soft failure process
α_i	Transmission parameter between Z_{Hj} to the shock size for the hard failure process of C_i
γ_i	Transmission parameter from Z_{Sj} to the shock damage for the soft failure process of C_i
β_i	Current degradation rate of C_i
T_i	Failure time of the i th component
m_i	Number of shocks have arrived in the time interval between T_i and T_{i+1}
$X(t)$	Degradation extent at t due to continuous degradation
$S(t)$	Cumulative degradation damage increments
$X_S(t)$	Overall degradation at t due to continuous degradation and shock damages
H	Critical degradation threshold
D	Maximum fracture strength for hard failures
$F_{W_i}(w_i)$	CDF of W_{ij} for C_i
$f_{W_i}(w_i)$	PDF of W_{ij} for C_i
$F_{Y_i}(y_i)$	CDF of Y_{ij} for C_i
$f_{Y_i}(y_i)$	PDF of Y_{ij} for C_i
$f_{Y_i}^{<m>}(y_i)$	PDF of the sum of m i.i.d. Y_{ij} variables
$f_{Z_{Sj}}(z_{Sj})$	PDF of Z_{Sj}

$f_{Z_S}^{<m>}(z_S)$	PDF of the sum of m i.i.d. Z_{Sj} variables
$f_{Z_{Hj}}(z_{Hj})$	PDF of Z_{Hj}

References

1. Babishin V, Hajipour Y, Taghipour S. Optimisation of non-periodic inspection and maintenance for multicomponent systems. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20(2): 327-342, https://doi.org/10.17531/ein.2018.2.20.
2. Bocchetti D, Giorgio M, Guida M, Pulcini G. A competing risk model for the reliability of cylinder liners in marine Diesel engines. Reliability Engineering & System Safety 2009; 94(8):1299-1307, https://doi.org/10.1016/j.ress.2009.01.010.
3. Cha JH, Finkelstein M, Levitin G. Bivariate preventive maintenance of systems with lifetimes dependent on a random shock process. European Journal of Operational Research 2018; 266(1):122-134, https://doi.org/10.1016/j.ejor.2017.09.021.
4. Che H, Zeng S, Guo J. Reliability Analysis of Load-Sharing Systems Subject to Dependent Degradation Processes and Random Shocks. IEEE Access 2017; 5: 23395-23404, https://doi.org/10.1109/ACCESS.2017.2762727.
5. Che H, Zeng S, Guo J, Wang Y. Reliability modeling for dependent competing failure processes with mutually dependent degradation process and shock process. Reliability Engineering & System Safety 2018;180:168-178, https://doi.org/10.1016/j.ress.2018.07.018.
6. Griffith WS. Optimal Reliability Modeling: Principles and Applications. Technometrics 2004;46(1):112-112, https://doi.org/10.1198/tech.2004.s742.
7. Guida M, Pulcini G. A continuous-state Markov model for age- and state-dependent degradation processes. Structural Safety 2011; 33(6): 354-366, https://doi.org/10.1016/j.strusafe.2011.06.002.
8. Guo C, Wang W, Guo B, Peng R. Maintenance optimization for systems with dependent competing risks using a copula function. Eksploatacja i Niezawodnosc -Maintenance and Reliability 2013;15(1):9-17.
9. He G, Geng Z. Dynamics and Robust Control of an Underactuated Torsional Vibratory Gyroscope Actuated by Electrostatic Actuator. IEEE/ASME Transactions on Mechatronics 2015;20(4):1725-1733, https://doi.org/10.1109/TMECH.2014.2350535.
10. Hu Z, Du X. System reliability prediction with shared load and unknown component design details. Artificial Intelligence for Engineering Design Analysis & Manufacturing 2017;31(3):223-234, https://doi.org/10.1017/S0890060417000130.
11. Jiakai C, Yan H, Wei W. Reliability analysis and optimization of equal load-sharing k-out-of-n phased-mission systems. Eksploatacja i Niezawodnosc -Maintenance and Reliability 2015;17(2):250-259, https://doi.org/10.17531/ein.2015.2.12.
12. Kong Y, Ye Z. Interval estimation for k-out-of-n load-sharing systems. IISE Transactions 2017;49(3):344-353, https://doi.org/10.1080/0740817X.2016.1217102.
13. Liu B, Xie M, Kuo W. Reliability modeling and preventive maintenance of load-sharing systems with degrading components. IIE Transactions 2016(8):699-709, https://doi.org/10.1080/0740817X.2015.1125041.
14. Liu B, Xie M, Xu Z, Kuo W. An imperfect maintenance policy for mission-oriented systems subject to degradation and external shocks. Computers & Industrial Engineering 2016;102:21-32, https://doi.org/10.1016/j.cie.2016.10.008.
15. Liu H, Yeh R, Cai B. Reliability modeling for dependent competing failure processes of damage self-healing systems. Computers & Industrial Engineering 2017;105:55-62, https://doi.org/10.1016/j.cie.2016.12.035.
16. Młynarski S, Pilch R, Smolnik M, Szybka J, Wiązania G. A Method for rapid evaluation of k-out-of-n systems reliability. Eksploatacja i Niezawodnosc -Maintenance and Reliability 2018;21(1):170-176, https://doi.org/10.17531/ein.2019.1.20.
17. Nelson W. Accelerated Testing: Statistical Models, Test Plans, and Data Analysis; 2008.
18. Peng H, Feng Q, Coit DW. 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.
19. Rafiee K, Feng Q, Coit DW. Reliability modeling for dependent competing failure processes with changing degradation rate. IIE Transactions 2014;46(5):483-496, https://doi.org/10.1080/0740817X.2013.812270.
20. Rothgang S, Baumhöfer T, Hoek HV, Lange T, Doncker RWD, Sauer DU. Modular battery design for reliable, flexible and multi-technology energy storage systems. Applied Energy 2015;137:931-937, https://doi.org/10.1016/j.apenergy.2014.06.069.
21. Ruiz-Castro JE. Markov counting and reward processes for analysing the performance of a complex system subject to random inspections. Reliability Engineering & System Safety 2016;145:155-168, https://doi.org/10.1016/j.ress.2015.09.004.
22. Song S, Coit DW, 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.
23. Song S, Coit DW, 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, https://doi.org/10.1080/0740817X.2016.1140922.
24. Suprasad AV, Krishna MB, Hoang P. Tampered Failure Rate Load-Sharing Systems: Status and Perspectives; 2008.
25. Taghipour S, Kassaei ML. Periodic Inspection Optimization of a k-out-of-n Load-Sharing System. IEEE Transactions on Reliability 2015;64(3):1116-1127, https://doi.org/10.1109/TR.2015.2421819.
26. Tanner DM, Dugger MT. Wear mechanisms in a reliability methodology. In: Ramesham R. Proceedings of the society of photo-optical instrumentation engineers(SPIE) 2003: 22-40, https://doi.org/10.1117/12.476345.
27. Tanner DM, Miller WM, Peterson KA, et al. Frequency dependence of the lifetime of a surface micromachined microengine driving a load. Microelectronics Reliability 1999;39(3):401-414, https://doi.org/10.1016/S0026-2714(98)00248-0.
28. Tanner DM, Walraven JA, Helgesen K, et al. MEMS reliability in shock environments. IEEE 2000: 129-138.
29. Walraven JA, Headley TJ, Campbell AB, Tanner DM. Failure analysis of worn surface-micromachined microengines. Proc SPIE 1999;3880:30-39, https://doi.org/10.1117/12.359369.
30. Wang Y, Guo L, Wen M, Yang Y. Availability analysis for a multi-component system with different k-out-of-n:G warm standby subsystems subject to suspended animation. Ekspolatacja i Niezawodnosc - Maintenance and Reliability 2019;21(2):289-300, https://doi.org/10.17531/

ein.2019.2.14.
31. Wang Z, Huang H, Li Y, Xiao N. 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.
32. Yang C, Zeng S, Guo J. Reliability Analysis of Load-Sharing K-out-of-N System Considering Component Degradation. Mathematical Problems in Engineering 2015;2015:1-10, https://doi.org/10.1155/2015/726853.
33. Yang L, Zhao Y, Peng R, Ma X. Hybrid preventive maintenance of competing failures under random environment. Reliability Engineering & System Safety 2018;174:130-140, https://doi.org/10.1016/j.ress.2018.02.017.
34. Yang L, Ma X, Zhao Y. A condition-based maintenance model for a three-state system subject to degradation and environmental shocks. Computers & Industrial Engineering 2017;105:210-222, https://doi.org/10.1016/j.cie.2017.01.012.
35. Yang L, Zhao Y, Ma X. Multi-level maintenance strategy of deteriorating systems subject to two-stage inspection. Computers & Industrial Engineering 2018;118:160-169, https://doi.org/10.1016/j.cie.2018.02.038.
36. Ye ZS, Tang LC, Xu HY. A Distribution-Based Systems Reliability Model Under Extreme Shocks and Natural Degradation. IEEE Transactions on Reliability 2011;60(1):246-256, https://doi.org/10.1109/TR.2010.2103710.
37. Yu H, Eberhard P, Zhao Y, Wang H. Sharing behavior of load transmission on gear pair systems actuated by parallel arrangements of multiple pinions. Mechanism & Machine Theory 2013;65(65):58-70, https://doi.org/10.1016/j.mechmachtheory.2013.02.010.
38. Zhang T, Zhang Y, Du X. Reliability analysis for k-out-of-n systems with shared load and dependent components. Structural and Multidisciplinary Optimization 2018;57(3):913-923, https://doi.org/10.1007/s00158-017-1893-z.
39. 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, https://doi.org/10.17531/ein.2018.4.9.
40. Zhao X, Wu C, Wang S, Wang X. Reliability analysis of multi-state k-out-of-n: G system with common bus performance sharing. Computers & Industrial Engineering 2018;124:359-369, https://doi.org/10.1016/j.cie.2018.07.034.
41. Tian Z, Zuo M, Yam RM. Multi-state k-out-of-n systems and their performance evaluation. IIE Transactions 2008;41(1):32-44, https://doi.org/10.1080/07408170802322655.

Haiyang CHE
Shengkui ZENG
Jianbin GUO

School of Reliability and Systems Engineering
Beihang University
Beijing 100191, China

E-mail: chehaiyang@buaa.edu.cn, zengshengkui@buaa.edu.cn, guojianbin@buaa.edu.cn

Article citation info:
NIKOŃCZUK P, ROSOCHACKI W. The concept of reliability measure of recuperator in spray booth. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 265–271, http://dx.doi.org/10.17531/ein.2020.2.9.

Piotr NIKOŃCZUK
Włodzimierz ROSOCHACKI

THE CONCEPT OF RELIABILITY MEASURE OF RECUPERATOR IN SPRAY BOOTH

KONCEPCJA MIARY NIEZAWODNOŚCI REKUPERATORA KABINY LAKIERNICZEJ*

Overspray sediments deposited on the recuperator fins gradually reduce the cross-section of the recuperator channels. The result of this process is the increase in airflow resistance and thermal resistance during heat transfer. Both phenomena have a negative impact on the reliability of the device. This paper presents the concept of recuperator reliability measures. For this purpose, the essential requirement of reliability (indestructibility) was formulated and damage was defined by identifying it with the loss of air flow reserve and reserve of heat transfer efficiency. On this basis ability features of the heat recovery unit were assessed. Limits of features and critical time of recuperator loss of ability were also assessed.

Keywords: *reliability, spray booth, recuperator, overspray sediments.*

Odkładające się na lamelach rekuperatora osady lakiernicze powodują stopniowe zmniejszanie przekroju poprzecznego kanałów rekuperatora. Skutkiem tego procesu są wzrosty oporów przepływu powietrza oraz oporu termicznego przy wymianie ciepła. Oba zjawiska wpływają negatywnie na niezawodność urządzenia. W artykule przedstawiono koncepcję miary niezawodności rekuperatora. W tym celu sformułowano podstawowe wymaganie niezawodnościowe (nieuszkadzalność) oraz zdefiniowano uszkodzenia utożsamiając je z utratą zapasu strumienia powietrza oraz zapasu efektywności wymiany ciepła. Na tym tle określono cechy zdatności urządzenia, granice ich obszarów oraz krytyczny czas utraty zdatności rekuperatora.

Słowa kluczowe: *niezawodność, kabina lakiernicza, rekuperator, osady lakiernicze.*

1. Introduction

The recuperator is a technical device used in ventilation systems, also in spray booths. Due to the technological requirements related to coating technology, it is important that this process is carried out in appropriate conditions, determined primarily by the right temperature and air purity [30]. The purpose of the use of a recuperator is to recover the waste heat from the exhaust air from the working chamber spray booth and preheat the fresh air taken in from the outside. Inside the recuperator there are alternately hot and cold air ducts separated from each other by thin aluminum fins. In the cross-flow recuperator, streams of warm and cold air flow perpendicularly to each other. Heat is exchanged between the air streams via fins. Figure 1 shows a spray booth with a cross recuperator and a diagram of air circulation during the booth operation in the painting mode. Fresh air taken from the outside is pre-heated in the recuperator (1) then, after having been cleaned in the prefilter (2) it is heated to the required temperature by a burner with a heat exchanger (3). The heated air is finally cleaned in the supply filter (4) and blown into the working chamber (5). Coating takes place in the working chamber and during this process the overspray is formed. Volatile organic compounds (VOC) and varnish particles, that are not located on the varnished surface, float in the overspray. The air passing through the working chamber takes the overspray with it and leaves the booth through the paint stop filter (6) which retains the paint particles. Then the purified air through the exhaust duct, goes to the recuperator (1) where it partially transfers the heat to the drawn-in fresh air.

Operation of the recuperator in the spray booth is accompanied by the process of sedimentation of overspray particles on the recuperator fins inside warm air ducts. Research and modeling of heat exchanger pollution are carried out [6]. A model of overspray sediment formation inside a cross recuperator is presented in [14]. A direct consequence of this phenomenon is the reduction of the cross section of the hot air ducts in the heat exchanger. This has been described in more details in [12]. As a result, it leads to an increase of airflow resistance

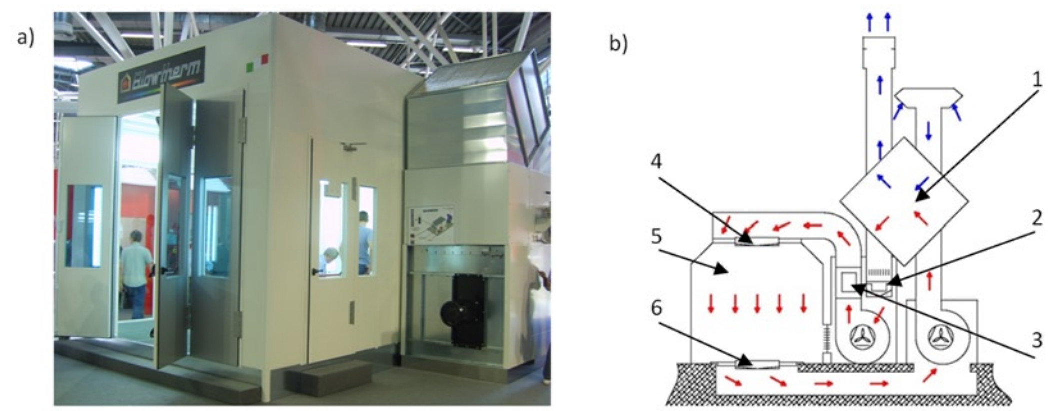


Fig. 1 Spray booth with recuperator a) real object b) Air circulation diagram, 1- cross recuperator, 2 – prefilter, 3 – burner with heat exchanger, 4 –supply filter, 5 – working chamber, 6 – paint stop filter

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

and thermal resistance in the heat exchange process. The latter causes a decrease in the efficiency of heat recovery in the recuperator, while a decrease in the volume of exchanged air leads to the risk of explosion. The process of sedimentation of overspray particles on recuperator fins is destructive, having a significant impact on the reliability of recuperator operation.

Safety requirements for paint and varnish application have been raised in publication [18], defined and updated in relevant European Union regulations [29] and other global legal acts, among others in Australia [32], in the United States [34] and in New Zealand [33]. The requirements of local regulations have been presented in the Guideline for Spray Finishing Particulate Recommended Practice developed by the National Air Filtration Association [30]. A similar study was published in the United Kingdom [31]. The impact of regulations on the car coating industry is presented in the study [35]. Two main hazards have been identified in the spray booths: intoxication of the painter and formation of an explosive mixture. The method of determining risk explosion for spray booths used for powder coating is described in [25].

Reducing the growth rate of paint deposits is possible, among others, by improving the efficiency of cleaning the air removed from the spray. The efficiency of paint stop filters depends on their type [4] and the size of the paint particles carried in the overspray. The analysis of the efficiency of air purification from varnish particles as a function of their size is presented in [1]. Comparative results of filter efficiency are presented in [4], while a broader scope of research work is included in the summary of the completed research project [3]. The size of the particles depends on the kind of varnish and application parameters. The analysis of varnish particle size is presented in publication [20], while significantly expanded results are contained in [21]. A separate analysis of the formation of paint mist and air purification for technology without compressed air (airless spray painting) is presented in [23].

The issue of purifying the air removed from the paint shop is still valid; a state of the art review of the matter is presented in [22]. Works are carried out on new technologies of air filtration in spray booths [7]. Wet gravity cleaning techniques [10] and medialess dynamic filtration [26] methods are considered. So far, no air purification technology that ensures complete removal of overspray particles has been developed. Biofiltration technologies are being considered for the removal of volatile organic compounds [8]; and among others the use of biological stream filtration [24] or fungal biofilters [16] has been proposed.

Deterioration in the level of the device reliability is the consequence of sedimentation of varnish particles on the recuperator fins. As of today, the technical documentation of spray booths equipped with cross recuperators does not contain guidelines for periodic inspections of the recuperator condition and the frequency of its cleaning. Identifying the main features of the recuperator's ability and assessment of their limits will help determine the frequency of inspections and cleaning of the recuperator to ensure the safety of coating process. This paper proposes measures of the recuperator reliability and estimation of the critical ability time of the recuperator t_{kr}, after which it gets damaged, assuming that its important reliability requirement is indestructibility.

The critical time of the recuperator ability t_{kr} is also an indicator for the frequency of inspections and maintenance works.

2. Reliability features

The following discussion assumes that the recuperator reliability is significantly affected by the process of overspray sediment deposition. It is a basic assumption which simplifies reality, neglecting other, less important processes that may lead to other forms of damage (e.g. mechanical damage). The presented influence of overspray sediments on the recuperator working parameters allows the assumption of two features determining its ability. For further analysis heat exchange efficiency reserve and air flow reserve are used. Loss of reserve in relation to each of these features is identified with the occurrence of damage and the transition of the device to a state of unreliability.

Bearing in mind the nature of the phenomenon leading to recuperator damage, in order to determine its measure of reliability, two features of ability were assumed: *heat exchange efficiency reserve* and *pressure drop reserve*.

2.1. Heat exchange efficiency reserve

This feature refers to damage identified with the state of the recuperator, in which the limit value k_{gr} of the thermal conductivity coefficient is reached. It refers to economic aspects related to waste heat recovery.

The efficiency of heat exchange in the recuperator is related to the heat flux $\dot{\mathrm{Q}}(t)$ [1], which is dependent on the coefficient of thermal conductivity $\boldsymbol{k}(t)$ and the temperature difference ΔT between the air streams on both sides of the recuperator lamellas (assuming that this difference is determined and unchangeable in time):

$$\dot{Q}(t) = k(t)\Delta T \tag{1}$$

Heat exchange efficiency reserve $\boldsymbol{h}_1(t)$ is defined as:

$$h_1(t) = \dot{Q}(t) - \dot{Q}_{gr} \tag{2}$$

where $\dot{Q}_{gr}$ means the value of heat exchange efficiency reserve.

Noting equation (1), the relationship describing the heat exchange efficiency reserve $\boldsymbol{h1}\ (t)$ takes the form:

$$h_1(t) = (k(t) - k_{gr})\Delta T \tag{3}$$

The value of the thermal conductivity coefficient $\boldsymbol{k}(t)$ is determined for the recuperator, taking into account the impact of overspray sediments growing on the lamellas. Thermal conductivity $\boldsymbol{k}(t)$ is a random variable, because the growth of sediment layers on the lamella surfaces is a random phenomenon. For any point of time τ and the lamella's surface point described by coordinates (x_o, y_o). The $\{\boldsymbol{k}(x,y,t)\}$ process realization is described by the following relationship [17]:

$$k\left(x_o, y_o, \tau\right) = \frac{1}{\frac{1}{\alpha_1} + \frac{\delta_R}{\lambda_{TR}} + \frac{\delta_S\left(x_o, y_o, \tau\right)}{\lambda_{TS}} + \frac{1}{\alpha_2}} \tag{4}$$

where:

- α_1, α_2 – convective heat transfer coefficient [W/(m^{2}K)]; it is assumed that the value is determined,
- δ_R – thickness of fin [m]; it is assumed that the value is determined,
- $\delta_S(x_o, y_o, \tau)$ – thickness of sediment at the point of the lamella surface determined by coordinates $(x_o,\ y_o)$ [m]; stochastic process $\{\boldsymbol{\delta_S}(x,y,t)\}$ realization in point of time τ,
- λ_{TR} – fin thermal conductivity [W/(mK)]; it is assumed that the value is determined,

1 In this work, the symbols of random variables are written in **bold**.

λ_{TS} – overspray sediment thermal conductivity [W/(mK)]; it is assumed that the value is determined.

To determine the feature of ability $\boldsymbol{h_1(t)}$ in point of time τ, the realization of the coefficient $\boldsymbol{k(t)}$ is determined by the formula:

$$k(\tau)=\frac{1}{xy}\int_0^y\int_0^x k(x,y,\tau)\,dydx \tag{5}$$

2.2. Pressure drop reserve

This feature refers to the damage identified with the state in which the limit value of pressure drop ΔP_{gr} in the recuperator channels is reached. The air flowing through the ventilation duct overcomes the frictional resistance appearing on the walls of the ventilation duct. This resistance, together with the diminishing cross-section of the duct, causes the pressure drop over its entire length, leading to a decrease in the volumetric air flow. Reduction of the volume of exchanged air results in the danger of creating an explosive mixture in the working chamber. It also leads to the risk of poisoning of the painter working inside [32].

Pressure drop reserve $\boldsymbol{h_2}(t)$ is determined by formula:

$$\mathrm{h}_2(t)=\Delta P_{gr}-\Delta \mathrm{P}(t) \tag{6}$$

where:

$\Delta \boldsymbol{P}(t)$ – pressure drop.

The pressure drop in the ventilation duct is a stochastic process $\{\Delta \boldsymbol{P}(t)\}$, whose realizations depend on the length of the channel and randomly time-varying unit resistance $\boldsymbol{r}(t)$:

$$\Delta P(t)=r(t)l \tag{7}$$

The resistance coefficient $\boldsymbol{r}(t)$ also known as unit pressure drop [17] depends on many parameters, including two that change their values over time:

$$\mathrm{r}(t)=\frac{\lambda_{\mathrm{F}}(t)\varsigma w^2}{2\mathrm{d}(t)} \tag{8}$$

where:

$\boldsymbol{\lambda_{\mathrm{F}}}(t)$ – dimensionless friction resistance coefficient; random value,

ς – air density [kg/m3]; it is assumed that the value is determined,

w – average air flow rate [m/s]; it is assumed that the value is determined

$\boldsymbol{d}(t)$ – hydraulic diameter of channel [m]; random value.

The hydraulic diameter for the channel cross-section in general form is determined using the equation [17]:

$$d=\frac{2ab}{a+b} \tag{9}$$

where a and b denote the dimensions of the rectangular cross-section of the channel.

Taking into account the time-varying randomly thickness of the overspray sediments $\boldsymbol{\delta_S(t)}$, the formula describing the realization of the equivalent diameter $\boldsymbol{d(t)}$ at the point of time τ takes the form:

$$d(\tau)=\frac{2ab-4(a+b)\delta_S(\tau)+8\delta_S{}^2(\tau)}{a+b-4\delta_S(\tau)} \tag{10}$$

The dimensionless friction resistance coefficient $\boldsymbol{\lambda_F(t)}$ changes its value over time due to the dependence on Reynolds number. For turbulent flow, the friction coefficient is described as follows:

$$\lambda_{\mathrm{F}}(t)=\frac{0{,}3164}{\sqrt[4]{\mathrm{Re}(t)}} \tag{11}$$

The Reynolds number $\boldsymbol{Re(t)}$ depends on the hydraulic diameter equivalent to the cross section in the ventilation duct $\boldsymbol{d(t)}$. At the point of time τ it is determined basing on the formula:

$$\mathrm{Re}(\tau)=\frac{wd(\tau)}{\upsilon} \tag{12}$$

where υ is the kinematic coefficient of viscosity [m^2/s].

3. Limits of ability features

The above defined features create a basis for demarcation of the following areas of ability

– for *heat exchange efficiency reserve*:

when the recuperator is able (no damage)

$$\mathrm{h}_1\left(\dot{Q}(t),\dot{Q}_{gr}\right)>0 \tag{13}$$

when the recuperator is disable

$$\mathrm{h}_1\left(\dot{Q}(t),\dot{Q}_{gr}\right)\le 0 \tag{14}$$

– for *pressure drop reserve*:

when the recuperator is able (no damage)

$$\mathrm{h}_2\left(\Delta \mathrm{P}(t),\Delta P_{gr}\right)>0 \tag{15}$$

when the recuperator is disable

$$\mathrm{h}_2\left(\Delta \mathrm{P}(t),\Delta P_{gr}\right)\le 0 \tag{16}$$

4. Measure of reliability

It is assumed that the basic reliability requirement of the recuperator construction is its functioning without damage within a specified period of time. This approach is justified by the function that this device performs. Achieving the critical states described above identified in this work with damage, is tantamount to unacceptable deterioration of the functionality of the device. This significantly affects the safety and operational cost of the spray booth and the quality of the paint application process.

Given the above, it is assumed that the measure of reliability that characterizes the reliability requirement formulated above, is the probability of its fulfillment in the analyzed time period:

$$R(t)=P\left(\left(\mathrm{h}_1(t)>0\right)\cap\left(\mathrm{h}_2(t)>0\right)\right) \tag{17}$$

Therefore, the probability of recuperator operation with both analyzed ability features is determined. It should be noted that the relationship (17) determines the probability of two dependent events. This relation results from the dependence of both features on the process of particle sedimentation, represented here by the stochastic process $\{\delta_S(t)\}$.

5. An analysis of ability features

It is assumed that the presented ability features depend on the constant construction parameters of the recuperator and the time varying thickness of overspray sediments $\delta_S(t)$.

A preliminary analysis of ability features was carried out on the example of a recuperator dedicated to spray booths. The unit is a part of the offer documentation supplied by one of the entrepreneurs operating on the refinishing market [28]. This recuperator itself also was an object of study presented in [13].

For the presented recovery unit, the impact of sediments thickness on changes in formulated ability features was analyzed. Figure 2 shows the changes in thermal conductivity coefficient $k(\delta_s)$ according to equation (4) and the corresponding changes in heat flux $\dot{Q}(\delta_S)$ according to formula (1). They affect the value of the ability feature $h_1(t)$. The value of thermal conductivity coefficient $\lambda_{TS} = 0.082 \pm 0.003$ [W/(mK)] was taken for calculations. Measurement methodology and value results with error analysis for thermal conductivity of sediments are described in [15]. For calculations according to equation (4) the following values were used: thermal conductivity of aluminum $\lambda_{TR} = 200$ [W/(mK)], equal convective heat transfer coefficient for air on both sides of the fin $\alpha_1 = \alpha_2 = 50$ [W/(m^{2}K)], fallowing the documentation [28] the thickness of the recuperator fins was determined as $\delta_R = 2e{-}4$ [m]. Values of heat flux $\dot{Q}(\delta_S)$ were calculated for temperature difference $\Delta T = 40$ [K].

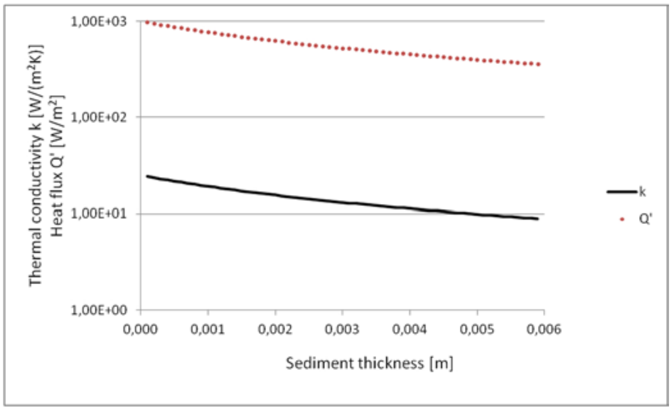


Fig. 2. Calculated changes in thermal conductivity coefficient $k(\delta_s)$ and heat flux $\dot{Q}(\delta_S)$

Figure 3 shows the increase in pressure drop as a function of sediment thickness $\Delta P(\delta_s)$. The pressure drop is associated with the ability feature $h_2(t)$. Calculations were made according to equation (7). The following variable values were used: air density $\varsigma = 1.2$ [kg/m^3], average air flow rate $w = 5.56$ [m/s], kinematic coefficient of viscosity $\upsilon = 1.5e{-}5$ [m^2/s]. Fallowing the documentation [28] the following parameters of recuperator channels were determined: dimensions of the rectangular cross-section of the channel $a = 1.2e{-}2$ [m], $b = 1$ [m] channel length $l = 1$ [m], number of channels in the recuperator separately for hot and cold air $n = 60$. Overspray sediments grow up only in the ducts with warm air removed from the spray booth. Calculations of the pressure drop $\Delta P(\delta_s)$ were carried out for a single warm air channel, assuming that there is a uniform distribution of air velocity in all cross-sections of the channels and turbulent flow occurs. For the hydraulic diameter $d(t)$ described by equation (10), a homogeneous, average value of sediments thickness $\delta_S(t)$ was assumed.

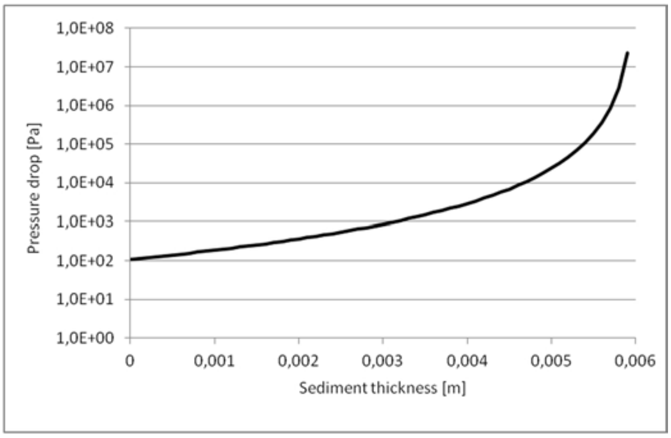


Fig. 3. Calculated pressure drop $\Delta P(\delta_s)$

Figure 4 shows the percentage changes of heat transfer efficiency and the inverse of pressure drop as a function of sediment thickness. The graph shows the inverse of the pressure drop of $1/\Delta P(\delta_s)$ to improve transparency in comparison with the percentage changes of the heat flux $\dot{Q}(\delta_S)$. The starting points of 100% for both parameters indicate their values for the clean condition of fins, not covered with sediments. The analysis of the presented graph indicates a much greater impact of the sediment growth process on the change of pressure drop, and as a result on the change in the value of the feature $h_2(t)$.

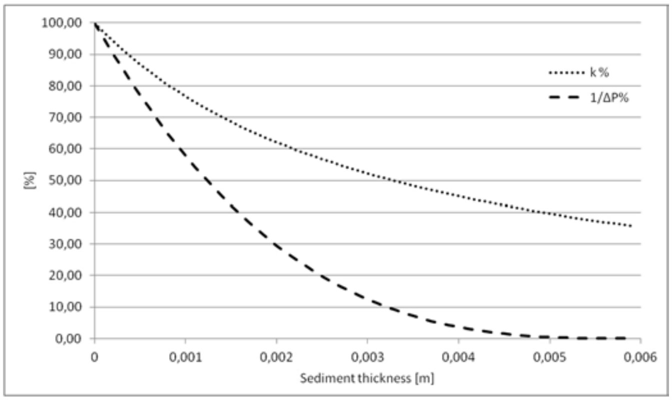


Fig. 4. Percentage changes in heat flux $\dot{Q}(\delta_S)$ and inverse pressure drop $1/\Delta P(\delta_s)$ depending on the thickness of the sediments

The growth rate of overspray sediments depends on many parameters and it is a process with variable dynamics. Figure 5 presents the results of measurements of sediments in three paint booths. The points presented in the graph represent average values from measurements after a given period of booth operation time. The trend lines show the average growth rates of sediments in each of the spray booth. The research methodology and their conditions were described in [13]. Measurements were carried out in spray booths not equipped with recuperators. Measurement points for technical reasons were located in each cabin on the air damper cover in the exhaust channel. This is the place where it is customary to install a recuperator (Figure 1). The measurement results were a basis for development of a simulation model of sediment deposition on the recuperator fins. The numerical model and simulation results are presented in [14]. The model assumes that the air velocity is the same in all cross-sections of the recuperator channels and that the flow is turbulent.

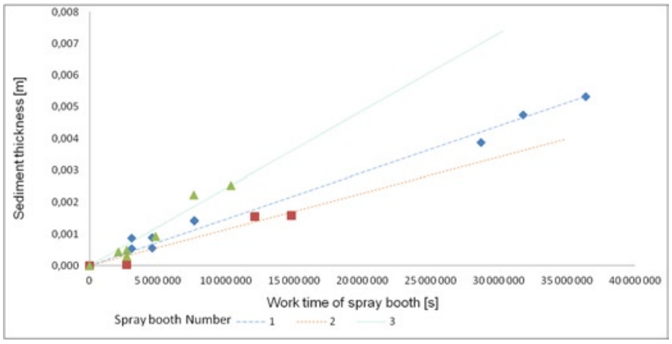


Fig. 5. Growth rate of paint deposits in three paint booths [13]

An analysis of the results presented in Figure 5 indicates a strongly random nature of the sediment build-up process. For example, trend lines of sediment thickness measurement results in booths No. 2 and 3 indicate more than twice the growth rate in booth No. 3 than in No. 2. This is primarily due to the random impact on this process of factors such as: total of summary working times in the painting and drying modes in the total operation time of the spray booth, setting parameters and transfer efficiency of the spray gun, efficiency of the paint stop filter, skills of the painter, shape and sizes of painted objects.

Based on the trend lines presented in Figure 5, the percentage changes in pressure drop and thermal conductivity coefficients in the time domain were developed for individual booths. The results are shown in Figure 6. On diagrams for booths 1, 2 and 3 are respectively marked pressure drops as *ΔP1*, *ΔP2* and *ΔP3* and the heat flux as *Q'1*, *Q'2* and *Q'3*. The calculations were carried out using the equations presented in section 2.

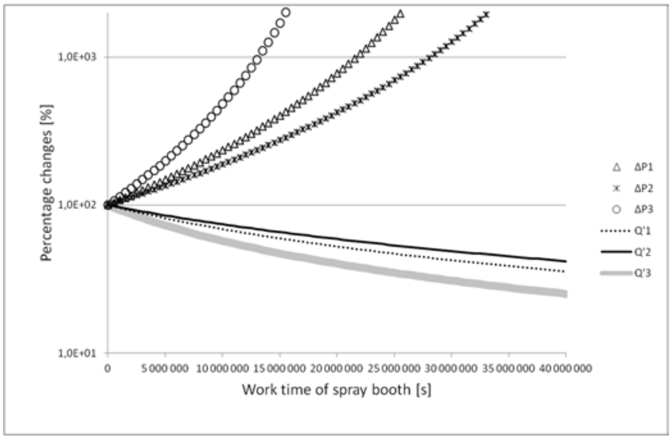


Fig. 6. Percentage changes in pressure drop (ΔP1, ΔP2, ΔP3) and heat flux (Q'1, Q'2, Q'3) in individual spray booths

Due to the exponential increase in pressure drop shown in Figure 3, the diagram in Figure 6 shows the fragments of curves that do not exceed the 2000 [%] value of changes. These values are the results of theoretical calculations that will not be achievable in a normal operation of the spray booth with a recuperator.

Comparing the dynamics of the percentage changes in the thermal conductivity coefficient and the pressure drop, a significant increase in the percentage change in the pressure drop relative to the percentage change in the thermal conductivity coefficient is noticeable.

The study of the paint booth operation process described in the work [13], observations and interviews with paint booth users provide the basis for estimating the values that are proposed in the ability features analyzed here as limits.

In relation to the ability feature $\boldsymbol{h_1(t)}$ identified with a feature of heat exchange efficiency reserve it is proposed to initially take the following as a limit value

$$\dot{Q}_{gr} = 0.5\dot{Q}_N = 500[W/m^2] \tag{18}$$

wherein Q is the nominal value of the heat flux for the new non-sedimented recuperator. The nominal value of the heat flux can be read from Figure 2 $\dot{Q}_N$ = 1000 [W/m^2]. At this value of the heat flux, the energy efficiency of the recuperator reaches half its nominal value, which reduces by half the estimated economic benefits of the spray booth user. It was considered that half of the savings obtained due to recovered heat constituted the profitability limit of investment costs related to the purchase and installation of a recuperator.

Regarding the $\boldsymbol{h_2(t)}$ characteristic, twice the nominal pressure drop is proposed as the limit value for the pressure ΔP_N

$$\Delta P_{gr} = 2\Delta P_N = 216[Pa] \tag{19}$$

As a nominal value of $\Delta P_{N,}$ a pressure drop on the recuperator in a clean state, when the recuperator fins are not covered with overspray sediments, was assumed. Figure 3 shows the changes in pressure drop calculated according to equation (7) depending on the thickness of the sediments. The initial value of the pressure drop for the sediments thickness δ_s = 0 [mm] is equal ΔP_N = 108 [Pa]. The total pressure drop in the ventilation ducts of the spray booth is individual for each booth. It is associated with many parameters and, above all, the length and cross-sections of the ducts, the number and type of fittings in ventilation systems, the construction of the heat exchanger for air heating, the types and cleanliness of air filters as well as the recuperator. On this basis, it was accepted that twice the nominal pressure drop on the recuperator is its critical value.

For the critical values proposed above and on the basis of faster changes in pressure drop as a function of sediment thickness, the $\boldsymbol{h_2(t)}$ feature was indicated as the leading feature in estimating the time of the loss of ability.

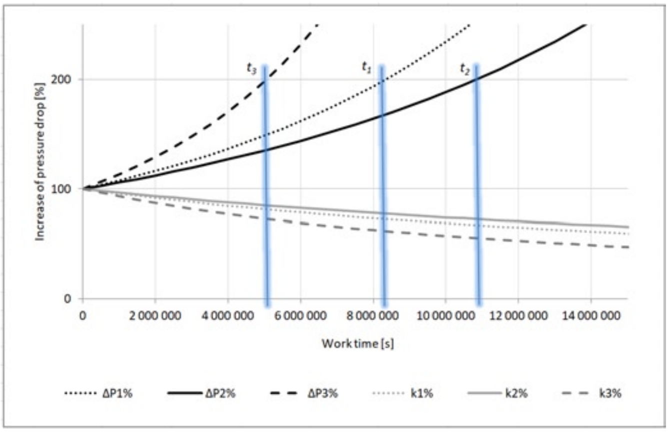


Fig. 7. Estimated doubling times of pressure drop

Figure 7 shows the estimated times at which the ΔP_{gr} limits were reached for individual booths. The times when the pressure drop doubled for each booth were marked successively as t_1, t_2 and t_3 adequately to the booth number. The doubling times for individual booths are respectively:

$$t_1 = 8346234 \text{ [s]}$$

$$t_2 = 10837105 \text{ [s]}$$

$$t_3 = 5061598 \text{ [s]}$$

These values are varied and it is particularly noticeable that the time t_2 is almost twice as large as the time t_3. The diversity of values is due to the strong randomness of the sediment grow up process.

Based on the above calculations, the t_3 value was adopted as the critical time of the recuperator's loss of ability, this is the shortest period in which damage will not occur

$$t_{kr} = t_3 = 5061598[s]$$

6. Summary

In the proposed model of the recuperator reliability in the spray booth, two ability features have been indicated: ***h₁(t)*** *heat transfer efficiency reserve* and ***h₂(t)*** *pressure drop reserve*. The features ***h₁(t)*** and ***h₂(t)*** have different variations depending on the thickness of the sediments. As indicated by the analysis the overspray sedimentation has a random character. The ***h₁(t)*** feature is associated with a decrease in heat recovery efficiency in the recuperator. It is economic in nature. However, the ***h₂(t)*** feature is associated with an increase in airflow resistance through the recuperator. Reducing the volume of air exchange in the spray booth can lead to increased concentration of overspray and VOC. This can result in poisoning of the painter or formation of an explosive mixture. This study indicates a much faster pace of changes in the ***h₂(t)*** feature as compared to the ***h₁(t)*** in the longer period of operation. Finally, the ***h₂(t)*** feature was recognized as dominant, which has a significant impact on determining the periodicity of inspection and cleaning of the recuperator.

The working time of the recuperator in undamaged condition was estimated at t_{kr} = 5061598 seconds, which is equivalent to 1406 hours of spray booth operation. After this time, it is required to carry out an inspection and clean the recuperator of overspray sediments. The results obtained relate to three spray booths and constitute preliminary values. Due to the heterogeneous growth rate of overspray sediments, it is difficult to estimate the exact times to reach the limit values by the ability characteristics. Acquiring and organizing the results of measurements of the growth rate of sediments in many spray booths will facilitate the determination of average and critical intervals of the paint booth operation time in which recuperator damage should be expected.

The proposed ability features illustrate the assumption that the reliability of the recuperator is significantly affected by the thickness of the overspray sediments without any other possible damage. A similar approach was presented in [19]. The condition of recuperator damage results in the loss of ability of the entire spray booth. The indicated features of the recuperator ability become also the features of the spray booths, but they are not the only features indicating the level booth's ability. By averaging the growth rate of sediments, an individual model of a spray booth as a multi-element system described in [27] can be created. The work [11] also presents a method of rapid assessment of the reliability of a complex technical system, where the components have different renewal times. Modern technologies of industry 4.0 relying on connecting industrial devices with the internet and data storage in the network bring about the possibility of automatic acquisition and storage in the cloud of the results of measurements of spray booths operating parameters. Selected parameters may indirectly indicate the cleanliness of the recuperator. An analysis of the collected results in the cloud enables remote determination of the recuperator's ability [5]. An analysis of the ability can be carried out by machine learning method [2] or fuzzy set logic [9].

References

1. Chan TL, D'Arcy JB, Schreck RM. High-solids paint overspray aerosols in a spray painting booth: particle size analysis and scrubber efficiency. American Industrial Hygiene Association Journal 1986 ;7: 411-417, https://doi.org/10.1080/15298668691389964.
2. Chen X, Zhang Z, Zhang Z. Real-time equipment condition assessment for a class-imbalanced dataset based on heterogeneous ensemble learning. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21(1): 68-80, https://doi.org/10.17531/ein.2019.1.9.
3. Darvin CH. Fractional penetration of paint overspray arrestors. Project summary, United States Environmental Protection Agency, April 1997, https://permanent.access. gpo.gov/ websites/epagov/www.epa.gov/ORD/WebPubs/projsum/600sr97011.pdf.
4. Darvin CH, Cox LS. & Smith D. Comparative Study of Spray Booth Filter System Efficiency, Journal of the Air & Waste Management Association, 1999; 49(3): 339:344, https://doi.org/10.1080/10473289.1999.10463801.
5. Fang Y, Tao W, Tee KF. A new computational method for structural reliability with big data. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21(1): 159-163, https://doi.org/10.17531/ein.2019.1.18.
6. Hassan Al-Haj Ibrahim Fouling in Heat Exchangers, MATLAB - A Fundamental Tool for Scientific Computing and Engineering Applications -2012; 3, Vasilios N. Katsikis, IntechOpen, DOI: 10.5772/46462. Available from: https://www.intechopen.com/books/matlab-a-fundamental-tool-for-scientific-computing-and-engineering-applications-volume-3/fouling-in-heat-exchangers, https://doi.org/10.5772/46462.
7. Howarth JL, Anand SC. Design, development and characterization of a novel and innovative exhaust filter media for the global automotive industry. Textile Research Journa, https://doi.org/10.1177/0040517515619349.
8. Kazenski SL, Kinney KA. Biofiltration of paint spray booth emissions: Packing media considerations and VOC interactions, in: Proceedings of the 2000 Annual Meeting and Exhibition of Air and Waste Management Association, June 2000, Salt Lake City, A & WMA, 1-14.
9. Khaniyev T, Baskir MB, Gokpinar F, Mirzayev F. Statistical distributions and reliability functions with type-2 fuzzy parameters. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21(2): 268-274, https://doi.org/10.17531/ein.2019.2.11.
10. Kim HT, Jung CH, Oh SN, Lee KW. Particle removal efficiency of gravitational wet scrubber considering diffusion, interception, and impaction, Journal of Environmental Engineering Science 2001; 2: 125-136, https://doi.org/10.1089/10928750151132357.
11. Młynarski S, Pilch R, Smolnik M, Szybka J, Wiązania G. A Method for rapid evaluation of k-out-of-n systems reliability. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (1): 170-176, https://doi.org/10.17531/ein.2019.1.20.
12. Nikończuk P. Study of Heat Recovery in Spray Booths, Metal Finishing 2013; 111(6): 37-39, https://doi.org/10.1016/S0026-0576(13)70284-8.
13. Nikończuk P. Preliminary analysis of heat recovery efficiency decrease in paint spray booths. Transactions of the Institute of Metal Finishing 2014; 92(5): 235-237, https://doi.org/10.1179/0020296714Z.000000000200.
14. Nikończuk P. Preliminary modeling of overspray particles sedimentation at heat recovery unit in spray booth. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (3): 387-393, https://doi.org/10.17531/ein.2018.3.6.
15. Nikończuk P, Dobrzyńska R. Preliminary Measurements of Overspray Sediment's Thermal Conductivity. Ochrona przed Korozją 2018; 61 (2): 40-42, https://doi.org/10.15199/40.2018.2.3.

16. Qi B, Moe WM, Kinney KA. Treatment of paint spray booth off-gases in a fungal biofilter. Journal of Environmental Engineering 2005; 2: 180-189, https://doi.org/10.1061/(ASCE)0733-9372(2005)131:2(180).
17. Recnagel - Sprenger - Hönmann - Schramek, Ogrzewanie i klimatyzacja, Poradnik, Wydanie 1, EWFE, Gdańsk 1994.
18. Ron J. Ventilation and paint filtration requirements for spray booths. Metal Finishing 2007; 105 (6), 82-84, https://doi.org/10.1016/S0026-0576(07)80185-1.
19. Rosochacki W. Basis of reliability model for offshore structures bearing. Scientific Journals Maritime University of Szczecin, 2009; 19 (91): 92-96.
20. Sabty-Daily RA, Hinds WC., Froines J. Size distribution of chromate paint aerosol generated in a bench-scale spray booth. The Annals of Occupational Hygiene 2005; (1): 33-45.
21. Salaimeh AA, Adornato A, Akafuah NK, Saito K. Assessment of Spray Envelope Using Droplet Breakup Regime and Mechanisms. ICLASS 2015, 13th Triennial International Conference on Liquid Atomization and Spray Systems, Tainan, Taiwan, August 23-2 2015.
22. Schweizer R. Technological Advancements in Overspray Collectors. Metal Finishing, 2010; 108(4): 38-45, https://doi.org/10.1016/S0026-0576(01)85352-6.
23. Settles GS, Miller JD, Hartranft TJ, Brandt AD. Visualization and collection of overspray from airless spray painting. In: 8th International Symposium on Flow Visualization, Sorrento, Italy 1998.
24. Webster TS, Tongna AP, Guarini WJ, McKnight L. Treatment of volatile organic compound emissions from a paint spray booth application using biological trickling filtration. In: 2000 USC-TRG Conference on Biofiltration (an air pollution control technology), Los Angeles, California, USA, October 2000, University of Southern California, Reynolds Group, 41-50.
25. Wiechuła BM. The assessment of explosion risk in cabins during spraying coating powder. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2005; 7 (25): 3-15.
26. Wright SR, Bond JH. Hazardous waste Minimization of paint overspray via medialess dynamic particle filtration. Final Report 25 Apr 97 -25 Jan 98, INNOVATECH, INC.
27. Zhang Y, Ma Y, Ouya ng 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, https://doi.org/10.17531/ein.2018.4.9.
28. Archiwum firmy Techniki systemów Aplikacyjnych Sp. z o.o.
29. Coating plants - Spray booths for application of organic liquid coating materials - Safety requirements, European Standard, Final Draft, EN 12215, June 2004.
30. Guideline for Spray Finishing Particulate Recommended Practice, Virginia Beach VA, National Air Filtration Association, 2012, http://www.nafahq.org/wp-content/uploads/Spray FinishingParticulateFiltrationa.pdf, dostęp: 11.05.2015.
31. Guidance on the application of the dangerous substances and explosive atmospheres regulations (DSEAR) 2002 to motor vehicle repair (MVR), Health and Safety Executive, Information Document HSE, 2002, 803/71, http://www.hse.gov.uk/foi/internalops/ocs/800-899/803_71/803_71id.htm#app2, dostęp: 04.12.2015.
32. Occupational safety and health act 1984 and occupational safety and health regulations 1996, Code of practice spray painting, West Perth, Australia WorkSafe Western Australia Commission, June 2009, https://www.commerce.wa.gov.au/sites/default/files/atoms/files/code_spray_painting.pdf, dostęp: 24.03.2016.
33. Spray coating a guide to the spray coating regulations 1962, Wellington, New Zealand, Occupational Safety and Health Service, Department of Labour, 1962, July 2013, http:// legislation. govt.nz/regulation/public/1962/0054/latest/whole.html#DLM16295.
34. United States Environmental Protection Agency (USEPA), National emission standards for hazardous air pollutants: Surface coating of miscellaneous metal parts and products; Final rule, Federal Register 69 (1), January 2, 2004, 130-192, https://www.gpo.gov/fdsys/pkg/FR-2004-01-02/pdf/03-21917.pdf.
35. United States Environmental Protection Agency (USEPA), Regulatory impact analysis for the proposed automobile and light duty truck coating NESHAP: Final Report, EPA-452/R-01-013, North Carolina, USEPA, Research Triangle Park, N.C., 2002.

Piotr NIKOŃCZUK
Włodzimierz ROSOCHACKI
West Pomeranian University of Technology Szczecin
Faculty of Maritime Technology and Transport
Al. Piastów 41, 71-065 Szczecin, Poland

E-mails: piotr.nikonczuk@zut.edu.pl,
wlodzimierz.rosochacki@zut.edu.pl

Article citation info:
LIU C, KRAMER A, NEUMANN S. Reliability assessment of repairable phased-mission system by Monte Carlo simulation based on modular sequence-enforcing fault tree model. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 272–281 http://dx.doi.org/10.17531/ein.2020.2.10.

Chenxi LIU
Achim KRAMER
Stephan NEUMANN

RELIABILITY ASSESSMENT OF REPAIRABLE PHASED-MISSION SYSTEM BY MONTE CARLO SIMULATION BASED ON MODULAR SEQUENCE-ENFORCING FAULT TREE MODEL

OCENA NIEZAWODNOŚCI NAPRAWIALNEGO SYSTEMU Z MISJAMI OKRESOWYMI ZA POMOCĄ SYMULACJI MONTE CARLO W OPARCIU O MODUŁOWY MODEL DRZEWA NIEZDATNOŚCI Z BRAMKAMI SEQ

Phased-mission system (PMS) is the system subject to multiple, consecutive and non-overlapping tasks. Much more complicated problems will be confronted when the PMS is repairable since the repairable system could perform the multi-phases mission with more diversity requirements. Besides, various maintenance strategies will directly influence the reliability analysis procedure. Most researches investigate those repairable PMSs that carry out the multi-phases mission with deterministic phase durations, and the mission fails once the system switches from up to down. In this case, one common maintenance strategy is that failed components are repairable as long as the system keeps in up state. However, many practical systems (e.g., construction machinery, agricultural machinery) may be involved in such multi-phases mission, which has uncertain phase durations but limited by a maximum mission time, within which failed components can be unconditional repaired, and the system can be restored from down state. Comparing with the former type of repairable PMS, the latter will also concern phase durations dependence, and both the system and components included have the state bidirectional transition. This paper makes new contributions to the reliability assessment of repairable PMSs by proposing a novel SEFT-MC method. Two types of repairable PMS mentioned above are considered. In our method, a specific sequence-enforcing fault tree (SEFT) is proposed to correctly depict failure logical relationships between the system and components included. In order to transfer the graphical fault tree (no matter its size and complexity) into a modular reliability model used in Monte Carlo (MC) simulation, an improved linear algebra representation (I-LAR) approach is introduced. Finally, a numerical example including two cases corresponding to the two types of repairable PMS is presented to validate the proposed method.

Keywords: *repairable, phased-mission system, modular reliability modeling, improved linear algebra representation; Monte Carlo simulation.*

System z misjami okresowymi (phased-mission system, PMS) to system, który wykonuje wiele następujących po sobie i nienakładających się na siebie zadań. W przypadku naprawialnych systemów PMS, analiza niezawodności jest o wiele bardziej skomplikowana, ponieważ system naprawialny może wykonywać misje wielofazowe o bardziej różnorodnych wymaganiach. Poza tym systemy takie wymagają zastosowania różnych strategii utrzymania ruchu, co ma bezpośredni wpływ na procedurę analizy niezawodności. Większość badaczy bada naprawialne systemy PMS, które wykonują misje wielofazowe, w których czas trwania fazy jest wielkością deterministyczną, a misja kończy się niepowodzeniem, gdy system przechodzi ze stanu zdatności do stanu niezdatności W takich przypadkach najczęściej przyjmuje się, że uszkodzone elementy można naprawić o ile system pozostaje w stanie zdatności. Jednak wiele systemów stosowanych w praktyce (t.j. maszyny budowlane czy maszyny rolnicze) może wykonywać misje wielofazowe, w których czas trwania fazy jest wielkością niepewną, ograniczoną jedynie przez maksymalny czas trwania misji, w którym to czasie uszkodzone komponenty mogą być bezwarunkowo naprawiane, dzięki czemu system może zostać przywrócony do stanu zdatności. W porównaniu z pierwszym rodzajem naprawialnego PMS, w drugim, czasy trwania faz są zależne od siebie. Ponadto, w systemie tego typu, zarówno poszczególne elementy, jak i cały system mogą przechodzić ze stanu zdatności do stanu niezdatności i odwrotnie. Niniejsza praca wnosi nowy wkład w ocenę niezawodności naprawialnych systemów PMS, proponując nowatorską metodę, która polega na wykorzystaniu dynamicznego drzewa niezdatności do przeprowadzenia symulacji Monte Carlo (SEFT-MC). Rozważane są dwa wymienione powyżej typy naprawialnego PMS. W naszej metodzie zaproponowano drzewo niezdatności z bramkami SEQ (SEFT), które pozwala poprawnie zobrazować logiczne zależności między systemem a jego komponentami w zakresie uszkodzeń. Do przeniesienia graficznego drzewa niezdatności (bez względu na jego rozmiar i złożoność) do modułowego modelu niezawodności wykorzystywanego w symulacji Monte Carlo, zastosowano udoskonaloną metodę reprezentacji algebry liniowej (I-LAR). Poprawność proponowanej metody wykazano na przykładzie numerycznym obejmującym dwa przypadki odpowiadające dwóm omawianym typom naprawialnego PMS.

Słowa kluczowe: *naprawialny, system z misjami okresowymi, modułowe modelowanie niezawodności, udoskonalona reprezentacja algebry liniowej; symulacja Monte Carlo.*

1. Introduction

Phased-mission systems (PMSs) are systems that perform multiple, consecutive and non-overlapping tasks [13]. Such systems are common in many fields, like power [4], spacecraft [5-6], distributed computing system [13], and military [26]. As the name suggests, the whole mission undertaken by PMS includes multiple tasks; each specified task lasts for a duration and the system has to withstand different stress loads. Usually, the system structure, as well as component failure behaviors are various among different phases; some components participate in more than one phase, and the cumulative damage caused in phase i have to be taken into account when determining the failure rate in phase j ($i<j$). Thus, challenges in analyzing PMS comprise of two aspects: dynamic behaviors among phases, and state dependence among phases.

For non-repairable PMS, methods and applications for reliability assessment have been extensively studied [19]. Basically, existing methodologies can be categorized into the simulation and the analytical methods. The simulation methods are outstanding in their wide applicability to a variety of scenarios [23, 28]; whereas the analytical methods, including binary decision diagram (BDD)-based method [17, 24, 25], multivalued decision diagram (MDD)-based method [13, 16], Markov chains-based method [18], Markov reward model-universal generating function (UGF) technique [7], Bayesian networks approach [4], recursive algorithm [3], have advantages in obtaining accurate results with high efficiency, but may not be suitable in large-scale PMS with complex dynamic behaviors.

In contrast, the investigation on the reliability of repairable PMSs has not been studied to the same extent, though they are commonly found in many real-world engineering applications. Comparing with non-repairable PMSs, there will be more challenges have to be confronted. On the one hand, the repairable system could perform the multi-phases mission with more diversity requirements; on the other hand, various maintenance strategies will directly influence the reliability analysis procedure.

Existing researches mostly investigate those repairable PMSs that carry out multi-phases missions with such requirement, i.e., phase durations are deterministic. Kim [2] supposed that failed components are repairable only when the system is up, and a Markov model is formulated to obtain the mission reliability. A series-parallel PMS is studied by the generic Monte Carlo simulator known as Raptor [15], in which only the non-critical component (i.e., generally a redundant component) can be repaired. Lu [10] proposed a decomposition approach combined with continuous-time Markov chains (CTMCs) to evaluate the reliability of PMS considering both combinatorial phase requirements and repairable components. The PMS consisting of a large number of phases and repairable components is studied in [9,11]. It is assumed that the failed component can only be repaired when the system is still operating, and it can be reused only in the next phase after its restoration. A truncation method based on the binary-decision-diagram (BDD) and Markov chains is proposed to solve the scaling issue. Considering multi-mission PMS with repairable components, and repairable PMS with common cause failures, Wu [20-21] proposed an extended object-oriented Petri net (EOOPN) model for mission reliability simulation. In Li et. al.'s research [5], redundant architecture such as cold standby (structural or functional) is applied to certain critical parts, and then, the Semi-Markov process is used to assess the reliability of the PMSs with non-exponential and partially repairable components. Zhao et.al. [27] introduced spare parts for every component to make the PMS repairable; an integrated modeling method based on the multistate multi-valued decision diagram (MMDD) and Markov chain is developed to evaluate the mission success; besides, the optimal allocation of spare parts is also studied. Overall, the PMS with deterministic phase durations refers to that each task has to be continuously executed for a specific duration. Some components in the system are allowed to be repaired or replaced to keep the system on, until a minimum cut set is triggered, resulting in the task (mission) interruption, i.e., mission failure. On the contrary, mission success is concluded if the system completes the whole mission in continuous operation.

However, it is not necessary to require the system to perform a multi-phases mission without interruption in many fields, such as construction machinery, agricultural machinery, printing equipment, machine tools, etc., since downtime of the system is allowable ascribed to the components' maintenance. A typical example is the tractor system that performs the grass harvest mission. The mission includes 3 phases: cutting the ripe grass; raking the grass that has been cut off; loading the grass up to the trailer and transporting the grass to the pasture faraway. During each phase, a certain task has to be carried out by general tractor equipping with the related implement, i.e., mower, rake, and trailer, respectively. Throughout the whole mission, failed components are repairable regardless of whether the system is up or down. However, it is required that the entire mission has to be completed within $T_{\max}$ days, including the system downtime (i.e. for repairs) due to certain component failures. The roles and encountered load condition of the tractor system varies in different phase; besides, the system configuration, success criteria, and component behavior change from phase to phase. Thus, the tractor system can be termed as the PMS. Moreover, even though the working time for each task is determined according to the normal operating ability, the duration of each phase is uncertain because the repair times for failure components are random variables. But all the three-phase durations have to satisfy the relationship, represented as $t_1 + t_2 + t_3 \leq T_{\max}$; otherwise, the required mission is determined as failure.

In consequence, the repairable PMS with uncertain phase durations but limited by a maximum mission time is also studied in this paper. Note that this type of PMS is different from those addressed in literature [2, 5, 9-11, 15, 20-21, 27] mentioned above, whose phase durations are assumed to be deterministic. It has to concern with phase durations dependence, except for dynamic behaviors among phases, and components state dependence among phases. Moreover, not only the repairable components have bidirectional transitions between states of up and down, but also the system has the bidirectional state transition.

In existing research, Monte Carlo (MC) simulation, as a typical simulation method, has been adopted in analyzing non-repairable PMS [23,28]. Since MC simulation is superior due to its strong adaptability, it could be taken to cope with the reliability analysis of repairable PMS regardless of the complexity of the system. MC procedure is a way of carrying out numerical trails and based on a mapping model between inputs and outputs. The accuracy of the analysis outcome could be guaranteed by the reasonable number of simulation trails. As for an individual trail, the correctness of output corresponding to certain inputs depends on the mapping model in use. Therefore, it is important to particularly explore a rational and efficient modeling method that is compatible with the problem being studied.

The fault tree is a graphical tool for system reliability analysis; it has the advantages of being straightforward, being clear logical, and having semantic specification. Thus, it is widely used in reliability analysis on system failure criteria during each phase of PMS. Additionally, some researches adopted the OR gate as the first-level logical connection to construct the whole fault tree of PMS [22], i.e., the output is the state of the system, whereas inputs are all phases subtrees. By using the OR gate, it can display the fact that the system is determined to fail once any one phase fails; however, it cannot display the sequence behavior among phases, i.e. phase j will not fail before phase i when $i<j$. Therefore, it is not appropriate for applying the OR gate as the first-level logical connection. (It must be noted that the sequence behavior and state dependence among phases do have been

taken into account in investigation [22], even though they are not be properly displayed in the fault tree.)

In fact, one type of logical gate, called sequence-enforcing gate (SEQ gate) [1] is introduced to express constraints that all inputs are forced to occur in the left-to-right order. Obviously, it just fit the sequence behavior of PMS, that system mission has to be carried out phase by phase. Thus, Sequence-enforcing Fault Tree (SEFT) is proposed in this paper, in which an SEQ gate is adopted as the first-level logical connection to construct the whole fault tree of PMS. In that case, a complete relationship between the system and components can be accurately displayed by logic gates. Furthermore, a fault tree can be regarded as a hierarchical combination of several logic modules [8]. Each logic module is centered on a gate unit, while linking an output event and more than one inputs. The existing literature [8,12] shows that once operating rules of all gates could be expressed in a standard unified form, the modular model of the whole fault tree can be established. Therefore, how to establish the unified form that is available in various static/dynamic gates including the SEQ gate will be specially studied in this paper.

This paper is organized as follows. In Section 2, the two types of repairable PMS being studied are introduced. In Section 3, the proposed SEFT-MC method (SEFT-MC is short for Monte Carlo Simulation based on Modular Sequence-enforcing Fault Tree Model) is described in detail. In Section 4, the application of the SEFT-MC method is presented, in which the influence of whether phase duration is deterministic is discussed. Finally, conclusions are drawn in Section 5, as well as the direction of future research.

2. System description

Two types of repairable PMS are considered in this paper, in which both types comply with the same system structure and failure criteria, in detail:

- N components are included in the system.
- The system is required to undertake a mission, which consists of n phases. The switching time between the two phases is negligible.
- Each component has binary states, i.e., up and down; up implies the component working normally, whereas down implies component failure or in repair.
- Component failure & repair times are mutually s-independent which can obey different distributions rather than just the exponential distribution.
- The system is either in up or down state, which is determined by related components states, as well as the structure function.

Moreover, the different characteristics of the two types of repairable PMS are listed in Table 1, including different mission requirements and maintenance strategies.

3. Proposed SEFT-MC method for repairable PMS analysis

To evaluate the reliability of repairable PMS, a SEFT-MC method is proposed in this investigation. Utilizing this method, the whole fault tree (i.e. SEFT) is constructed to distinctly express interrelationships between the system state and components states; at this point, a modular reliability model could be developed, which is used to effectively support the further MC simulation procedure. The highlight of this method is the proposal of SEFT and how to transfer this graphical expression into a modular reliability model that is unaffected by the size and complexity of the fault tree.

3.1. Basic structure of SEFT

SEFT is proposed as the whole fault tree of PMS. Take a 3-PMS (short for PMS with 3 phases) for example, the basic structure of SEFT is shown in Fig. 1. The top event represents the state of a system that has to carry out a 3-phases mission; utilizing the SEQ-OR gate, it connects to all phase subtrees. Each subtree can be further explored by analyzing system failure criteria during the related phase.

As the core of an SEFT, SEQ-OR gate is a kind of SEQ gate, which not only restricts that the inputs must occur from left to right but also determines the output failure as long as any one input fails. By using the SEQ-OR gate as the first-level logical connection, the basic structure of SEFT is suitable for reliability analysis on PMS in various practical fields.

3.2. Improved linear algebra representation approach

An SEFT can be regarded as a hierarchical combination of several logic modules. A logic module, as shown in Fig. 2, includes a gate unit, m inputs (short for input events), and 1 output (shorts for output event). For each logic module, once inputs state transition is given,

Table 1. Differences between two types of repairable PMS

	Items	Type I	Type II
Multi-phases mission requirements	Phase durations	Determined values, i.e., T_1, T_2, ..., T_n	Random variables, i.e., t_1, t_2, ..., t_n
	Time of system in up state	Determined values, i.e., $T_1^{\mathrm{up}}=T_1$, $T_2^{\mathrm{up}}=T_2$, ..., $T_n^{\mathrm{up}}=T_n$	Determined values, i.e., T_1^{up}, T_2^{up}, ..., T_n^{up}
	Maximum mission time	Determined values, i.e. $T_{\max}= T_1+T_2+\ldots+T_n$	Determined values, i.e. $T_{\max}> T_1^{\mathrm{up}}+T_2^{\mathrm{up}}+\ldots+T_n^{\mathrm{up}}$
Maintenance strategies	Failed components repairable	Only repaired when the system is up	Unconditional repaired immediately
	The extent of repair	As good as a new one	As good as a new one
	System state bidirectional transition	No, down→up is not allowed	Yes, down→up is allowed unless the limited mission time is reached

according to operating rules of the gate, the output state transition can be determined. If the output is not the top event of the whole tree, it is also an input belonging to a logic module of the higher hierarchical level. Thus, as long as operating rules of various gates are established in a standard unified form, the modular reliability model of SEFT can be obtained by means of substitution layer by layer.

Liu proposed the linear algebra representation (LAR) approach in literature [8]. According to LAR, each state of a certain event is denoted by a *state unit vector*, and then the transition between any two states can be expressed as a matrix multiplication. Besides, how to express the operating rules of logic gates in a standard unified form is also introduced in literature [8], including 3 static gates (OR, AND, VOTING gates) and 3 dynamic gates (PAND, SPARE, FDEP gates). Then, we wonder if the operating rule of SEQ-OR gate can be expressed in the unified form, by directly applying the LAR approach.

Compared to other logic gates, the SEQ-OR gate has a very special feature. As shown in Fig. 1, an SEQ-OR gate connects more than one phase (as inputs), and these phases are carried out one by one. In other words, at any time during the system mission period, only 1 phase event is active, as well as its subtree. Once a phase is accomplished, it should be non-activated, and the next phase will be activated unless the whole system mission is fulfilled. However, the existing LAR approach supposes that all events included in gate operating are active. It can neither be used to distinguish whether the event is active or not; nor to support expressions of activated/non-activated action.

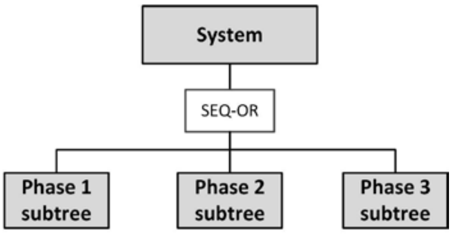


Fig. 1. Basic structure of SEFT

Thus, an improved linear algebra representation (I-LAR) approach is proposed as follows:

- For any event with 2 states, i.e., up and down, the *ordered index vector* is written as:

$$\boldsymbol{\alpha} = (1,2) \tag{1}$$

where each component of the row vector is called a *state number*[8]. Specifically, state1and 2 denote up and down, respectively.

- The *State unit vector* $\boldsymbol{S}_i$ s used to denote state i [8]. Here, $\mathbf{S}_i$ is a 2-dimensional unit column vector with "1" in the ith element and "0" in the other. In detail, corresponding to state1, i.e., up $\boldsymbol{S}_1 = \langle 1,0 \rangle$ corresponding to state2, i.e., down, $\boldsymbol{S}_2 = \langle 0,1 \rangle$. Then, the *state set* of a certain event is denoted as:

$$V = \{\boldsymbol{S}_1, \boldsymbol{S}_2\} \tag{2}$$

- In accordance with the statement in [8], the *state transition matrix* $\boldsymbol{T}_{pq}$ is used to express the instantaneous state transition by matrix multiplication. In detail, the transition from state p to state q can be represented by:

$$\boldsymbol{T}_{pq} \cdot \boldsymbol{S}_p = \boldsymbol{S}_q \tag{3}$$

where $\boldsymbol{S}_p$ and $\boldsymbol{S}_q$ are *state unit vector*s associated with state p and q, respectively; $\boldsymbol{T}_{pq}$ is an elementary switching matrix that transformed from the identity matrix by exchanging the pth and qth row vectors, and the dimensions of $\boldsymbol{T}_{pq}$, $\boldsymbol{S}_p$, $\boldsymbol{S}_q$ are the same.

In this paper, since p and q is either 1 or 2, bidirectional state transitions are specified as follows. Considering that a certain event transits from up to down at time t_i, *state transition* state(t_i^-) →state(t_i^+) can be represented by:

$$\boldsymbol{T}_{12} \cdot \boldsymbol{S}_1 = \boldsymbol{S}_2 \tag{4}$$

where $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are *state unit vector*s associated to state up and down, respectively $\boldsymbol{T}_{12}$; the *state transition matrix*, given by:

$$\boldsymbol{T}_{12} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{5}$$

Further, regarding the state transition from down to up, it can be represented by $\boldsymbol{T}_{21} \cdot \boldsymbol{S}_2 = \boldsymbol{S}_1$. Obviously, $\boldsymbol{T}_{21} = \boldsymbol{T}_{12}$, which is transformed from the 2-by-2 identity matrix by exchanging the 1st and 2nd row vectors.

- At any time during the system mission period, events included in SEFT may be either *inactive* or *active*.

① Definition 1. (Inactive event)

Inactive event is the event, whose state is impossible to transit.

② Definition 2. (*Active event*)

Active event is the event, whose state has the possibility to transit.

According to the definitions above, the top event is always an *active event* during the mission process. During the 1st phase, all events belonging to the phase 1 subtree are *active event*s, and events belonging to other phase subtrees are *inactive event*s. During other phases, things can be deduced in the same manner.

- For each event, it is assigned an *event vector* $\hat{\boldsymbol{H}}$

③ Definition 3. (*Event vector*)

Event vector $\hat{\boldsymbol{H}}$ is a 3-dimensional column vector, can be written as:

$$\hat{\boldsymbol{H}} = \begin{pmatrix} \delta \\ \boldsymbol{H} \end{pmatrix} \tag{6}$$

where $\boldsymbol{H}$ is a *state unit vector*, i.e. $\boldsymbol{H} \in V$; $\boldsymbol{\delta}$ is the index to distinguish whether the event is active or not, in specifically:

$$\delta = \begin{cases} 0 & \text{inactive} \\ 1 & \text{active} \end{cases} \tag{7}$$

- Thus, once an *inactive event* is activated, it can be expressed as:

$$\hat{\boldsymbol{H}}\Big|_{\delta=0} + \begin{pmatrix} 1 \\ \boldsymbol{o} \end{pmatrix} = \hat{\boldsymbol{H}}\Big|_{\delta=1} \tag{8}$$

where $\boldsymbol{o}$ s a 2-dimensional column vector with all 0 elements.

• On the contrary, once an *active event* is non-activated, it can be expressed as:

$$\hat{\boldsymbol{H}}\Big|_{\delta=1} - \begin{pmatrix} 1 \\ \boldsymbol{o} \end{pmatrix} = \hat{\boldsymbol{H}}\Big|_{\delta=0} \tag{9}$$

• According to the LAR approach proposed in literature [8], *state matrix* and *state number vector* are two concepts corresponding to the combination of m events. Among the m events, since *inactive event*s and *active event*s might co-exist, it is necessary to give new definitions.

④ Definition 4. (*State matrix*)

State matrix $\boldsymbol{X}$ is a matrix corresponding to m events. Only the *state unit vector*s of those *active event*s will be selected and sequentially combined into the *state matrix* $\boldsymbol{X}$.

In detail, it can be obtained as follows:

a) *Obtaining* $\boldsymbol{x}\left(j = 1, 2, \cdots, m\right)$

$\mathbf{x}_j$ is a 2-dimensional column vector, related to event j. It can be determined by the following equation:

$$\boldsymbol{x}_j = \Delta \cdot \hat{\boldsymbol{H}}_{\boldsymbol{j}} = \begin{pmatrix} 0 & \delta_j & 0 \\ 0 & 0 & \delta_j \end{pmatrix} \cdot \hat{\boldsymbol{H}}_{\boldsymbol{j}} \tag{10}$$

where $\hat{\boldsymbol{H}}_{\boldsymbol{j}}$ is the event vector of event j, and Δ s a 2-by-3 matrix, which mainly depends on the index δ_j

Obviously, Eq. (10) can be simplified as:

$$\boldsymbol{x}_j = \begin{cases} \boldsymbol{o} & \delta_j = 0 \\ \boldsymbol{H}_{\boldsymbol{j}} & \delta_j = 1 \end{cases} \tag{11}$$

where $\boldsymbol{H}_{\boldsymbol{j}}$ is the *state unit vector* of event j, and $\boldsymbol{o}$ s a 2-dimensional column zero vector.

b) *Obtaining* $\boldsymbol{X}$:

As long as $\boldsymbol{x}_j \left(j = 1, 2, \cdots, m\right)$ s not a zero vector, it will be selected in order as a column of $\boldsymbol{X}$.

Thus, the number of columns in the *state matrix* $\boldsymbol{X}$ may be less than m.

⑤ Definition 5. (*State number vector*)

Corresponding to *state matrix* $\boldsymbol{X}$, *state number vector* $\boldsymbol{XX}$ as a row vector is defined to denote the ordered collection of those *active event*s' *state number*s. It can be obtained by:

$$\boldsymbol{XX} = \alpha \cdot \boldsymbol{X} \tag{12}$$

3.3. Modular modeling of SEFT

For a logic module, as shown in Fig. 2, the state of output will not change unless one input has a state transition. Based on I-LAR approach introduced above, the operation process of a logic module can be described as follows:

• Given the following conditions:

a) At time t_i^-, the input *state matrix* is represented as $\boldsymbol{X}\left(t_i^-\right)$ and the corresponding *state number vector* is expressed as $\boldsymbol{XX}\left(t_i^-\right)$

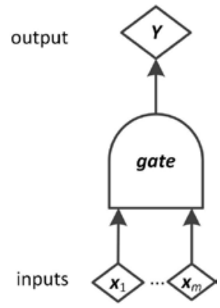


Fig. 2. Logic module of SEFT

b) At time t_i^+, the input *state matrix* is represented as $\boldsymbol{X}\left(t_i^+\right)$, and the corresponding *state number vector* is expressed as $\boldsymbol{XX}\left(t_i^+\right)$.

c) At time t_i^-, the output state is represented as $\boldsymbol{Y}\left(t_i^-\right)$, and the corresponding *state number* is expressed as p. In other words, the output state is $\boldsymbol{Y}\left(t_i^-\right) = \boldsymbol{S}_{\boldsymbol{p}}$

• To determine the output state at time. t_i^+ , represented as $\boldsymbol{Y}\left(t_i^+\right)$ it can be calculated by:

$$\boldsymbol{Y}\left(t_i^+\right) = \left(\boldsymbol{T}_{pq}\right)^k \boldsymbol{Y}\left(t_i^-\right) \tag{13}$$

where q is the output *state number* at time t_i^+ , and $k \in \{0,1\}$; the value of k is used to reveal whether the output state transition occurs or not, in detail, k =1 indicates the transition is triggered, whereas k =0 indicates that no state transition of the output happens instantly.

Compared to the statement in literature [8], the revised operation process has no difference but only those active input events are involved, owing to new definitions of *state matrix* and *state number vector*.

Obviously, variables k and q in Eq.(13) change as the gate unit in the logic module changes. The calculation of these two variables is determined by the operation rules of each gate.

• *OR gate*: Considering m inputs, as long as one active input is in the down state, the output is determined as down. In other words, the output state is the same as the worst active input state. Thus, the variable q is represented as:

$$q = \left\| \boldsymbol{XX}\left(t_i^+\right) \right\|_{\infty} \tag{14}$$

where $\| \ \|_{\infty}$ refers to the infinity norm of a certain vector.

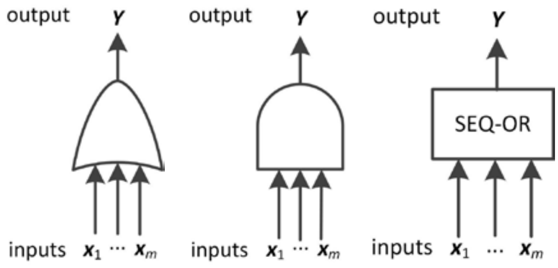
As for the variable k, since OR gate is a kind of static gate, the output state is only related to the combination of active inputs states at time t_i^+ , it can be determined that $k \equiv 1$

• *AND gate*: Considering m inputs, if and only if all active inputs are in the down state, the output is determined as down. In other words, the output state is the same as the best active input state. Thus, the variable q is represented as:

$$q = \left\| \boldsymbol{XX}\left(t_i^+\right) \right\|_{-\infty} \tag{15}$$

where $\| \ \|_{-\infty}$ refers to the negative infinity norm of a certain vector.

Since AND gate is a kind of static gate, it can be determined that $k \equiv 1$.



a) OR gate *b) AND gate* *c) SEQ-OR gate*

Fig. 3. Typical logical gate

- SEQ-OR gate: Considering m inputs, they have to be activated one by one from left to right; obviously, there is only 1 active input at any time; once the active input $\boldsymbol{x}_j$ is in down state, the output is determined as down. In other words, the output state is the same as the current active input state.

  Therefore, at time t_i^+, once the state of active input $\boldsymbol{x}_j$ transits, variables k and q are, respectively, represented as:

$$k = \min\left(\left\lfloor \frac{xx_1\left(t_i^+\right)}{2} \right\rfloor, 1\right) \tag{16}$$

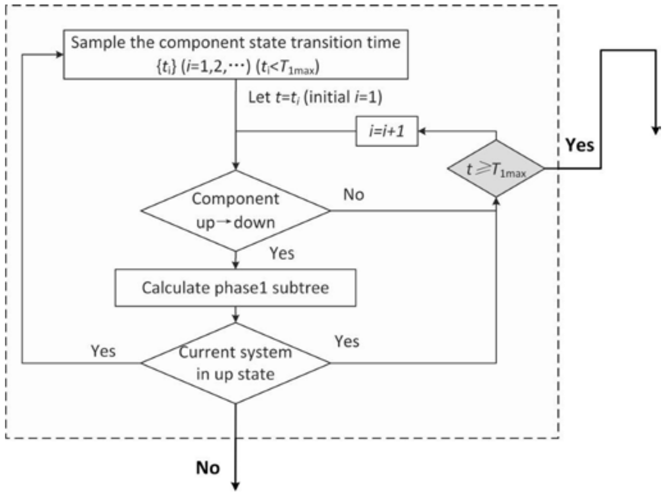
$$q = xx_1\left(t_i^+\right) \tag{17}$$

where $\lfloor \ \rfloor$ refers to the typical floor function.

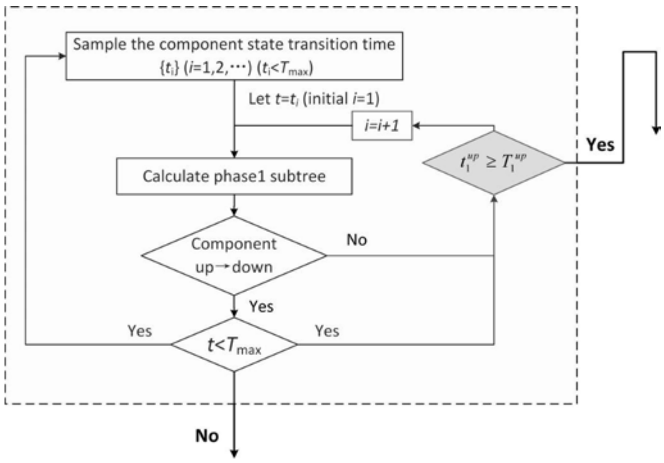
In summary, by means of the proposed I-LAR method, operation rules of different logic gates can be presented in unified forms of expression. And then, constructing a modular SEFT model for PMS reliability analysis is feasible by substitution of logic module layer by layer.

3.4. MC simulation based on the modular reliability model

Once the modular reliability model related to SEFT is obtained based on the statement above, MC simulation containing M trials is adopted to evaluate the reliability of repairable PMS. The basic flow chart is shown in Fig. 4, corresponding to a 3-PMS example. During the rth simulation trial, the procedure used to distinguish whether the current phase is successful or not may be repeated up to three times. The trial will not switch to the next phase until the current phase is completed, and the mission success is determined when the final phase has been fulfilled. Furthermore, take phase1 for example, the detailed distinguishing process for repairable PMS of Type I and Type II is different, as shown in Fig.5.



(a) Type I



(b) Type II

Fig. 5. Detailed distinguishing process for the current phase

(a) Type I

In the rth simulation trial, phase1 starts at the initial time t=0. Phase1 success is determined iff the trail time reaches the given phase duration $T_{1\max}$. It is regarded that the system keeps on operating until a component state transition time has reached. Since the component state transition might be bidirectional, two situations need to be discussed separately:

- Once one component state transits from up to down, the modular subtree of the current phase is calculated. As long as the system transits to the down state, the rth simulation trial ends. Otherwise, the repair time for a certain component is sampled according to its given distribution function; furthermore, the next failure time after its restoration is also sampled. Then, the trial moves to the next state transition time unless the trail time has come to the given phase duration $T_{1\max}$.
- Once one component state transits from down to up, obviously, the system will not change the status of normal operating. Thus, as long as the trail time is still shorter than the determined phase duration $T_{1\max}$, the trial will move to the next state transition time.

(b) Type II

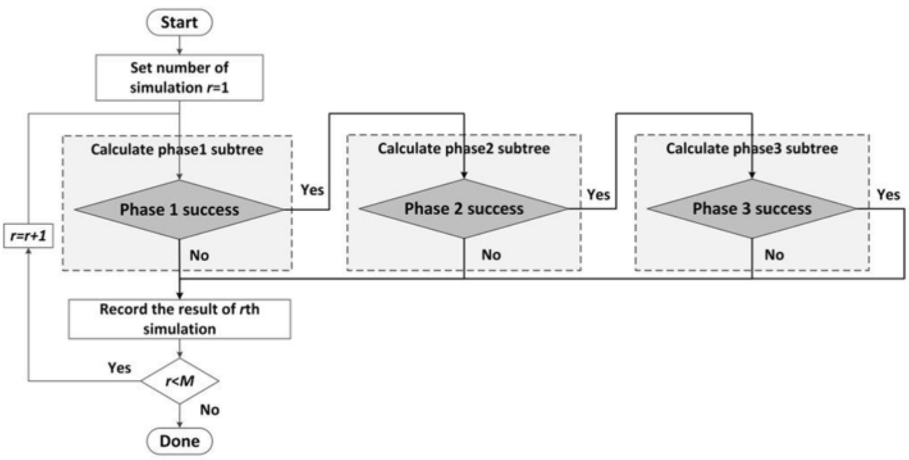


Fig. 4. Basic flow chart of MC simulation based on modular SEFT

In the rth simulation trial, phase1 also starts at initial time t=0. Phase1 success is determined iff the time of system in up state reaches the given value T_1^{up}. Different from Type I, the state of phase1 event has to be calculated once a component state transition time has reached. Then, regarding the two situations:

- Once the state transition is from down to up, as long as the system's operating time is still shorter than the given T_1^{up}, the trial in the current phase has to be continued.
- Once the state transition is from up to down, if the trail time has come to the maximum mission time $T_{\max}$, the rth simulation trial ends. Otherwise, the repair time for a certain component is sampled according to its given distribution function; furthermore, the next failure time after its restoration is also sampled. Then, the trial moves to the next state transition time unless the system's operating time in the current phase has come to the given T_1^{up}.

4. Numerical example

In this section, the application of the proposed SEFT-MC method is illustrated under two different cases corresponding to the two types of repairable PMS mentioned above. Furthermore, comparisons of the two cases are also be discussed afterward.

The system structure and failure criteria of both cases are identical, which is based on the example presented in the literature [2]. The system consists of 4 components, as shown in Fig.6; all the 4 components participate in phase1, whereas component D and B is not involved in phase 2 and 3, respectively. Each component has two states, i.e., up and down. The bidirectional state transition time of all components is exponentially distributed; the transition rates are shown in Table 2, and there is no difference in each phase. Here, failed components will be restored as good as a new one once it is repaired. Different mission requirements of the two cases are listed in Table 3.

Obviously, the SEFTs corresponding to two cases are identical, as shown in Fig. 7. The SEQ-OR gate is used as the specified first-level connection; besides, OR gates and AND gates are adopted to construct the phase subtrees. According to the I-LAR approach, the *state set* of each event is $V=\{\boldsymbol{S}_1,\boldsymbol{S}_2\}$, where the *state unit vectors* are $\boldsymbol{S}_1=\langle 1,0\rangle$ and $\boldsymbol{S}_2=\langle 0,1\rangle$. Then, *state transitions* of up→down and down→up are represented by $\boldsymbol{T}_{12}\cdot\boldsymbol{S}_1=\boldsymbol{S}_2$ and $\boldsymbol{T}_{21}\cdot\boldsymbol{S}_2=\boldsymbol{S}_1$, respectively, where $\boldsymbol{T}_{12}=\boldsymbol{T}_{21}=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$.

At the beginning of the mission, the initial state of all events is set as up. Besides, all events included in phase1 subtree are activated, as well as the top event. Thus, according to Eq.(6)-Eq.(7), the *event vector* of above events are $\hat{\boldsymbol{H}}=\langle 1\ \ 1\ \ 0\rangle$, others are $\hat{\boldsymbol{H}}=\langle 0\ \ 1\ \ 0\rangle$.

Once phase1 is successfully completed, the mission switches to phase2 instantly. Phase1 event is non-activated by subtraction according to Eq.(8), represented as $\hat{\boldsymbol{H}}=\langle 0\ \ 1\ \ 0\rangle=\langle 1\ \ 1\ \ 0\rangle-\langle 1\ \ 0\ \ 0\rangle=\langle 0\ \ 1\ \ 0\rangle$; whereas phase2 event is activated by addition according to Eq.(9), represented as $\hat{\boldsymbol{H}}=\langle 0\ \ 1\ \ 0\rangle+\langle 1\ \ 0\ \ 0\rangle=\langle 1\ \ 1\ \ 0\rangle$. As for component D, since it is not involved in phase2, during this phase, the corresponding event vector is set by subtraction, whereas the state will remain. Similar operations are also applied to phase3.

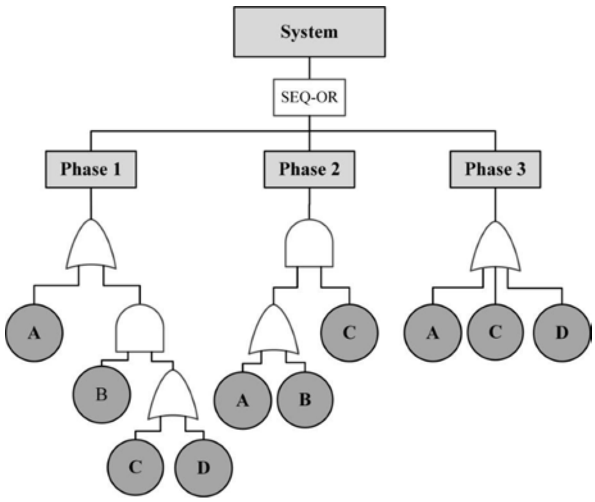
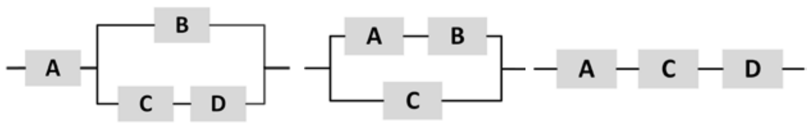


Fig. 7. SEFT of the 3-PMS example



(a) Phase 1 *(b) Phase 2* *(c) Phase 3*

Fig. 6. The structure for each phase of the discussed PMS

Table 2. State transition rates of components

State transition	Symbol	Component			
		A	B	C	D
Up→down	λ	0.1	0.2	0.3	0.4
Down→up	μ	0.2	0.3	0.4	0.5

Table 3. Differences of mission requirements between two cases

Items	Case I	Case II
Phase durations	T_1=1days, T_2=1days, T_3=2days	Random variables, represented as t_1, t_2, t_3
Time of system in up state	T_1^{up} =1days, T_2^{up} =1days, T_3^{up} =2days	T_1^{up} =1days, T_2^{up} =1days, T_3^{up} =2days
Maximum mission time	$T_{\max}$=4days	$T_{\max}$=6/10/14days

Then, according to the modular modeling introduced in section 3.3, once a state transition of any basic event occurs, the state of top event and intermediate event can be easily obtained through a series of matrix operations.

Further, MC simulation with M trials is carried out, in which the basic flow chart is shown in Fig. 4. As for one simulation trial, it switches to phase j (j=2, 3) iff the phase (j-1) has successfully completed, and the whole mission success is determined followed by the completion of phase 3. Corresponding to the case I and case II, the process to determine whether the current phase is succeeded or not is implemented by the flow chart, as shown in Fig. 5(a) and Fig. 5 (b), respectively. For case I, the criteria for success is that the trail time has come to the given phase duration; whereas the simulation trail will be interrupted, i.e., mission failed, once the system state transits from up to down. For case II, the criteria for success is that the time of system in up state reaches the given value; whereas the simulation trail will be interrupted once the trail time has come to the maximum mission time $T_{\max}$.

As for MC simulation, high accuracy and short computation time are contradict each other,

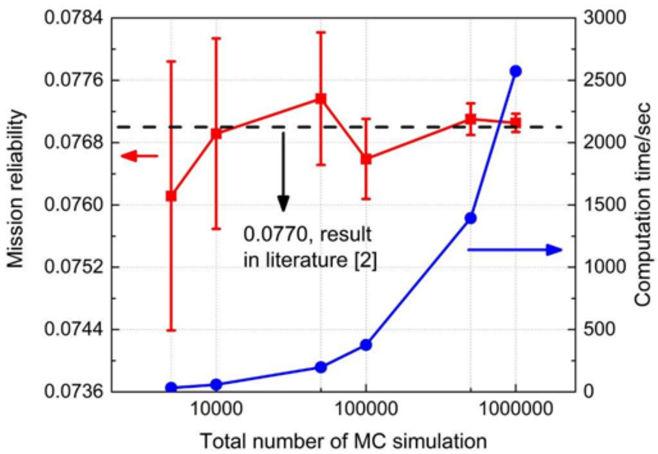


Fig. 8. Effect of total number of MC simulation on accuracy and computation time

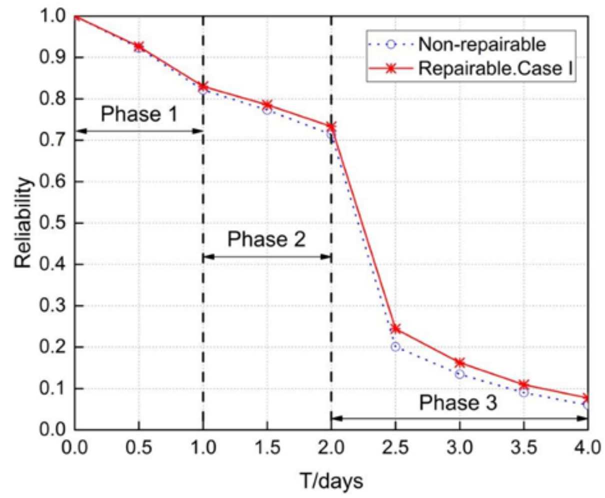


Fig. 9. Comparison of dynamic change in the reliability of the 3-PMS example

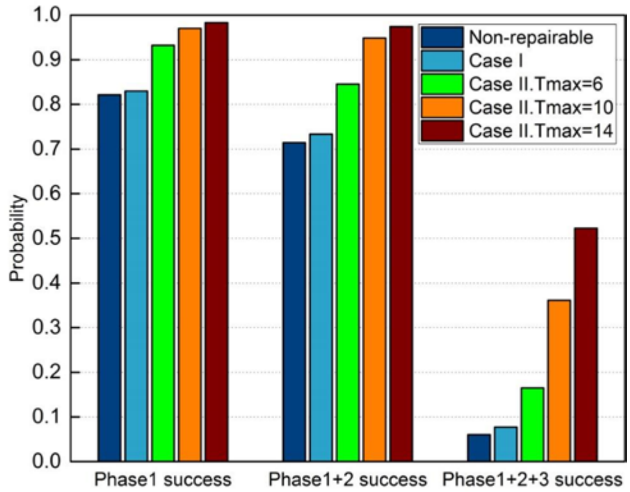


Fig. 10. Comparison of the success probability of the whole mission and each phase included

and they have different requirements for the total number M. The evaluation of mission reliability and computation time with increasing M are addressed using the proposed SEFT-MC method, as shown in Fig. 8. For a certain value of M, 10 repeated simulations are conducted and the corresponding results (including mean value and root mean squared error (RMSE) of evaluation, and average computation time) are given. With increasing M, the resultant values of mission reliability gradually tend to 0.077, which is consistent with the results in the study [2]. Meanwhile, the reducing RMSEs indicate improving convergence of results. When the value of M reaches 5×10^5, the reliability calculated by the SEFT-MC method is 0.077101 with the RMSE of 2.0276×10^{-4}, which is acceptable in this study.

Fig. 9 shows the dynamic changes in the reliability of repairable PMS discussed in case I. In order to discuss the effect of reliability improvement, the non-repairable PMS that has the same system structure and failure criteria is also considered according to SEFT-MC method. It is easy to find that maintenance strategy in case I can just slightly improve the system reliability, since only the component in the redundant structure may be repaired.

Due to the uncertainty of phase duration in repairable PMS discussed in case II, it is more meaningful to investigate the probability of success for each phase. As shown in Fig. 10, the system reliability in case II has significantly improved according to result comparison. Furthermore, the greater the maximum mission time is, the higher the probability to complete the whole mission and each phase included. Herein, in order to make sure the probability of mission success is higher than 50%, the maximum mission time should be set as 14 days, which is 3.5 times the required time of system in up state. That is to say, the reliability improvement is at the expense of increased mission time.

5. Conclusions and future work

Repairable PMSs abound in real-world applications. Due to the diversity of mission requirements and maintenance strategies, the analysis of repairable PMSs is much more complicated than that of non-repairable PMSs. In this paper, a novel SEFT-MC method is developed to evaluate the reliability of repairable PMS considering two types: to execute a multi-phases mission with deterministic phase durations, and within which failed components could be repaired only when the system is up; to execute a multi-phases mission with uncertain phase durations but limited by a maximum mission time, and within which failed components could be unconditional repaired immediately. The major characteristics of the proposed method are: the specific SEFT, whose core is the SEQ-OR gate, could be applied to a variety of PMS; the modular reliability modeling could make up for modeling inability of MC simulation itself; the manner to construct the modular reliability model has universal applicability due to the proposed I-LAR approach; the I-LAR approach allows the achievement that operational rules of various gates are expressed in standard form, and moreover, inputs included in the gate operating can be either active or inactive events. Furthermore, by means of a numerical example including two cases corresponding to the two types of repairable PMS, the application of the proposed method is demonstrated; in addition, the comparisons of two cases display that the significant improvement in reliability is at the expense of increasing mission time. This result could be useful for decision-makers on the optimal choice of maintenance strategies according to a comprehensive trade-off between reliability improvement and time cost. Consequently, a detailed study of such optimization problems will be conducted in our future work. Furthermore, how to improve the calculating efficiency by introducing some improved MC simulation methods will also be studied.

Considering the degradation of the system/components in PMS, the multi-state behavior will be introduced in analyzing PMS. In other words, the degradation process can be described in terms of transitions among multi-states (from perfectly working to totally failure). Therefore, the reliability assessment of non-repairable/repairable multi-state PMS is another direction of our future work.

Acknowledgment

This research is carried out with the Institute for Machine Elements and Systems Engineering (MSE) at RWTH Aachen University, which is led by Univ.-Prof. Dr.-Ing. Georg Jacobs; it is supported by the Jiangsu Government Scholarship for Overseas Studies Program and the Jiangsu Province "Double-Creation Project" Program. The authors are grateful for the reviewers' helpful comments.

References

1. Dugan JB, Bavuso SJ, Boyd MA. Dynamic Fault-Tree Models for Fault-Tolerant Computer Systems. IEEE Transactions on Reliability 1992; 41(3):363-377, https://doi.org/10.1109/24.159800.
2. Kim K, Park KS. Phased-Mission System Reliability under Markov Environment. IEEE Transactions on Reliability 1994; 43(2):301-309, https://doi.org/10.1109/24.295013.
3. Levitin G, Xing L, Amari, SV. Recursive algorithm for reliability evaluation of non-repairable phased mission systems with binary elements. IEEE Transactions on Reliability 2012; 61(2):533-542, https://doi.org/10.1109/TR.2012.2192060.
4. Li XT, Tao LM, Jia M. A Bayesian networks approach for event tree time-dependency analysis on phased-mission system. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2015; 17(2):273-281, https://doi.org/10.17531/ein.2015.2.15.
5. Li XY, Huang HZ, Li YF. Reliability analysis of phased mission system with non-exponential and partially repairable components. Reliability Engineering and System Safety 2018; 175:119-127, https://doi.org/10.1016/j.ress.2018.03.008.
6. Li XY, Li YF, Huang HZ, Zio E. Reliability assessment of phased-mission systems under random shocks. Reliability Engineering and System Safety 2018; 180:352-361, https://doi.org/10.1016/j.ress.2018.08.002.
7. Li YF, Peng R. Availability modeling and optimization of dynamic multi-state series-parallel systems with random reconfiguration. Reliability Engineering & System Safety 2014; 127:47-57, https://doi.org/10.1016/j.ress.2014.03.005.
8. Liu C, Chen N, Yang J. New method for multi-state system reliability analysis based on linear algebraic representation. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 2015; 229(5):469-482, https://doi.org/10.1177/1748006X15582129.
9. Lu JM, Lundteigen MA, Liu Y, Wu XY. Flexible truncation method for the reliability assessment of phased mission systems with repairable components. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18(2):229-236, https://doi.org/10.17531/ein.2016.2.10.
10. Lu JM, Wu XY. Reliability evaluation of generalized phased-mission systems with repairable components. Reliability Engineering & System Safety 2014; 121:136-145, https://doi.org/10.1016/j.ress.2013.08.005.
11. Lu JM, Wu XY, Liu Y, Lundteigen MA. Reliability analysis of large phased-mission systems with repairable components based on success-state sampling. Reliability Engineering and System Safety 2015; 142:123-133, https://doi.org/10.1016/j.ress.2015.05.010.
12. Merle G, Roussel JM, Lesage JJ, Perchet V, Vayatis N. Quantitative Analysis of Dynamic Fault Trees Based on the Coupling of Structure Functions and Monte Carlo Simulation. Quality and Reliability Engineering International 2016; 32(1):7-18, https://doi.org/10.1002/qre.1728.
13. Mo YC, Xing LD, Amari SV. A multiple-valued decision diagram based method for efficient reliability analysis of non-repairable phased-mission systems. IEEE Transactions on Reliability 2014; 63(1):320-330, https://doi.org/10.1109/TR.2014.2299497.
14. Mo YC, Zhong F, Zhao X. New results to BDD truncation method for efficient top event probability calculation. Nuclear Engineering and technology 2012; 44: 755-766, https://doi.org/10.5516/NET.03.2011.058.
15. Murphy KE, Carter CM, Malerich AW. Reliability analysis of phased-mission systems: A correct approach. 2007 Proceedings - Annual Reliability and Maintainability Symposium, RAMS 2007:7-12, https://doi.org/10.1109/RAMS.2007.328042.
16. Peng R, Zhai Q, Xing L, Yang J. Reliability of demand-based phased-mission systems subject to fault level coverage. Reliability Engineering and System Safety 2014; 121:18-25, https://doi.org/10.1016/j.ress.2013.07.013.
17. Tang Z, Dugan JB. BDD-based reliability analysis of phased-mission systems with multimode failures. IEEE Transactions on Reliability 2006; 55(2):350-360, https://doi.org/10.1109/TR.2006.874941.
18. Wang C, Xing L, Peng R, Pan Z. Competing failure analysis in phased-mission systems with multiple functional dependence groups. Reliability Engineering and System Safety 2017; 164:24-33, https://doi.org/10.1016/j.ress.2017.02.006.
19. Wu D, Peng R, Xing L. Recent Advances on Reliability of Phased Mission Systems. In: Li QL., Wang J., Yu HB. (eds) Stochastic Models in Reliability, Network Security and System Safety. JHC80 2019. Communications in Computer and Information Science, vol 1102. Springer, Singapore, https://doi.org/10.1007/978-981-15-0864-6_2.
20. Wu X, Wu X. Mission reliability modeling and evaluation of multi-mission phased mission system based on an extended object-oriented Petri net. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19 (2): 244-253, https://doi.org/10.17531/ein.2017.2.12.
21. Wu X, Wu Y. Extended object-oriented petri net model for mission reliability simulation of repairable PMS with common cause failures. Reliability Engineering & System Safety 2015; 136:109-119, https://doi.org/10.1016/j.ress.2014.11.012.
22. Xing L. Reliability evaluation of phased-mission systems with imperfect fault coverage and common-cause failures. IEEE Transactions on Reliability 2007; 56(1):58-68, https://doi.org/10.1109/TR.2006.890900.
23. Yang J, Liu F, Li F. Reliability Modeling and Simulation for Phased Mission System with Multi-mode Failures. Fire Control and Command Control 2011; 36(2):89-92.
24. Zhai Q, Xing L, Peng R, Yang J. Aggregated combinatorial reliability model for non-repairable parallel phased-mission systems. Reliability Engineering and System Safety 2018; 176:242-250, https://doi.org/10.1016/j.ress.2018.04.017.
25. Zhang S, Sun S, Si S, Wang P. A decision diagram based reliability evaluation method for multiple phased-mission systems. Eksploatacja i niezawodnosc - Maintenance and Reliability 2017; 19 (3):485-492, https://doi.org/10.17531/ein.2017.3.20.
26. Zhang T, Bai G, Guo B. Success probability model of phased mission systems with limited spares. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2012; 14(1):24-32.

27. Zhao J, Cai ZQ, Si W, Zhang S. Mission success evaluation of repairable phased-mission systems with spare parts. Computers and Industrial Engineering 2019; 132:248-259, https://doi.org/10.1016/j.cie.2019.04.038.
28. Zio E, Librizzi M. Direct Monte Carlo Simulation for the Reliability Assessment of a Space Propulsion System Phased Mission (PSAM-0067). Proceedings of the Eighth International Conference on Probabilistic Safety Assessment & Management (PSAM) 2006.

Chenxi LIU
School of Mechanical Engineering
Jiangsu University
Xuefu Road 301, 212013 Zhenjiang, China

Achim KRAMER
Stephan NEUMANN
Institute for Machine Elements and Systems Engineering
RWTH Aachen University
Schinkelstraße 10, 52062 Aachen, Germany

E-mails: liu_chenxi@ujs.edu.cn, achim.kramer@imse.rwth-aachen.de,
stephan.neumann@imse.rwth-aachen.de

Article citation info:
CUI X, LI T, WANG S, SHI J, MA Z. Reliability modeling based on power transfer efficiency and its application to aircraft actuation system. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 282–296, http://dx.doi.org/10.17531/ein.2020.2.11.

Xiaoyu CUI
Tongyang LI
Shaoping WANG
Jian SHI
Zhonghai MA

RELIABILITY MODELING BASED ON POWER TRANSFER EFFICIENCY AND ITS APPLICATION TO AIRCRAFT ACTUATION SYSTEM

MODEL NIEZAWODNOŚCI OPARTY NA WYDAJNOŚCI PRZESYŁU ENERGII I JEGO ZASTOSOWANIE DO OCENY LOTNICZEGO UKŁADU HYDRAULIKI SIŁOWEJ

The power transfer systems (PTS) has special reliability properties, including multiple states and fault dependence. Consequently, traditional binary-state reliability modeling methods cannot accurately evaluate the reliability of PTS. In order to resolve the contradiction between terminal energy demand and power transfer capability of PTS, this paper proposes a novel multi-state reliability model based on power transfer efficiency (PTE) for reliability evaluation of PTS. The multi-state model caused by performance degradation based on PTE is considered in this paper. In addition, the failure correlation in virtue of the system structure and energy allocation mechanism is analyzed in the proposed model, and the corresponding reliability evaluation result is obtained under different terminal energy requirements. The approach is verified on the example of a dual hydraulic actuation system (DHAS), in which the stochastic model based on the generalized stochastic Petri nets (GSPNs) is established and combined with the power transfer capability via universal generating function (UGF). Though changing flow rate to face the degradation rate of hydraulic pump, the reliability assessment of DHAS based on the proposed reliability model is effective and accurate.

Keywords: *reliability modeling, power transfer efficiency, multi-state performance, dual hydraulic actuation system, generalized stochastic Petri nets, universal generating function.*

Układy przesyłu energii (power transfer systems, PTS) charakteryzują się szczególnymi właściwościami niezawodnościowymi, w tym wielostanowością i zależnością między błędami. W związku z tym, tradycyjne metody modelowania niezawodności, które sprawdzają się w przypadku systemów dwustanowych, nie pozwalają na dokładną ocenę niezawodności PTS. W przedstawionej pracy zaproponowano nowatorski model niezawodności systemu wielostanowego, który do oceny niezawodności PTS wykorzystuje dane o wydajności przesyłu energii (PTE). Model ten wiążę niezawodność zarówno z zapotrzebowaniem na energię końcową jak i zdolnością przesyłową PTS. Rozważano model wielostanowy opisujący proces degradacji komponentów systemu w oparciu o PTE. W proponowanym modelu analizowano korelacje między uszkodzeniami w świetle struktury systemu i mechanizmu alokacji energii, a niezawodność oceniano dla różnych stopni zapotrzebowania na energię końcową. Podejście to zweryfikowano na przykładzie podwójnego układu hydrauliki siłowej (DHAS), dla którego ustalono model stochastyczny oparty na uogólnionych stochastycznych sieciach Petriego (GSPN), który łączono ze zdolnością przesyłową za pomocą uniwersalnej funkcji tworzącej (UGF). Badania pompy hydraulicznej prowadzone dla różnych prędkości przepływu i różnych szybkości degradacji wykazały, iż ocena niezawodności DHAS na podstawie proponowanego modelu cechuje się skutecznością i trafnością.

Słowa kluczowe: *modelowanie niezawodności, wydajność przesyłu energii, działanie systemu wielostanowego, podwójny układ hydrauliki siłowej, uogólnione stochastyczne sieci Petriego, uniwersalna funkcja tworząca.*

1. Introduction

The systems in engineering field, in terms of their basic functions, can be divided into signal transfer system (STS), power transfer system (PTS), and mass transfer system (MTS). An STS, e.g. a sensor network, receives and transfers a series of analog/digital signals between terminals [9]. A PTS, e.g. a power transmission network, transforms and transmits energy between energy generators and consumers. An MTS transports mass between terminals. Due to the similar transmission characteristics derived from conservation law, an MTS can also be regarded as a special type of PTS [3, 7]. In this paper, only STS and PTS are considered. The correct implementation of functions requires that both STS and PTS have a complete transfer channel during the process of signal and energy transmitting. The impacts of random internal and external stresses acting on STS and PTS result in stochastic behaviors of transmission function. Hence, the concept of reliability is adopted to describe STS and PTS's capability to fulfill their transmission function under stated conditions for a specific period of time [24]. However, there are big differences between reliability characteristics existing in STS and PTS. Firstly, an STS is a binary-state system, which means that the signal transmission is either "correct" or "incorrect". Therefore, "correct" or "incorrect" becomes a criterion to assess whether an STS fulfills its function or not. However, it is difficult to accurately assess a PTS's transmission quality based on "correct" and "incorrect", because the system exists obvious degradation states, viz. multi-state properties during power transmission process [5]. For example, the wear process of hydraulic pump results in multiple pressure and flow output probabilities. Hence, whether a hydraulic pump fulfills its function depends not only on its states, but also on the terminal's energy requirements. Furthermore, an STS is generally a weak fault-coupling system, which means that a failed component in STS has small probability to influence other components, thus resulting in the component and/or system loss of function. The situation is completely different for a PTS, where a failed unit will change the energy distribution which may result in overload on other units, and ultimately these overloaded units will fail. For example, the hydraulic unit's sticking or jamming electro-hydrostatic actuator(EHA) on an aircraft may reduce hydraulic component's energy transmission efficiency, which will increase its forward motor's load, causing portion of the energy entering into the motor to be transformed to heat, which increases the probability of the motor failure [33, 35].

As discussed above, the different characteristics between STS and PTS raise challenges to assess and model the reliability uniformly. The traditional reliability model and assessment is established and conducted based on fault data with statistical theory [37]. This method can provide the probability density function based on the historical fault data and is suitable for STS' reliability assessment. Due to the lack of considering the fault behavior process, this reliability model cannot accurately describe the fault coupling among the components in PTS, and it also fails to assess the impact of multiple states on terminal's demands. In order to describe fault coupling relationship, researchers have developed numerous fault correlation models, e.g. copula models [14]. However, since these models describe the probability correlation among multiple faults based on data statistics, in the case of insufficient statistical data these models cannot reflect the probability characteristics of the faults. In order to build an accurate model of component's failure behavior in PTS, the physics of failure (PoF) method [31] is a good alternative. The PoF methods have been used quite successfully in modeling mechanical, civil, and aerospace structures. Reliability models based on failure physics can effectively capture failure processes, like fatigue, creep, wear, plastic deformation and instability. Peng et al. [22] applied PoF model to calculate the reliability of a micro-motor and determined the optimal maintenance strategy. Subsequently, it is found that the transfer of power along PTS will also affect its performance.

As a typical case of multi-state PTS, aircraft actuation system has distinct component degradation process and multiple power transfer states. Generally, aircraft actuation system with similar redundancy consists of three cross-linking units, viz. three independent hydraulic power sources, dissimilar redundant flight control computers and multiple redundant hydraulic actuators [26]. The effect of redundancy on risk and reliability can be calculated by using fault tree analysis (FTA) and reliability block diagram (RBD), but they are not very suitable for dynamic systems [1]. The critical components in aircraft power and actuation system are designed to be highly reliable to ensure safety, since the internal degradation and external events can cause extreme damage [29]. The Markov process method is more feasible to model both internal technical failures and external events dynamically [6]. Cross linking in three independent redundant hydraulic actuation systems can increase the survive channel under failures, then the reliability model based on failure monitoring and failure handoff appears to evaluate the system reliability under multiple faults [34]. More reliable and safety technologies make the flight control system more complex with dissimilar redundancy and multi separate control surfaces, e.g. A380 and B787 [25], in which surfaces are driven by both hydraulic power and electrical power. Accordingly, the aircraft actuation system based on multiple power resources and heterogeneous driving system becomes complicated cyber-physical system, in which the power supply systems, control computers and the actuators are different and complex fault tolerant strategies are designed. Under these circumstances, even if there are multiple faults, the performance of the actuation system can remain normal or slightly degraded, which is a specific multi-state system (MSS). The research on MSS began in the 1970s [8]. In recent years, multi-state systems have been widely studied to describe multiple states of performance or the degradation process of components and systems. Various approaches have been proposed to estimate the reliability of MSS [19], including universal generating function (UGF) technique [11], stochastic process approach [23], and Monte Carlo simulation technique [30]. X. Li. et.al [15] even consider the multi-phased characteristics in the multi-state reliability assessment of the spacecraft. The control surface requires fast, accurate and smooth movement according to the control command. A specific amount of power flow transmitted from the input to the output measures the reliability considering flight performance. The reliability evaluation based on stochastic-flow network is first presented by Lee S H [12] and developed by Lin [17]. Lin [16] presented multi-state reliability method where Markov process was used to calculate the probability of states. If the state transfer is considered by time interval, the Markov process can be described as a Poisson process. Since Markovian method cannot characterize stress induced failure, Malinowski [21] presented corresponding numerical algorithm. Ushakov [27] presented UGF to solve the NP-hard problem of multi-state methods by defining the universal generating operator (UGO). Levitin [13] and Lisnianski [18] extended UGF to z-transform and verified this method on the physical system.

Aiming at modeling the reliability characteristics of aircraft actuation system, this paper presents a power flow-based reliability model, in which the power can be adjusted by pilot command and the fault reflected power loss. The multiple combination among power supply, computer and actuator under fault needs multi-state method to model its internal fault and fault transfer in the system. Multiple states in actuation system is reflected by the combinations of components' power transfer efficiency (PTE) which can be figured out by PoF analysis [10]. As an intuitive and effective method to analyze the dynamic processes of multi-elements degradation interior state, general stochastic Petri nets (GSPNs) [32]are adopted in this paper to model the functional states and degradation states of each component. In order to avoid solving massive differential equations, this paper adopts the UGF technique to calculate the system probability through algebraic calculus on the probability of component [28]. Through discretizing the continuous random variables and defining their probability distributions, a UGO can be obtained, then the availability can be calculated based on the UGF model [13, 18]. So, the multi-state reliability model via UGF is quite suitable for the case we studied.

The rest of this paper is presented as follows. In section 2, the reliability model based on PTE is formulated and the corresponding solution algorithm is introduced. With the UGF method, this paper calculates the probability of inner state of PTS. In Section 3, the case study of DHAS' reliability, which is a redundant actuation system composed of dual hydraulic source/hydraulic actuator and utilized in large aircraft [4], is computed and analyzed. In Section 4, conclusions are drawn.

2. Reliability assessment based on power transfer efficiency

Conventional reliability is defined as the probability that the component can perform its required function under stated conditions for a specified period of time [36]. Assuming that T is a random variable representing the failure time of a component, then the reliability is defined as the probability that the system will perform its expected function under the specified conditions of a given environment over a specified period of time t, namely:

$$R^{c} = P(T > t) = \int_{t}^{\infty} f(t)dt \tag{1}$$

where $f(t)$ is the failure probability density of the component, and $R^{c}(t)$ is the reliability of component at time t. This reliability model is based on random failure mechanism under binary states, which is suitable for the reliability assessment of STS.

Practically, the performance of dynamic system is strongly related to its power supply and transmission that it is difficult to assess the system with only by "correct" or "incorrect". For example, the hydraulic actuator of aircraft drives the control surface in a very wide frequency band, but the leakage of valve and cylinder will consume its power supply and degrade its performance. Therefore, whether a PTS fulfills its function depends not only on the states of its components, but also on the terminals' energy demands. Assuming that W is the actual power supply provided for the PTS, and W_{th} is the minimum power requirement to guarantee its normal operation , then the reliability of PTS can be described as:

$$R^{\text{p}} = P(t \mid W \geq W_{\text{th}}). \tag{2}$$

Eq. (2) implies that the system power supply is sufficient to consumers' requirements over the whole lifetime of the system. However, any component of PTS will dissipate the power during the operating process. It is difficult to keep enough power for the terminal output after experiencing power losses from multiple components. For instance, the aircraft actuation system could not drive the control surface according to the corresponding demand because of the internal and external leakage of system. In this situation, it is necessary to build the reliability model based on power supply and power transfer. This kind of reliability model can describe the failure process of PTS with the time increasing and the power being consumed.

Example 1: For a specific aircraft aileron control (shown in Fig. 1), its dynamic design processes are shown as follows in order to meet the actual requirements during each phase of flight profile. Suppose the control signal of aileron actuator is $x = x_0 \cdot \sin\omega t$, in which x_0 is the stroke of cylinder and ω is the frequency of actuator. According to the rated requirement, the maximum motion speed of actuator is $\dot{x}_{\max} = 0.0838\text{m/s}$ and the minimum output force ensuring the normal movement of surface is F=7500N , then the reliability of aileron actuator can be described as:

$$R^{\text{p}}(t) = P(t \mid W \geq F \cdot \dot{x}_{\max}) = P(t \mid W \geq 628.5\ \text{W}). \tag{3}$$

where W is the actual power transferred by the actuation system to drive the aileron and $W_{\text{th}} = F \cdot \dot{x}_{\max} = 628.5\text{W}$ is the minimal allowed power that can drive the surface under normal operational condition.

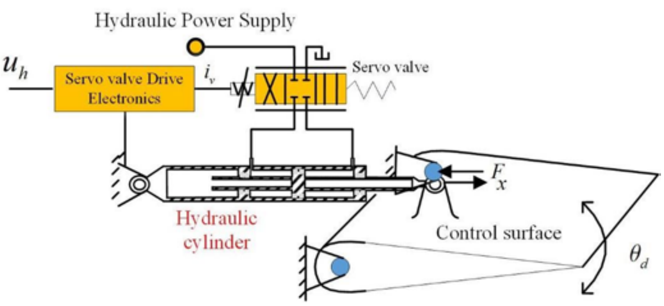


Fig. 1. The movement of aircraft aileron control

2.1. Assumptions and the general model

The reliability model of a PTS is developed based on the following assumptions:

- A PTS consists of n power transfer units (PTUs), in which the input power, the output power and the dissipated power of the $\text{PTU}_j (j = 1,2,\cdots,n)$ obey the power conservation law.
- Let η_j be the PTE of the $\text{PTU}_j (j = 1,2,\cdots,n)$. η_j decreases with the operational time increasing.
- The degradation of PTU_j can be described as a multi-state process with the occurrence and development of failure.
- The state of the $\text{PTU}_j (j = 1,2,\cdots,n)$ is statistically independent and the state duration obeys exponential distribution.
- c_j is the maximum allowed power capability under normal condition of the PTU_j , which is a constant.

As a fundamental physical component, the PTU and its performance determines PTS' power transfer capability. In terms of power transmission process, $\text{PTU}_j (j = 1,2,\cdots,n)$ in PTS can be described as is shown in Fig. 2.

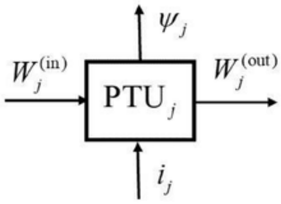


Fig. 2. A generic model based on power

For PTU_j , the input power $W_j^{(\text{in})}$ entering PTU_j is transmitted to its downstream PTUs with the amount of $W_j^{(\text{out})}$, while portion of power Ψ_j, is dissipated in the forms of heat and vibration. i_j represents the control signal of the unit according to the system requirement. Hence:

$$W_j^{(\text{in})} = W_j^{(\text{out})} + \psi_j \tag{4}$$

where $\eta_j = W_j^{(\text{out})} \big/ W_j^{(\text{in})}$ defines the PTE of PTU_j

2.2. Stochastic power transmission model of PTU

Assuming that the initial power transmission capability of a PTU_j meets the PTS power requirements, the unit degrades with the operational time, as is shown in Fig. 3. The degradation process of PTU_j is divided into four states, viz. normal operation, light degradation, serious degradation and failure and the PTE η_j can be described as $\boldsymbol{\eta}_j = \left[\eta_j^{(0)}, \eta_j^{(1)}, \eta_j^{(2)}, \eta_j^{(3)}\right]$. As a result, it is necessary to consider the multi-state process in power transmision model of PTU.

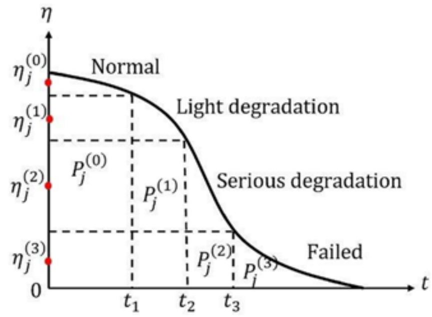


Fig. 3. Multi-state process of PTU_j

$\mathrm{PTU}(\boldsymbol{\eta}_j, \mathbf{A}_j, \mathbf{P}_j(t))$ is defined as a stochastic power transmission model for PTU_j, where $\boldsymbol{\eta}_j = \left[\eta_j^{(0)}, \eta_j^{(1)}, \eta_j^{(2)}, \eta_j^{(3)}\right]^{\mathrm{T}}$ denotes the PTE of PTU_j at normal state, light degradation state, serious degradation state and failed state, respectively. $\boldsymbol{P}_j(t) = \left[P_j^{(0)}(t), P_j^{(1)}(t), P_j^{(2)}(t), P_j^{(3)}(t)\right]^{\mathrm{T}}$ denotes the corresponding probabilities related to PTU_j at each state and $\boldsymbol{A}_j$ defines a stochastic process model relating PTU_j's degradation with failure development.

In terms of its PTE, PTU_j's operating states s_j are divided into four discrete intervals, $s_j \in \{s_0, s_1, s_2, s_3\}$, where different states denote that PTU_j has different power transmission capabilities, i.e.

$$\boldsymbol{W}_j^{(\mathrm{out})} = \begin{cases} w_j^{(0)} = \eta_j^{(0)} W_j^{(\mathrm{in})} & s_j = s_0 \\ w_j^{(1)} = \eta_j^{(1)} W_j^{(\mathrm{in})} & s_j = s_1 \\ w_j^{(2)} = \eta_j^{(2)} W_j^{(\mathrm{in})} & s_j = s_2 \\ w_j^{(3)} = \eta_j^{(3)} W_j^{(\mathrm{in})} & s_j = s_3 \end{cases} \tag{5}$$

Example 2: A hydraulic pump's PTE decreases as its components experiencing wear and tear. Fig. 4 (a) illustrates the wear process of slippers inside a hydraulic pump. According to the Archard's model, the abrasion loss of slippers can be described as follows:

$$\frac{d\varphi_w}{dt} = k_w \frac{q_w}{A_w} \tag{6}$$

where φ_w is the abrasion loss, $d\varphi_w / dt$ is the wear rate, k_w is constantly related to the surface condition and lubrication within the friction pair, q_w is the load on the wear surface, and A_w is the contact area of the wear surface. The abrasion loss curve is illustrated in Fig. 4 (b). As the slippers wear over time, the volumetric efficiency (VE) of the pump decreases, and it leads to the reduction of the effective power of the output, as shown in Fig. 4 (c).

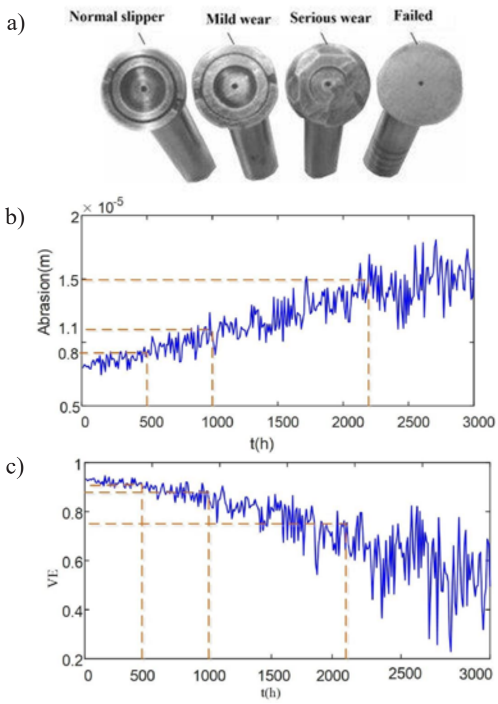


Fig. 4. Illustration of multi-state for hydraulic pump caused by wear, a) Wear process of slipper in pump, b) Abrasion loss of slippers, c) VE of the pump

According to Fig. 4 (b), during the time interval $[0, 500\mathrm{h})$, the slippers' abrasion is less than $0.8 \times 10^{-5}\,\mathrm{m}$. The corresponding VE of the pump is in the interval of [0.92, 0.95], and can be regarded as a normal state of the hydraulic pump. During the time interval [500h,1000h), the abrasion loss of slippers is in the interval of [$0.8 \times 10^{-5}\,\mathrm{m}, 1.1 \times 10^{-5}\,\mathrm{m}$) and the VE of the pump is [0.9,0.92). This state is defined as a light degradation state. As the pump continue to be used, in the time interval [1000h,2000h) the slippers' abrasion is in the interval of [$1.1 \times 10^{-5}\,\mathrm{m}, 1.5 \times 10^{-5}\,\mathrm{m}$) and the corresponding VE of the pump is [0.75, 0.9). Fig. 4 (c), describes a serious degradation state of the hydraulic pump. After 2000 hours, the slippers' abrasion becomes serious and the corresponding VE is less than 0.75, which cannot meet the demand of the downstream PTUs. We define this state as a failed state for the hydraulic pump.

In order to characterize the development of failure, GSPN model is used to establish the internal state transition process of PTU_j. Defining $\boldsymbol{A}_j$ as the state transition matrix, the probabilities of PTU_j, $\boldsymbol{P}_j(t) = \left[P_j^{(0)}(t), P_j^{(1)}(t), P_j^{(2)}(t), P_j^{(3)}(t)\right]^{\mathrm{T}}$ can be obtained by:

$$\boldsymbol{P}_j(t+1) = \boldsymbol{A}_j \bullet \boldsymbol{P}_j(t) \tag{7}$$

$\boldsymbol{s}_j = \left\{s_j^{(k)} \mid \left(P_j^{(k)}(t), \eta_j^{(k)}\right), k = 0,1,2,3\right\}$ describes the power transfer state of PTU_j at time *t*. UGF method is applied to simplify the calculation of the PTU state and its corresponding probability. For PTU_j, the PTE-based model, $\mathrm{PTU}(\boldsymbol{\eta}_j, \boldsymbol{A}_j, \boldsymbol{P}_j(t))$, can be written as:

$$\mathrm{PTU}(\boldsymbol{\eta}_j, \boldsymbol{A}_j, \boldsymbol{P}_j(t)) = P_j^{(0)}(t) z^{\eta_j^{(0)}} + P_j^{(1)}(t) z^{\eta_j^{(1)}} + P_j^{(2)}(t) z^{\eta_j^{(2)}} + P_j^{(3)}(t) z^{\eta_j^{(3)}} \tag{8}$$

2.3. Reliability model based on power transfer efficiency

For a PTS with n PTUs, as is shown in Fig. 5, the input power $W_{\mathrm{sys}}^{(\mathrm{in})}$ is transmitted to terminal users in channels composed of individual PTUs in the form of parallel or serial layout. The PTE of a PTS also depends on multiple states that $s_{\mathrm{sys}}^{(i)} = \left(P_{\mathrm{sys}}^{(i)}(t), \eta_{\mathrm{sys}}^{(i)}\right), i = 0,1,2,\ldots,4^n - 1$ represents the i^{th} state of PTS at time *t*, which means that the PTS has PTE $\eta_{\mathrm{sys}}^{(i)}$ with the probability of $P_{\mathrm{sys}}^{(i)}(t)$.

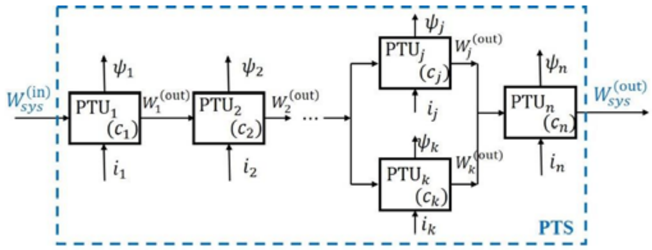


Fig. 5. Illustration of PTS structure

Given PTU's PTE-based models, $\mathrm{PTU}(\boldsymbol{\eta}_j, \boldsymbol{A}_j, \boldsymbol{P}_j(t))$, $j = 1,2,\ldots,n$, the PTS' model can be described as:

$$\mathrm{PTS}(\eta_{\mathrm{sys}}, \boldsymbol{P}_{\mathrm{sys}}(t)) = \mathrm{PTS}(\Theta(\eta_1, \eta_2, \ldots, \eta_n), g(\boldsymbol{P}_1(t), \boldsymbol{P}_2(t), \ldots \boldsymbol{P}_n(t))) \tag{9}$$

where the structural functions Θ and g describe mathematical relationships of PTE and corresponding probabilities of PTS and its components. The forms of Θ and g are determined by layout structure among PTUs. In engineering application, there are two typical connecting structures among PTUs, i.e. series layout and parallel layout.

2.3.1. PTE-base reliability model for series system

For a structure serialized by m PTUs, as shown in Fig. 6, the model of PTS can be described by Eq. (10).

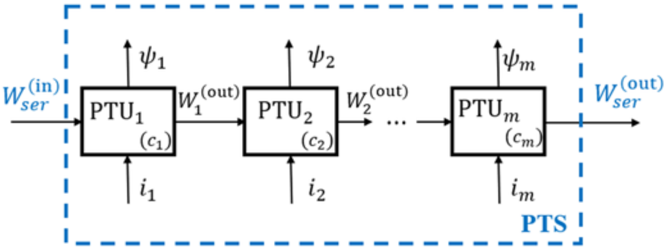


Fig. 6. Illustration of a structure serialized by m PTUs

$$\mathrm{PTS}(\eta_{ser}, \boldsymbol{P}_{ser}(t)) = \mathrm{PTU}(\eta_1, \boldsymbol{A}_1, \boldsymbol{P}_1(t)) \otimes \ldots \otimes \mathrm{PTU}(\eta_m, \boldsymbol{A}_m, \boldsymbol{P}_m(t))$$
$$= \sum_{i_1=1}^{4}\sum_{i_2=1}^{4}\ldots\sum_{i_m=1}^{4}\left(\left(\prod_{j=1}^{m}\prod_{k=1}^{4} P_j^{(i_k)}(t)\right) z^{\prod_{j=1}^{m}\prod_{k=1}^{4}\eta_j^{(i_k)}}\right) \tag{10}$$

The i^{th} state in $s_{ser}^{(i)} = \left(P_{ser}^{(i)}(t), \eta_{ser}^{(i)}\right)$ is $s_{ser}^{(i)} \in (s_1^{i_1}, s_2^{i_2}, \cdots s_m^{i_m})$ and its PTE can be calculated by:

$$\begin{aligned} W_1^{(\mathrm{out})} &= W_{ser}^{(\mathrm{in})} \cdot \eta_1^{\mathrm{i}_1} \\ W_2^{(\mathrm{out})} &= W_1^{(\mathrm{out})} \cdot \eta_2^{\mathrm{i}_2} \\ &\cdots \\ W_{ser}^{(\mathrm{out})} &= W_{m-1}^{(\mathrm{out})} \cdot \eta_{\mathrm{m}}^{\mathrm{i}_{\mathrm{m}}} \end{aligned} \tag{11}$$

The PTE for the serial system shown in Fig. 6 is $\eta_{ser}^{(i)} = W_{ser}^{(\mathrm{out})} / W_{ser}^{(\mathrm{in})} = \eta_1^{(i_1)} \cdot \eta_2^{(i_2)} \cdots \eta_m^{(i_m)}$ and the corresponding probability is $\mathrm{P}_{\mathrm{sys}}^{(i)}(t) = \mathrm{P}_1^{(i_1)}(t) \bullet P_2^{(i_2)}(t) \cdots P_m^{(i_m)}(t)$.

Example 3: Assume that two PTUs are serially connected. The operator $\otimes$ is defined to describe the serial operation and the i^{th} state probability of the PTS and its corresponding PTE is given by Eq. (12), with four states in each PTU.

$$\begin{aligned} \mathrm{PTS}(\eta_{ser}, \boldsymbol{P}_{ser}(t)) &= \mathrm{PTU}(\eta_1, \boldsymbol{A}_1, \boldsymbol{P}_1(t)) \otimes \mathrm{PTU}(\eta_2, \boldsymbol{A}_2, \boldsymbol{P}_2(t)) \\ &= P_1^{(0)}(t)P_2^{(0)}(t)z^{\eta_1^{(0)}\eta_2^{(0)}} + P_1^{(0)}(t)P_2^{(1)}(t)z^{\eta_1^{(0)}\eta_2^{(1)}} + P_1^{(0)}(t)P_2^{(2)}(t)z^{\eta_1^{(0)}\eta_2^{(2)}} + P_1^{(0)}(t)P_2^{(3)}(t)z^{\eta_1^{(0)}\eta_2^{(3)}} \\ &+ P_1^{(1)}(t)P_2^{(0)}(t)z^{\eta_1^{(1)}\eta_2^{(0)}} + P_1^{(1)}(t)P_2^{(1)}(t)z^{\eta_1^{(1)}\eta_2^{(1)}} + P_1^{(1)}(t)P_2^{(2)}(t)z^{\eta_1^{(1)}\eta_2^{(2)}} + P_1^{(1)}(t)P_2^{(3)}(t)z^{\eta_1^{(1)}\eta_2^{(3)}} \\ &+ P_1^{(2)}(t)P_2^{(0)}(t)z^{\eta_1^{(2)}\eta_2^{(0)}} + P_1^{(2)}(t)P_2^{(1)}(t)z^{\eta_1^{(2)}\eta_2^{(1)}} + P_1^{(2)}(t)P_2^{(2)}(t)z^{\eta_1^{(2)}\eta_2^{(2)}} + P_1^{(2)}(t)P_2^{(3)}(t)z^{\eta_1^{(2)}\eta_2^{(3)}} \\ &+ P_1^{(3)}(t)P_2^{(0)}(t)z^{\eta_1^{(3)}\eta_2^{(0)}} + P_1^{(3)}(t)P_2^{(1)}(t)z^{\eta_1^{(3)}\eta_2^{(1)}} + P_1^{(3)}(t)P_2^{(2)}(t)z^{\eta_1^{(3)}\eta_2^{(2)}} + P_1^{(3)}(t)P_2^{(3)}(t)z^{\eta_1^{(3)}\eta_2^{(3)}} \end{aligned} \tag{12}$$

2.3.2. PTE-based reliability model for parallel system

Fig. 7 shows a parallel structure of a PTS, in which PTUs transfer the input power to terminal customers. Assume that the input of j^{th} PTU is $c_j \cdot W_{par}^{(\mathrm{in})} \Big/ \sum_{i=1}^{m} c_i$ where c_j denotes the capacity of PTU$_j$, which indicates the ability to cope with the input power. The capacity of each PTU determines the power allocation among components.

The parallel structure in a PTS' model can be described as:

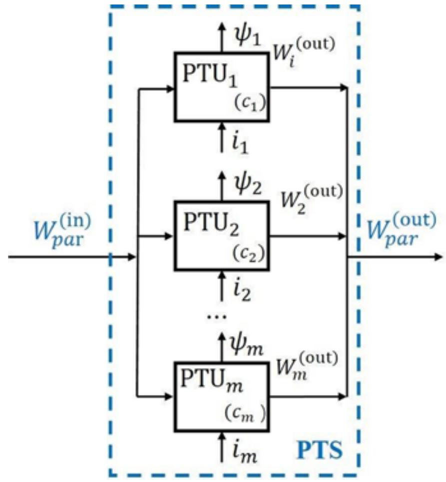


Fig. 7. Illustration of a parallel structure by m PTUs

$$\mathrm{PTS}(\eta_{par}, \boldsymbol{P}_{par}(t)) = \mathrm{PTU}(\eta_1, \boldsymbol{A}_1, \boldsymbol{P}_1(t)) \oplus \ldots \oplus \mathrm{PTU}(\eta_m, \boldsymbol{A}_m, \boldsymbol{P}_m(t))$$
$$= \sum_{i_1=1}^{4}\sum_{i_2=1}^{4}\ldots\sum_{i_m=1}^{4}\left(\left(\prod_{j=1}^{m}\prod_{k=1}^{4} P_j^{(i_k)}(t)\right) z^{\frac{c_1\eta_1^{(i_1)} + c_2\eta_2^{(i_2)} + \cdots + c_m\eta_m^{(i_m)}}{\sum_{i=1}^{m} c_i}\sum_{j=1}^{m}\sum_{k=1}^{4}\eta_j^{(i_k)}}\right) \tag{13}$$

The i^{th} state in PTS $s_{par}^{(i)} = \left(p_{par}^{(i)}(t), \eta_{par}^{(i)}\right)$ is $s_{par}^{(i)} \in (s_1^{i_1}, s_2^{i_2}, \cdots s_m^{i_m})$ and its PTE can be obtained by:

$$\begin{aligned} W_1^{(\mathrm{in})} &= W_{par}^{(\mathrm{in})} \cdot \frac{c_1}{\sum_{i=1}^{m} c_i}, W_1^{(\mathrm{out})} = \eta_1^{(\mathrm{i}_1)} \cdot W_1^{(\mathrm{in})} \\ W_2^{(\mathrm{in})} &= W_{par}^{(\mathrm{in})} \cdot \frac{c_2}{\sum_{i=1}^{m} c_i}, W_2^{(\mathrm{out})} = \eta_2^{(\mathrm{i}_2)} \cdot W_2^{(\mathrm{in})} \\ &\cdots \\ W_m^{(\mathrm{in})} &= W_{par}^{(\mathrm{in})} \cdot \frac{c_m}{\sum_{i=1}^{m} c_m}, W_m^{(\mathrm{out})} = \eta_m^{(\mathrm{i}_m)} \cdot W_m^{(\mathrm{in})} \end{aligned} \tag{14}$$

Then,

$$W_{par}^{(\mathrm{out})} = W_1^{(\mathrm{out})} + W_2^{(\mathrm{out})} + \cdots W_m^{(\mathrm{out})} = W_{par}^{(\mathrm{in})}\left(\frac{c_1}{\sum_{i=1}^{m} c_i}\eta_1^{(i_1)} + \frac{c_2}{\sum_{i=1}^{m} c_i}\eta_2^{(i_2)} + \cdots + \frac{c_m}{\sum_{i=1}^{m} c_i}\eta_m^{(i_m)}\right).$$

The PTE of the i^{th} state for the parallel structure, as is shown in Fig. 7, is $\eta_{par}^{(\mathrm{i})} = \frac{W_{par}^{(out)}}{W_{par}^{(\mathrm{in})}} = \frac{1}{\sum_{i=1}^{m} c_i}\left(\mathrm{c}_1\eta_1^{(\mathrm{i}_1)} + \mathrm{c}_2\eta_2^{(\mathrm{i}_2)} + \cdots + \mathrm{c}_m\eta_m^{(\mathrm{i}_m)}\right)$ and the corresponding probability can be calculated as $P_{par}^{(i)}(t) = P_1^{(i_1)}(t) \cdot P_2^{(i_2)}(t) \cdots P_m^{(i_m)}(t)$.

Example 4: Two PTUs are connected by parallel structure. The operator $\oplus$ is defined to describe the parallel operation. The PTS' i^{th} state probability and its corresponding PTE can be obtained by:

$$\begin{aligned}\mathrm{PTS}(\boldsymbol{\eta}_{par},\boldsymbol{P}_{par}(t)) &= \mathrm{PTU}(\eta_1,\boldsymbol{A}_1,\boldsymbol{P}_1(t)) \oplus \mathrm{PTU}(\eta_2,\boldsymbol{A}_2,\boldsymbol{P}_2(t)) \\
&= P_1^{(0)}(t)P_2^{(0)}(t)z^{\frac{c_1\eta_1^{(0)}+c_2\eta_2^{(0)}}{c_1+c_2}} + P_1^{(0)}(t)P_2^{(1)}(t)z^{\frac{c_1\eta_1^{(0)}+c_2\eta_2^{(1)}}{c_1+c_2}} + P_1^{(0)}(t)P_2^{(2)}(t)z^{\frac{c_1\eta_1^{(0)}+c_2\eta_2^{(2)}}{c_1+c_2}} + P_1^{(0)}(t)P_2^{(3)}(t)z^{\frac{c_1\eta_1^{(0)}+c_2\eta_2^{(3)}}{c_1+c_2}} \\
&+ P_1^{(1)}(t)P_2^{(0)}(t)z^{\frac{c_1\eta_1^{(1)}+c_2\eta_2^{(0)}}{c_1+c_2}} + P_1^{(1)}(t)P_2^{(1)}(t)z^{\frac{c_1\eta_1^{(1)}+c_2\eta_2^{(1)}}{c_1+c_2}} + P_1^{(1)}(t)P_2^{(2)}(t)z^{\frac{c_1\eta_1^{(1)}+c_2\eta_2^{(2)}}{c_1+c_2}} + P_1^{(1)}(t)P_2^{(3)}(t)z^{\frac{c_1\eta_1^{(1)}+c_2\eta_2^{(3)}}{c_1+c_2}} \\
&+ P_1^{(2)}(t)P_2^{(0)}(t)z^{\frac{c_1\eta_1^{(2)}+c_2\eta_2^{(0)}}{c_1+c_2}} + P_1^{(2)}(t)P_2^{(1)}(t)z^{\frac{c_1\eta_1^{(2)}+c_2\eta_2^{(1)}}{c_1+c_2}} + P_1^{(2)}(t)P_2^{(2)}(t)z^{\frac{c_1\eta_1^{(2)}+c_2\eta_2^{(2)}}{c_1+c_2}} + P_1^{(2)}(t)P_2^{(3)}(t)z^{\frac{c_1\eta_1^{(2)}+c_2\eta_2^{(3)}}{c_1+c_2}} \\
&+ P_1^{(3)}(t)P_2^{(0)}(t)z^{\frac{c_1\eta_1^{(3)}+c_2\eta_2^{(0)}}{c_1+c_2}} + P_1^{(3)}(t)P_2^{(1)}(t)z^{\frac{c_1\eta_1^{(3)}+c_2\eta_2^{(1)}}{c_1+c_2}} + P_1^{(3)}(t)P_2^{(2)}(t)z^{\frac{c_1\eta_1^{(3)}+c_2\eta_2^{(2)}}{c_1+c_2}} + P_1^{(3)}(t)P_2^{(3)}(t)z^{\frac{c_1\eta_1^{(3)}+c_2\eta_2^{(3)}}{c_1+c_2}}\end{aligned} \tag{15}$$

The effective implementation of the function for the system depends on whether the input energy can meet the energy requirements of the drive system. Through the expression of the PTE-based model mentioned above, the reliability that the system can meet the energy demand $W^{(\mathrm{d})}$ can be introduced as:

$$R^{\mathrm{P}}\left(t \mid \boldsymbol{W}_{\mathrm{sys}}^{(\mathrm{out})} \geq W^{(\mathrm{d})}\right)=\sum_{i=1}^{m}\left\{p_{\mathrm{sys}}^{(i)}(t)\cdot 1\left(\eta_{\mathrm{sys}}^{(i)}\cdot W_{\mathrm{sys}}^{(\mathrm{in})}-W^{(\mathrm{d})}\geq 0\right)\right\} \tag{16}$$

where $1\left(\eta_{\mathrm{sys}}^{(i)}\cdot W_{\mathrm{sys}}^{(\mathrm{in})}-W^{(\mathrm{d})}\geq 0\right)=\begin{cases}1,\eta_{\mathrm{sys}}^{(i)}\cdot W_{\mathrm{sys}}^{(\mathrm{in})}-W^{(\mathrm{d})}\geq 0\\ 0,\eta_{\mathrm{sys}}^{(i)}\cdot W_{\mathrm{sys}}^{(\mathrm{in})}-W^{(\mathrm{d})}<0\end{cases}$.

Fig. 8 shows the flow chat of the reliability assessment method based on PTE. Firstly, the degradation analysis based on PTE model with multiple states is carried out and the corresponding probability of the states can be calculated by GSPN. Then, UGF is used to calculate the probability of PTU states and the system with u-function. Finally, the reliability assessment can be obtained based on the power threshold boundary.

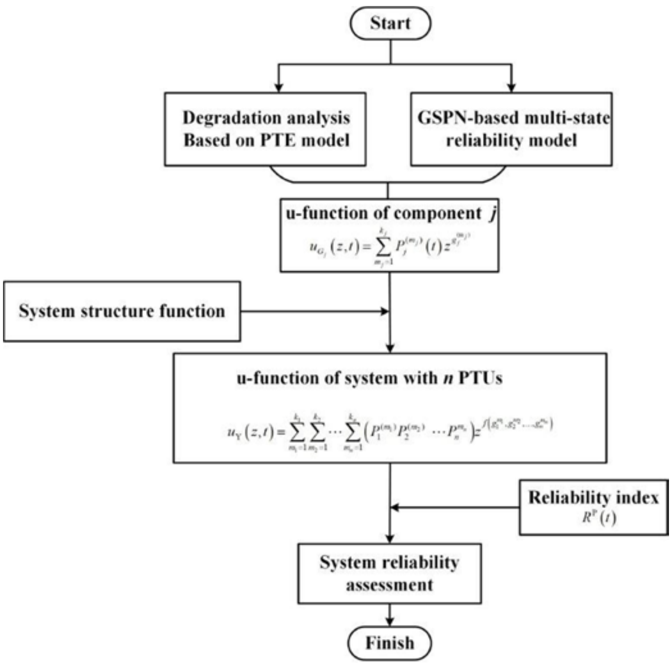


Fig. 8. Flow chart of the reliability assessment of PTS based on PTE

3. Case study of dual hydraulic actuation system

3.1. System description

Dual hydraulic actuation system (DHAS) is a typical PTS which drives the control surfaces of civil aircraft, shown in Fig. 9, in which the hydraulic pump transfers the energy from engine through accessory gearbox, servo valve controls the flow of cylinder according to command and the cylinder drives the control surface with high pressure oil. Components in DHAS can be regarded as PTUs with power dissipation. In order to keep high reliability, DHAS adopts dual redundant hydraulic power supply systems and actuations.

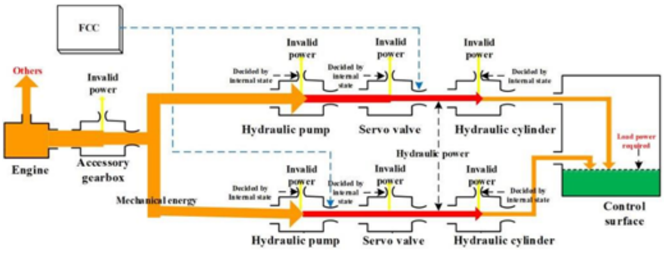


Fig. 9. Structure diagram and power transmission of DHAS

Hydraulic pump transfers a portion of aircraft engine power to high pressure hydraulic power supply, which drives the control surface according to the DHAS control command. Whether the DHAS can provide enough power to drive the control surface with very quick response depends on components' performance degradation and its external destruction. According to Section 2.3, a PTE-based model for DHAS can be built as is shown in Fig. 10.

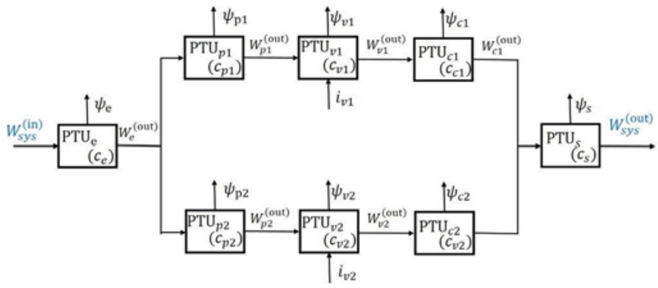


Fig. 10. PTE -based model of each PTU in DHAS

In Fig. 10, the output power of aircraft engine $W_{\mathrm{sys}}^{(\mathrm{in})}$ is firstly transferred into accessory gearbox $\mathrm{PTU}(\eta_{\mathrm{G}},\boldsymbol{P}_{\mathrm{G}}(t))$, then the output power of accessory gearbox $W_{\mathrm{G}}^{(\mathrm{out})}$ is inputted to DHAS channels. In DHAS channel, the primary hydraulic pump $\mathrm{PTU}(\eta_{\mathrm{p}_1},\boldsymbol{A}_{\mathrm{p}_1},\boldsymbol{P}_{\mathrm{p}_1}(t))$ and the secondary hydraulic pump $\mathrm{PTU}(\eta_{\mathrm{P}_2},\boldsymbol{A}_{\mathrm{P}_2},\boldsymbol{P}_{\mathrm{P}_2}(t))$ receive the power and transfer the power into pressurized hydraulic oil. The servo valves $\mathrm{PTU}(\eta_{\mathrm{V}_1},\boldsymbol{A}_{\mathrm{V}_1},\boldsymbol{P}_{\mathrm{V}_1}(t))$ and $\mathrm{PTU}(\eta_{\mathrm{V}_2},\boldsymbol{A}_{\mathrm{V}_2},\boldsymbol{P}_{\mathrm{V}_2}(t))$ receive control signals from flight control computers and drive the corresponding hydraulic cylinder, $\mathrm{PTU}(\eta_{\mathrm{C}_1},\boldsymbol{A}_{\mathrm{C}_1},\boldsymbol{P}_{\mathrm{C}_1}(t))$ or $\mathrm{PTU}(\eta_{\mathrm{C}_2},\boldsymbol{A}_{\mathrm{C}_2},\boldsymbol{P}_{\mathrm{C}_2}(t))$. The dual hydraulic cylinders integrate to drive the control surface $(\eta_{\mathrm{sur}},\boldsymbol{P}_{\mathrm{sur}}(t))$. The output power $W_{\mathrm{sys}}^{(\mathrm{out})}$ works as the speed and output force on control surfaces. Supposing that the $W_{\mathrm{sur}}^{(\mathrm{d})}$ is the minimum required power of control surface under normal performance, the reliability of DHAS can be described as:

$$R_{\mathrm{DHAS}}^{\mathrm{p}}(t) = P(t \mid W_{\mathrm{sys}}^{(\mathrm{out})} \geq W_{\mathrm{sur}}^{(\mathrm{d})}) \tag{17}$$

For each channel of the DHAS, according to Eq. (10), the PTE-based model can be written as:

$$\begin{aligned}
&\mathrm{DHAS}_1(\eta_{\mathrm{DHAS}_1}, \boldsymbol{P}_{\mathrm{DHAS}_1}(t)) = \mathrm{PTU}(\eta_{\mathrm{P}_1}, A_{\mathrm{P}_1}, \boldsymbol{P}_{\mathrm{P}_1}(t)) \otimes \mathrm{PTU}(\eta_{\mathrm{V}_1}, A_{\mathrm{V}_1}, \boldsymbol{P}_{\mathrm{V}_1}(t)) \otimes \mathrm{PTU}(\eta_{\mathrm{C}_1}, A_{\mathrm{C}_1}, \boldsymbol{P}_{\mathrm{C}_1}(t)) \\
&\mathrm{DHAS}_2(\eta_{\mathrm{DHAS}_2}, \boldsymbol{P}_{\mathrm{DHAS}_2}(t)) = \mathrm{PTU}(\eta_{\mathrm{P}_2}, A_{\mathrm{P}_2}, \boldsymbol{P}_{\mathrm{P}_2}(t)) \otimes \mathrm{PTU}(\eta_{\mathrm{V}_2}, A_{\mathrm{V}_2}, \boldsymbol{P}_{\mathrm{V}_2}(t)) \otimes \mathrm{PTU}(\eta_{\mathrm{C}_2}, A_{\mathrm{C}_2}, \boldsymbol{P}_{\mathrm{C}_2}(t))
\end{aligned} \tag{18}$$

Then, the PTE-based model for the whole DHAS can be written as:

$$\begin{aligned}
&\mathrm{DHAS}(\eta_{\mathrm{DHAS}}, \boldsymbol{P}_{\mathrm{DHAS}}(t)) = \mathrm{PTU}(\eta_{\mathrm{G}}, \boldsymbol{P}_{\mathrm{G}}(t)) \otimes \\
&\qquad \left(\mathrm{DHAS}_1(\eta_{\mathrm{DHAS}_1}, \boldsymbol{P}_{\mathrm{DHAS}_1}(t)) \oplus \mathrm{DHAS}_2(\eta_{\mathrm{DHAS}_2}, \boldsymbol{P}_{\mathrm{DHAS}_2}(t))\right) \otimes (\eta_{\mathrm{sur}}, \boldsymbol{P}_{\mathrm{sur}}(t))
\end{aligned} \tag{19}$$

Here the DHAS of Airbus380 is selected as a case study in real flight profile. The worst scenario is considered, whereby a failure of the control surface will result in the loss of the aircraft, and the demand is that such event must have a probability of less than 1×10^{-9} per flight hour. The failure rates assumed for each type of component are listed in Table 1 [2].

Table 1. Failure rates for each component

Component	Mechanical	Controller	Power supply
Failure rate λ	2.2×10^{-6} / h	8.6×10^{-5} / h	5.4×10^{-5} / h

3.2. PTE-based reliability model of individual PTU in DHAS

In Fig. 10, the main components in DHAS include hydraulic pump, servo valves and hydraulic cylinders. The PTE-based reliability model of above PTU is built as follows.

3.2.1. Hydraulic pump

The structure of an aircraft axial piston pump is shown in Fig. 11. A conventional piston pump provides the pressurized oil through the pistons reciprocating within their bores, in which the displaced volume from the pump is controlled by the inclination angle of the hanger. The failure behavior of the pump is a progressive failure process caused by wear [2].

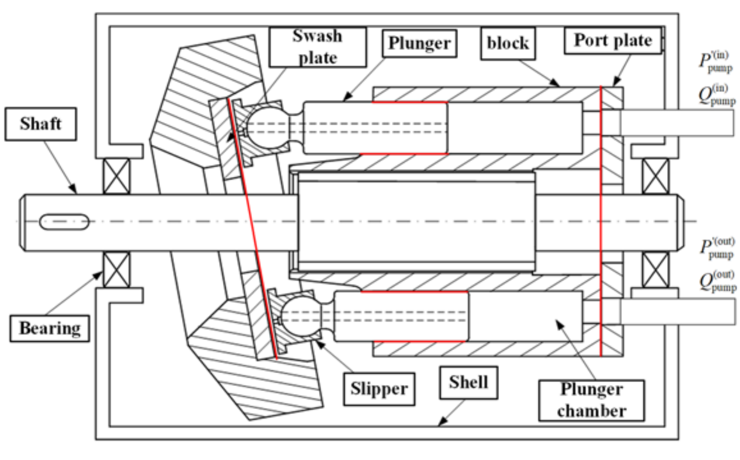


Fig. 11. Operating principle and components of a piston hydraulic pump

The power output of hydraulic pump, $W_{\mathrm{P}}^{(\mathrm{out})}$, is influenced by two different random effects:

1) the input power comes from the engine accessory gearbox $W_{\mathrm{G}}^{(\mathrm{in})}$.
2) the efficiency of hydraulic pump, $\eta_{\mathrm{P}} = \eta_{\mathrm{P}}^{(\mathrm{M})}\eta_{\mathrm{P}}^{(\mathrm{V})}$ wherein, $\eta_{\mathrm{P}}^{(\mathrm{M})}$ denotes the mechanical efficiency and $\eta_{\mathrm{P}}^{(\mathrm{V})}$ denotes the volumetric efficiency. Therefore:

$$W_{\mathrm{P}}^{(\mathrm{out})} = \eta_{\mathrm{P}}^{(\mathrm{M})}\eta_{\mathrm{P}}^{(\mathrm{V})} \cdot W_{\mathrm{P}}^{(\mathrm{in})} \tag{20}$$

For the volumetric efficiency of the pump:

$$\eta_{\mathrm{P}}^{(\mathrm{V})} = 1 - \frac{\Delta Q_{\mathrm{P}}}{n_{\mathrm{P}} V} \tag{21}$$

wherein, ΔQ_{P} is the internal leakage flow of pump and V is theoretical displacement of pump.

The wear of the friction pairs in a hydraulic pump, e.g. , the wear of the plunger–plunger cavity on rotor, will increase the clearance of the friction pairs which enlarges the oil leakage ΔQ_{P} . $\eta_{\mathrm{P}}^{(\mathrm{V})}$ ill then decrease accordingly, as shown in Fig. 4.

The mechanical efficiency, $\eta_{\mathrm{P}}^{(\mathrm{M})}$, is determined by the conditions of mechanical-linkage system between the hydraulic pump and driving devices. This paper ignores the mechanical efficiency and only discusses the impacts of the volumetric efficiency.

According to the value of $\eta_{\mathrm{P}}^{(\mathrm{V})}$ the states of pump are divided into four discrete cases, i.e.:

$$\mathbf{s}_{\mathrm{P}} = \begin{cases} s_{\mathrm{P}0} & \eta_{\mathrm{P}}^{(0)} \in [0.92, 1] \\ s_{\mathrm{P}1} & \eta_{\mathrm{P}}^{(1)} \in [0.9, 0.92) \\ s_{\mathrm{P}2} & \eta_{\mathrm{P}}^{(2)} \in [0.75, 0.9) \\ s_{\mathrm{P}3} & \eta_{\mathrm{P}}^{(3)} \in [0, 0.75) \end{cases} \tag{22}$$

When $\eta_{\mathrm{P}} \geq 0.92$, the pump operates very well, and $s_{\mathrm{P}0}$ is named as the normal state of the hydraulic pump. With the development of internal wear, the value of η_{P} decreases into the interval of [0.9,0.92). In this case, the pump operates well despite the small amount of wear, so $s_{\mathrm{P}1}$ is named as the light degradation state. The value of η_{P} decreases into the interval of [0.75,0.9) with continuous accumulation of wear, the performance of hydraulic pump degrades obviously, thus $s_{\mathrm{P}2}$ is named as serious degradation state. Once $\eta_{\mathrm{P}} < 0.75$, the pump will be regarded as failed. Median value of each interval is used as the indicator and we get $\boldsymbol{\eta}_{\mathrm{P}} = [\eta_{\mathrm{P}}^{(0)}, \eta_{\mathrm{P}}^{(1)}, \eta_{\mathrm{P}}^{(2)}, \eta_{\mathrm{P}}^{(3)}]^{\mathrm{T}} = [0.96, 0.91, 0.825, 0.375]^{\mathrm{T}}$.

In order to get the probabilities corresponding to $\mathbf{s}_{\mathrm{P}} = [s_{\mathrm{P}0}, s_{\mathrm{P}1}, s_{\mathrm{P}2}, s_{\mathrm{P}3}]$, a GSPN-based state transition model is established as shown in Fig. 12.

According to the GSPN model, at time 0, pump starts from the normally functioning state, #(PUMP.up)=1. After a period of time $t_{\mathrm{P}1}$, pump may directly fail with the probability $P(t_{\mathrm{P}i3})$, which is described by a token being transmitted directly from PUMP.up to PUMP.dn. Alternatively, the pump can degrades through a gradual failure mode with probability $P(t_{\mathrm{P}i1})$, and

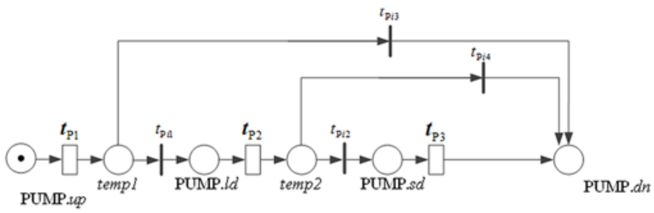


Fig. 12. The GSPN model of hydraulic pump performance degradation process

then pump enters the light degradation state, i.e. #(PUMP.ld)=1. Here, $P(t_{Pi3})+P(t_{Pi1})=1$. The transfer rate of the timed transition t_{P1} is λ_{P1}, so that $P(t_{Pi3})\lambda_{P1}$ describes the failure rate of pump from normal operating state to complete failure, and $P(t_{Pi1})\lambda_{P1}$ describes the failure rate of pump from normal operating state to light degradation state. As operational time increasing, pump may suffer additional performance degradation leading directly to complete failure, i.e., entering the down state (PUMP.dn), or firstly reaching to the serious degradation state (PUMP.sd) and the token is correspondingly transferred with probability of $P(t_{Pi4})$. Here, $P(t_{Pi4})+P(t_{Pi2})=1$ and $P(t_{Pi1})+P(t_{Pi3})=1$. The $tempi(i=1,2)$ represents temporary states in the GSPN modeling of the hydraulic pump performance degradation process. Given the initial work condition, the steady state probabilities of the pump states can be obtained by simulation. This part is adapted from the original work proposed by authors previously [29].

GSPN-based model of hydraulic pump only describes its degradation process. To obtain $\mathbf{P}_P(t)$, the Markov process model which equivalent to GSPN model is illustrated in Fig. 13.

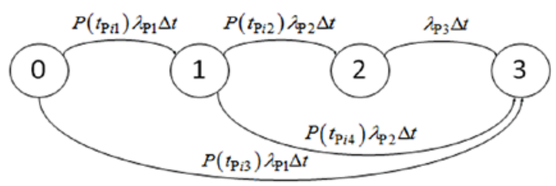


Fig. 13. The mechanical degradation states of hydraulic pump

The corresponding relationship between states of GSPN model and Markov model is illustrated in Table 2.

Table 2. The corresponding relationship between states of GSPN model and Markov model

No.	States in Markov Model	States in GSPN model	Description
1	'0'	#(PUMP.up)=1	The normally state of pump
2	'1'	#(PUMP.ld)=1	The light degradation state of pump
3	'2'	#(PUMP.ld)=1	The serious degradation state of pump
4	'3'	#(PUMP.dn)=1	The failure state of pump

The state transition matrix of hydraulic pump, $\boldsymbol{A}_P$, can be written as:

$$\boldsymbol{A}_P=\begin{bmatrix} -\left(P(t_{Pi1})\lambda_{P1}+P(t_{Pi3})\lambda_{P1}\right) & 0 & 0 & 0 \\ P(t_{Pi1})\lambda_{P1} & -\left(P(t_{Pi2})\lambda_{P2}+P(t_{Pi4})\lambda_{P2}\right) & 0 & 0 \\ 0 & P(t_{Pi2})\lambda_{P2} & -\lambda_{P3} & 0 \\ P(t_{Pi3})\lambda_{P1} & P(t_{Pi4})\lambda_{P2} & \lambda_{P3} & 0 \end{bmatrix}=\begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & 0 \end{bmatrix} \tag{23}$$

The probabilities that the hydraulic pump stays in $\boldsymbol{s}_P=[s_{P0},s_{P1},s_{P2},s_{P3}]$ can be obtained according to Eq. (7) and the derivation is shown below:

$$\begin{aligned} P_P^{(0)}(t)&=e^{a_{11}t} \\ P_P^{(1)}(t)&=\frac{a_{21}}{a_{11}-a_{22}}(e^{a_{11}t}-e^{a_{22}t}) \\ P_P^{(2)}(t)&=\frac{a_{21}a_{32}}{(a_{11}-a_{22})(a_{11}-a_{33})}e^{a_{11}t}+\frac{a_{21}a_{32}}{(a_{22}-a_{11})(a_{22}-a_{33})}e^{a_{22}t}+\frac{a_{21}a_{32}}{(a_{33}-a_{11})(a_{33}-a_{22})}e^{a_{33}t} \\ P_P^{(3)}(t)&=\left[-\frac{a_{41}}{a_{11}}+\frac{a_{21}a_{42}}{a_{11}a_{22}}-\frac{a_{21}a_{32}a_{43}}{a_{11}a_{22}a_{33}}\right]+\left[\frac{a_{41}}{a_{11}}+\frac{a_{21}a_{42}}{a_{11}(a_{11}-a_{22})}+\frac{a_{21}a_{32}a_{43}}{a_{11}(a_{11}-a_{22})(a_{11}-a_{33})}\right]e^{a_{11}t} \\ &+\left[\frac{a_{21}a_{42}}{a_{22}(a_{22}-a_{11})}+\frac{a_{21}a_{32}a_{43}}{a_{22}(a_{22}-a_{11})(a_{22}-a_{33})}\right]e^{a_{22}t}+\frac{a_{21}a_{32}a_{43}}{a_{33}(a_{33}-a_{11})(a_{33}-a_{22})}e^{a_{33}t} \end{aligned} \tag{24}$$

The parameters of the hydraulic pump model of Fig. 13 are listed in Table 3 [29].

Table 3. Parameters of Pump model

Parameter	Value	Parameter	Value
$P(t_{Pi1})$	0.9	$P(t_{Pi2})$	0.3
$P(t_{Pi3})$	0.1	$P(t_{Pi4})$	0.7
λ_{P1}	$5.4\times10^{-5}/h$	λ_{P2}	$6.0\times10^{-5}/h$
λ_{P3}	$6.8\times10^{-5}/h$		

Then, the probabilities of the pump states can be obtained:

$$\boldsymbol{P}_P(t)=\begin{bmatrix} P_P^{(0)}(t) \\ P_P^{(1)}(t) \\ P_P^{(2)}(t) \\ P_P^{(3)}(t) \end{bmatrix}=\begin{bmatrix} e^{-5.4\times10^{-5}t} \\ 8.1\times(e^{-5.4\times10^{-5}t}-e^{-6.0\times10^{-5}t}) \\ 10.41e^{-5.4\times10^{-5}t}-18.23e^{-6.0\times10^{-5}t}+7.82e^{-6.8\times10^{-5}t} \\ 1-19.51e^{-5.4\times10^{-5}t}+26.33e^{-6.0\times10^{-5}t}-7.82e^{-6.8\times10^{-5}t} \end{bmatrix} \tag{25}$$

The probability curves of the pump states are illustrated in Fig. 14.

Therefore, for a hydraulic pump, the stochastic power transmission model, $\mathrm{PTU}(\boldsymbol{\eta}_P,\boldsymbol{A}_P,\boldsymbol{P}_P(t))$, can be written as:

$$\begin{aligned} \mathrm{PTU}(\boldsymbol{\eta}_P,\boldsymbol{A}_P,\boldsymbol{P}_P(t))&=P_P^{(0)}(t)z^{\eta_P^{(0)}}+P_P^{(1)}(t)z^{\eta_P^{(1)}}+P_P^{(2)}(t)z^{\eta_P^{(2)}}+P_P^{(3)}(t)z^{\eta_P^{(3)}} \\ &=e^{-5.4\times10^{-5}t}z^{0.96}+8.1\times(e^{-5.4\times10^{-5}t}-e^{-6.0\times10^{-5}t})z^{0.91} \\ &+\left[10.41e^{-5.4\times10^{-5}t}-18.23e^{-6.0\times10^{-5}t}+7.82e^{-6.8\times10^{-5}t}\right]z^{0.825} \\ &+\left[1-19.51e^{-5.4\times10^{-5}t}+26.33e^{-6.0\times10^{-5}t}-7.82e^{-6.8\times10^{-5}t}\right]z^{0.375} \end{aligned} \tag{26}$$

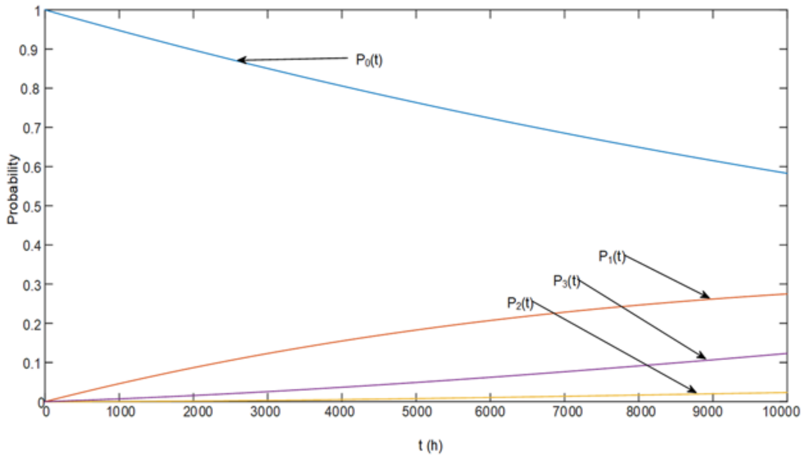


Fig. 14. The probability curves of the pump states

3.2.2. Servo valve

The servo valve is the control component in DHAS, which receives the command from the actuator control electronics (ACE) through the aviation bus. The power output of the servo valve is influenced by the command signal i_c from the ACE (externally) and mechanical degradation (internally). The schematic representation of the servo valve shown in Fig. 15 indicates the information flow and hydraulic power which together determine the power output of the servo valve.

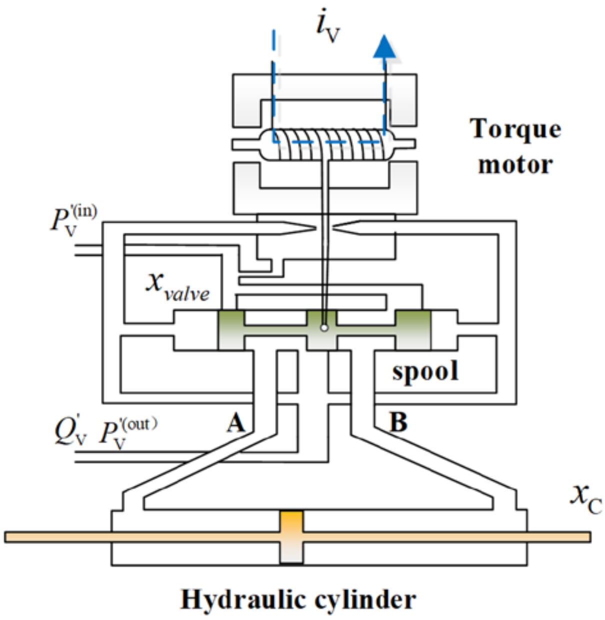


Fig.15. Control current and power output of the servo valve

The servo valve consists of four main parts connected in series: electrical-mechanical conversion, mechanical-hydraulic conversion, hydraulic amplifier and feedback device. The failure of a single sub-assembly will lead to the failure of the entire servo valve. Each part's degradation/failure/repair behavior contributes to the mechanical states of the servo valve.

The PTE of servo valve can be formulated as:

$$\begin{cases} \eta_V = \dfrac{W_V^{(out)}}{W_V^{(in)}} \\ W_V^{(out)} = P_L \cdot Q_L = P_L \cdot C_q \cdot C_W \cdot K_v \cdot i_V \cdot \sqrt{\dfrac{1}{\rho}\left(P_{sys} - P_L\right)} \\ W_V^{(in)} = W_P^{(out)} = P_{sys} \cdot Q_{sys} \end{cases} \tag{27}$$

in which, P_L and Q_L are the load pressure and load flow rate, respectively, which are directly related to the actual load. P_{sys} and Q_{sys} are the output pressure and output flow rate of the hydraulic pump. C_q, C_W, K_v are discharge coefficient of a servo valve, a gradient of valve orifice and amplification coefficient of servo valve respectively, ρ is density of hydraulic oil, which can be assumed to be a constant during operation, i_V is the controlled signal by actuator control electronics.

Given load pressure P_L, the PTE of servo valve η_V is determined by both control current states and mechanical states of the servo valve, i.e. $\eta_V = \eta_V^{(i_c)} \cdot \eta_V^{(M)}$, where $\eta_V^{(i_V)}$ indicates the control efficiency and $\eta_V^{(M)}$ indicates the mechanical efficiency.

The attenuation is inevitable in the process of control current of i_V transmission, which influences the output flow rate of servo valve. To simplify analysis, $\boldsymbol{\eta}_V^{(i_V)} = [\eta_V^{i_V(0)}, \eta_V^{i_V(1)}, \eta_V^{i_V(2)}, \eta_V^{i_V(3)}]^T = [1.0, 0.8, 0.3, 0]^T$ is used to characterize the actual effects of different attenuation rates of control current acting on the efficiency of servo valve, and $\boldsymbol{P}_V^{(i_V)} = [P_V^{i_V(0)}, P_V^{i_V(1)}, P_V^{i_V(2)}, P_V^{i_V(3)}]^T = [0.864, 0.118, 0.016, 0.002]^T$ is given to denote the corresponding probabilities that control current has different attenuation rates. Therefore, the sub-PTE model of servo valve which is linked with control current part, $\mathrm{PTU}(\boldsymbol{\eta}_V^{(i_V)}, \boldsymbol{P}_V^{(i_V)})$, can be written as:

$$\begin{aligned} \mathrm{PTU}(\boldsymbol{\eta}_V^{(i_V)}, \boldsymbol{P}_V^{(i_V)}) &= P_V^{i_V(0)} z^{\eta_V^{i_V(0)}} + P_V^{i_V(1)} z^{\eta_V^{i_V(1)}} + P_V^{i_V(2)} z^{\eta_V^{i_V(2)}} + P_V^{i_V(3)} z^{\eta_V^{i_V(3)}} \\ &= 0.864 z^1 + 0.118 z^{0.8} + 0.016 z^{0.3} + 0.002 z^0 \end{aligned} \tag{28}$$

Similar to hydraulic pump, wear between spool and sleeves in servo valve will reduce the output flow rate, and then reduce the mechanical efficiency $\eta_V^{(M)}$. According to the actual value of $\eta_V^{(M)}$ the mechanical states of the servo valve are divided into four discrete cases, i.e.:

$$\boldsymbol{s}_V^{(M)} = \begin{cases} s_{V0} & \eta_V^{M(0)} \in [0.5, 0.667] \\ s_{V1} & \eta_V^{M(1)} \in [0.35, 0.5) \\ s_{V2} & \eta_V^{M(2)} \in [0.2, 0.35) \\ s_{V3} & \eta_V^{M(3)} \in [0, 0.2) \end{cases} \tag{29}$$

For a servo valve, the maximum mechanical efficiency can be 0.667 in the engineering field. When $\eta_V^{(M)} \in [0.5, 0.667]$, the servo valve operates very well, and s_{V0} is named as the normal state. With the increase of internal wear, the value of $\eta_V^{(M)}$ decreases into the interval of [0.35,0.5). In this case, the servo valve operates well despite the small amount of wear, so s_{V1} is named as the light degradation state. The value of $\eta_V^{(M)}$ enters into the interval of [0.2,0.35) with

$$
\begin{aligned}
\mathrm{PTU}(\boldsymbol{\eta}_\mathrm{V},\boldsymbol{A}_\mathrm{V},\boldsymbol{P}_\mathrm{V}(t)) &= \mathrm{PTU}_\mathrm{V}^{i_\mathrm{V}}(\boldsymbol{\eta}_\mathrm{V}^{(i_\mathrm{V})},\boldsymbol{P}_\mathrm{V}^{(i_\mathrm{V})}) \otimes \mathrm{PTU}(\boldsymbol{\eta}_\mathrm{V}^{(\mathrm{M})},\boldsymbol{A}_\mathrm{V},\boldsymbol{P}_\mathrm{V}^{(\mathrm{M})}(t)) \\
&= P_\mathrm{V}^{i_\mathrm{V}(0)}P_\mathrm{V}^{\mathrm{M}(0)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(0)}\eta_\mathrm{V}^{\mathrm{M}(0)}} + P_\mathrm{V}^{i_\mathrm{V}(0)}P_\mathrm{V}^{\mathrm{M}(1)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(0)}\eta_\mathrm{V}^{\mathrm{M}(1)}} + P_\mathrm{V}^{i_\mathrm{V}(0)}P_\mathrm{V}^{\mathrm{M}(2)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(0)}\eta_\mathrm{V}^{\mathrm{M}(2)}} + P_\mathrm{V}^{i_\mathrm{V}(0)}P_\mathrm{V}^{\mathrm{M}(3)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(0)}\eta_\mathrm{V}^{\mathrm{M}(3)}} \\
&+ P_\mathrm{V}^{i_\mathrm{V}(1)}P_\mathrm{V}^{\mathrm{M}(0)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(1)}\eta_\mathrm{V}^{\mathrm{M}(0)}} + P_\mathrm{V}^{i_\mathrm{V}(1)}P_\mathrm{V}^{\mathrm{M}(1)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(1)}\eta_\mathrm{V}^{\mathrm{M}(1)}} + P_\mathrm{V}^{i_\mathrm{V}(1)}P_\mathrm{V}^{\mathrm{M}(2)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(1)}\eta_\mathrm{V}^{\mathrm{M}(2)}} + P_\mathrm{V}^{i_\mathrm{V}(1)}P_\mathrm{V}^{\mathrm{M}(3)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(1)}\eta_\mathrm{V}^{\mathrm{M}(3)}} \\
&+ P_\mathrm{V}^{i_\mathrm{V}(2)}P_\mathrm{V}^{\mathrm{M}(0)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(2)}\eta_\mathrm{V}^{\mathrm{M}(0)}} + P_\mathrm{V}^{i_\mathrm{V}(2)}P_\mathrm{V}^{\mathrm{M}(1)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(2)}\eta_\mathrm{V}^{\mathrm{M}(1)}} + P_\mathrm{V}^{i_\mathrm{V}(2)}P_\mathrm{V}^{\mathrm{M}(2)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(2)}\eta_\mathrm{V}^{\mathrm{M}(2)}} + P_\mathrm{V}^{i_\mathrm{V}(2)}P_\mathrm{V}^{\mathrm{M}(3)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(2)}\eta_\mathrm{V}^{\mathrm{M}(3)}} \\
&+ P_\mathrm{V}^{i_\mathrm{V}(3)}P_\mathrm{V}^{\mathrm{M}(0)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(3)}\eta_\mathrm{V}^{\mathrm{M}(0)}} + P_\mathrm{V}^{i_\mathrm{V}(3)}P_\mathrm{V}^{\mathrm{M}(1)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(3)}\eta_\mathrm{V}^{\mathrm{M}(1)}} + P_\mathrm{V}^{i_\mathrm{V}(3)}P_\mathrm{V}^{\mathrm{M}(2)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(3)}\eta_\mathrm{V}^{\mathrm{M}(2)}} + P_\mathrm{V}^{i_\mathrm{V}(3)}P_\mathrm{V}^{\mathrm{M}(3)}(t)z^{\eta_\mathrm{V}^{i_\mathrm{V}(3)}\eta_\mathrm{V}^{\mathrm{M}(3)}}
\end{aligned} \tag{33}
$$

continuous accumulation of inner wear, the performance of servo valve degrades obviously, thus s_{V2} is named as serious degradation state. Once $\eta_\mathrm{V}^{(\mathrm{M})} < 0.2$, the servo valve will be regarded as failed. Median value of each interval is utilized as the indicator of $\eta_\mathrm{V}^{(\mathrm{M})}$, and we get $\boldsymbol{\eta}_\mathrm{V}^{(\mathrm{M})}=\left[\eta_\mathrm{V}^{\mathrm{M}(0)}\ \ \eta_\mathrm{V}^{\mathrm{M}(1)}\ \ \eta_\mathrm{V}^{\mathrm{M}(2)}\ \ \eta_\mathrm{V}^{\mathrm{M}(3)}\right]^\mathrm{T}=\left[0.5835\ \ 0.425\ \ 0.275\ \ 0.1\right]^\mathrm{T}$.

Given the reliability parameters of servo valve as listed in Table 4, the probabilities that the servo valve stays in $\boldsymbol{s}_\mathrm{V}^{(\mathrm{M})}$ can be obtained with the similar GSPN model and Markov model of hydraulic pump shown in Fig. 13.

Table 4. Parameters of servo valve model

Parameter	Value	Parameter	Value
$P(t_{\mathrm{V}i1})$	0.9	$P(t_{\mathrm{V}i2})$	0.5
$P(t_{\mathrm{V}i3})$	0.1	$P(t_{\mathrm{V}i4})$	0.5
λ_{V1}	$8.6\times10^{-5}/h$	λ_{V2}	$9.0\times10^{-5}/h$
λ_{V3}	$9.8\times10^{-5}/h$		

According to GSPN model, the state transition matrix of servo valve, , can be written as:

$$
\boldsymbol{A}_\mathrm{V} = \begin{bmatrix} -8.6\times10^{-5} & 0 & 0 & 0 \\ 7.74\times10^{-5} & -9.0\times10^{-5} & 0 & 0 \\ 0 & 4.5\times10^{-5} & -9.8\times10^{-5} & 0 \\ 9\times10^{-6} & 4.5\times10^{-5} & 9.8\times10^{-5} & 0 \end{bmatrix} \tag{30}
$$

Then, the probabilities of mechanical states of the servo valve are obtained as:

Then, the probabilities $\boldsymbol{P}_\mathrm{V}^{(\mathrm{M})}(t)$ of mechanical states of the servo valve are obtained as:

$$
\boldsymbol{P}_\mathrm{V}^{(\mathrm{M})}(t) = \begin{bmatrix} P_\mathrm{V}^{\mathrm{M}(0)}(t) \\ P_\mathrm{V}^{\mathrm{M}(1)}(t) \\ P_\mathrm{V}^{\mathrm{M}(2)}(t) \\ P_\mathrm{V}^{\mathrm{M}(3)}(t) \end{bmatrix} = \begin{bmatrix} e^{-8.6\times10^{-5}t} \\ 19.35\times(e^{-8.6\times10^{-5}t}-e^{-9.0\times10^{-5}t}) \\ 72.56e^{-8.6\times10^{-5}t}-108.84e^{-9.0\times10^{-5}t}+36.28e^{-9.8\times10^{-5}t} \\ 1-92.91e^{-8.6\times10^{-5}t}+128.19e^{-9.0\times10^{-5}t}-36.28e^{-9.8\times10^{-5}t} \end{bmatrix} \tag{31}
$$

Therefore, the sub-PTE model of servo valve relating with mechanical states part, $\mathrm{PTU}(\boldsymbol{\eta}_\mathrm{V}^{(\mathrm{M})},\boldsymbol{A}_\mathrm{V},\boldsymbol{P}_\mathrm{V}^{(\mathrm{M})}(t))$, can be written as:

$$
\begin{aligned}
\mathrm{PTU}(\boldsymbol{\eta}_\mathrm{V}^{(\mathrm{M})},\boldsymbol{P}_\mathrm{V}^{(\mathrm{M})}(t)) &= p_\mathrm{V}^{\mathrm{M}(0)}z^{\eta_\mathrm{V}^{\mathrm{M}(0)}}+p_\mathrm{V}^{\mathrm{M}(1)}z^{\eta_\mathrm{V}^{\mathrm{M}(1)}}+p_\mathrm{V}^{\mathrm{M}(2)}z^{\eta_\mathrm{V}^{\mathrm{M}(2)}}+p_\mathrm{V}^{\mathrm{M}(3)}z^{\eta_\mathrm{V}^{\mathrm{M}(3)}} \\
&= e^{-8.6\times10^{-5}t}z^{0.5835}+19.35\times(e^{-8.6\times10^{-5}t}-e^{-9.0\times10^{-5}t})z^{0.425} \\
&+\left[72.56e^{-8.6\times10^{-5}t}-108.84e^{-9.0\times10^{-5}t}+36.28e^{-9.8\times10^{-5}t}\right]z^{0.275} \\
&+\left[1-92.91e^{-8.6\times10^{-5}t}+128.19e^{-9.0\times10^{-5}t}-36.28e^{-9.8\times10^{-5}t}\right]z^{0.1}
\end{aligned} \tag{32}
$$

The complete PTE model of servo valve can be written as Equation (33) [above].

3.2.3. Hydraulic cylinder

Fig. 16 shows the inner structure of a hydraulic cylinder in which the leakage due to the friction and wear is the main failure behavior. Since the wear and tear between the piston and cylinder are progressive, the multiple degradation states based on a GSPN model can used to establish the reliability model.

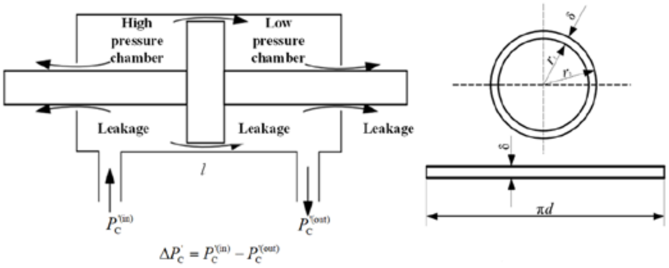


Fig. 16. Working principle of hydraulic cylinder

The output power of hydraulic cylinder is:

$$
W_\mathrm{C}^{(\mathrm{out})} = F_L\cdot V = \eta_\mathrm{C}\cdot P_L\cdot Q_L = \eta_\mathrm{C}\cdot W_\mathrm{V}^{(\mathrm{out})} \tag{34}
$$

where η_C is volumetric efficiency of hydraulic cylinder.

The value of η_C is determined by wear status of the hydraulic cylinder. According to the actual value of η_C, the states of the cylinder are divided into four discrete cases mentioned above, i.e.:

$$
\boldsymbol{s}_\mathrm{C} = \begin{cases} s_{\mathrm{C0}} & \eta_\mathrm{C}^{(0)}\in[0.95,1] \\ s_{\mathrm{C1}} & \eta_\mathrm{C}^{(1)}\in[0.85,0.95) \\ s_{\mathrm{C2}} & \eta_\mathrm{C}^{(2)}\in[0.75,0.85) \\ s_{\mathrm{C3}} & \eta_\mathrm{C}^{(3)}\in[0,0.75) \end{cases} \tag{35}
$$

Table 5. Parameters of the hydraulic cylinder model

Parameter	Value	Parameter	Value
$P(t_{Ci1})$	0.8	$P(t_{Ci2})$	0.6
$P(t_{Ci3})$	0.2	$P(t_{Ci4})$	0.4
λ_{C1}	2.2×10^{-6}/h	λ_{C2}	3.6×10^{-6} / h
λ_{C3}	6.7×10^{-6} / h		

Median value of each interval is used as the indicator of η_C, and we can get $\boldsymbol{\eta}_C = \left[\eta_C^{(0)} \quad \eta_C^{(1)} \quad \eta_C^{(2)} \quad \eta_C^{(3)}\right]^T = [0.975 \quad 0.9 \quad 0.8 \quad 0.375]^T$.

Given the reliability parameters of hydraulic cylinder as listed in Table 5, the probabilities that the hydraulic cylinder stays in $\mathbf{s}_C$ can be obtained with the similar GSPN model and Markov model same to hydraulic pump shown in Fig. 13.

According to GSPN model, the state transition matrix of hydraulic pump $\boldsymbol{A}_C$ can be written as:

$$\boldsymbol{A}_C = \begin{bmatrix} -2.2\times10^{-6} & 0 & 0 & 0 \\ 1.76\times10^{-6} & -3.6\times10^{-6} & 0 & 0 \\ 0 & 2.16\times10^{-6} & -6.7\times10^{-6} & 0 \\ 4.4\times10^{-7} & 0 & 6.7\times10^{-6} & 0 \end{bmatrix} \tag{36}$$

Then, the steady probabilities $\boldsymbol{P}_C(t)$ of hydraulic cylinder are listed as:

$$\boldsymbol{P}_C(t) = \begin{bmatrix} P_C^{(0)}(t) \\ P_C^{(1)}(t) \\ P_C^{(2)}(t) \\ P_C^{(3)}(t) \end{bmatrix} = \begin{bmatrix} e^{-2.2\times10^{-6}t} \\ 1.26\times(e^{-2.2\times10^{-6}t} - e^{-3.6\times10^{-6}t}) \\ 0.60e^{-2.2\times10^{-6}t} - 0.87e^{-3.6\times10^{-6}t} + 0.27e^{-6.7\times10^{-6}t} \\ 1 - 2.86e^{-2.2\times10^{-6}t} + 2.13e^{-3.6\times10^{-6}t} - 0.27e^{-6.7\times10^{-6}t} \end{bmatrix} \tag{37}$$

Therefore, for a hydraulic cylinder, the stochastic power transmission model, $\text{PTU}(\eta_C, \boldsymbol{A}_C, \boldsymbol{P}_C(t))$, can be written as:

$$\begin{aligned} \text{PTU}(\eta_C, \boldsymbol{A}_C, \boldsymbol{P}_C(t)) &= P_C^{(0)}(t) z^{\eta_C^{(0)}} + P_C^{(1)}(t) z^{\eta_C^{(1)}} + P_C^{(2)}(t) z^{\eta_C^{(2)}} + P_C^{(3)}(t) z^{\eta_C^{(3)}} \\ &= e^{-2.2\times10^{-6}t} z^{0.975} + 1.26\times(e^{-2.2\times10^{-6}t} - e^{-3.6\times10^{-6}t}) z^{0.9} \\ &\quad + \left[0.60e^{-2.2\times10^{-6}t} - 0.87e^{-3.6\times10^{-6}t} + 0.27e^{-6.7\times10^{-6}t}\right] z^{0.8} \\ &\quad + \left[1 - 2.86e^{-2.2\times10^{-6}t} + 2.13e^{-3.6\times10^{-6}t} - 0.27e^{-6.7\times10^{-6}t}\right] z^{0.375} \end{aligned} \tag{38}$$

3.2.4. Control surface

Control surface is one of the terminal customers, so the output power of DHAS is used to drive the control surface to implement the aircraft flight control. To simplify the PTE-based model, we defined the mechanical efficiency of the control surface as $\eta_{sur} = 0.95$. Its mechanical failure rate $\lambda_{sur} = 2.2\times10^{-6}$ / h is selected from Table 1, then we can get the reliability of control surface as $P_{sur}(t) = e^{-3.2\times10^{-6}t}$, therefore:

$$\text{PTU}(\eta_{sur}, P_{sur}) = e^{-3.2\times10^{-6}t} \times z^{0.95} \tag{39}$$

3.2.5. Engine accessory gearbox

The aircraft engine provides initial power input for aircraft power supply and actuation system through engine accessory gearbox. The mechanical energy is provided by the engine accessory gearbox. Similar with the control surface, it is supposed that the mechanical efficiency of the accessory gearbox is $\eta_G = 0.853$, and the reliability of accessory gearbox is $P_G(t) = e^{-3.0\times10^{-6}t}$, then the PTE-based model of engine accessory gearbox is obtained as:

$$\text{PTU}(\eta_G, P_G) = e^{-3.0\times10^{-6}t} \cdot z^{0.853} \tag{40}$$

3.3. PTE-based reliability model of DHAS

The PTE-based model for single hydraulic actuation system (HAS) is given by:

$$\begin{aligned} \text{HAS}(\eta_{HAS_1}, \boldsymbol{P}_{HAS_1}(t)) &= (\eta_{P_1}, \boldsymbol{A}_{P_1}, \boldsymbol{P}_{P1}(t)) \otimes (\eta_{V_1}, \boldsymbol{A}_{V_1}, \boldsymbol{P}_{V_1}(t)) \otimes (\eta_{C_1}, \boldsymbol{A}_{C_1} \boldsymbol{P}_{C_1}(t)) \\ &= \sum_{i=0}^{3} \sum_{j=0}^{15} \sum_{k=0}^{3} P_P^{(i)}(t) P_V^{(j)}(t) P_C^{(k)}(t) \cdot z^{\eta_P^{(i)} \eta_V^{(j)} \eta_C^{(k)}} \\ \text{HAS}(\eta_{HAS_2}, \boldsymbol{P}_{HAS_2}(t)) &= (\eta_{P_2}, \boldsymbol{A}_{P_2}, \boldsymbol{P}_{P_2}(t)) \otimes (\eta_{V_2}, \boldsymbol{A}_{V_2}, \boldsymbol{P}_{V_2}(t)) \otimes (\eta_{C_2}, \boldsymbol{A}_{C_2} \boldsymbol{P}_{C_2}(t)) \\ &= \sum_{i=0}^{3} \sum_{j=0}^{15} \sum_{k=0}^{3} P_P^{(i)}(t) P_V^{(j)}(t) P_C^{(k)}(t) \cdot z^{\eta_P^{(i)} \eta_V^{(j)} \eta_C^{(k)}} \end{aligned} \tag{41}$$

The complete PTE-based model of DHAS is composed of dual HASs, engine accessory gearbox and control surface, thus:

$$\begin{aligned} \text{DHAS}(\eta_{DHAS}, \boldsymbol{P}_{DHAS}(t)) &= (\eta_G, P_G(t)) \otimes \left(\text{HAS}(\eta_{HAS_1}, \boldsymbol{P}_{HAS_1}(t)) \oplus \text{HAS}(\eta_{HAS_2}, \boldsymbol{P}_{HAS_2}(t))\right) \otimes (\eta_{sur}, P_{sur}(t)) \\ &= (\eta_G, P_G(t)) \otimes \left(\sum_{i=0}^{n_{HAS_1}-1} \sum_{j=0}^{n_{HAS_2}-1} P_i^{(HAS_1)}(t) P_j^{(HAS_2)}(t) z^{\frac{\eta_{HAS_1}^{(i)} + \eta_{HAS_2}^{(j)}}{2}}\right) \otimes (\eta_{sur}, P_{sur}(t)) \\ &= e^{-2.2\times10^{-6}t} \cdot z^{0.853} \otimes \left(\sum_{i=0}^{n_{HAS_1}-1} \sum_{j=0}^{n_{HAS_2}-1} P_i^{(HAS_1)}(t) P_j^{(HAS_2)}(t) z^{\frac{\eta_{HAS_1}^{(i)} + \eta_{HAS_2}^{(j)}}{2}}\right) \otimes e^{-2.2\times10^{-6}t} \cdot z^{0.95} \end{aligned} \tag{42}$$

Given the power input of DHAS, $W_{DHAS}^{(in)} = 66.5\text{kW}$, $\text{DHAS}(\boldsymbol{W}_{DHAS}^{(out)}, \boldsymbol{P}_{DHAS}(t))$ can be expressed as:

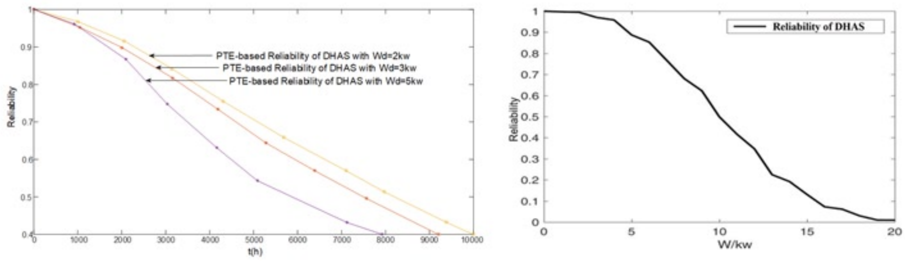
$$\text{DHAS}(\boldsymbol{W}_{DHAS}^{(out)}, \boldsymbol{P}_{DHAS}(t)) = e^{-2.2\times10^{-6}t} \cdot z^{56.72} \otimes \left(\sum_{i=0}^{n_{HAS_1}-1} \sum_{j=0}^{n_{HAS_2}-1} P_i^{(HAS_1)}(t) P_j^{(HAS_2)}(t) z^{\frac{\eta_{HAS_1}^{(i)} + \eta_{HAS_2}^{(j)}}{2}}\right) \otimes e^{-2.2\times10^{-6}t} \cdot z^{0.95} \tag{43}$$

According to Eq. (43), we can get combination items of $(W_{DHAS}^{(out)i}, P_{DHAS}^{(i)}(t))$, $i = 0,1,2,...,n_{DHAS} - 1$. Then the PTE-based reliability model of DHAS can be written as:

$$R^p\left(t \mid \boldsymbol{W}_{DHAS}^{(out)} \geq W^{(d)}\right) = \sum_{i=1}^{m} \left\{P_{DHAS}^{(i)}(t) \cdot 1\left(W_{DHAS}^{(out)i} - W^{(d)} \geq 0\right)\right\} \tag{44}$$

where $W^{(d)}$ is actual control surface driving power demand.

From Eq. (43) and Eq. (44), we can calculate the reliability of DHAS with the constant demand of $W^{(d)} \in [0,20]$ (kW) as shown in Fig. 17.



(a) *PTE-based Reliability of DHAS*

(b) *Reliability of DHAS under changeable demand,* $t = 100$ h

Fig. 17. PTE-based Reliability curves of DHAS

The power transfer process of a component is different from fault mode transmission process, which will pose a different impact on DHAS. In addition, the variable demand is another key factor that reflects the performance of DHAS. From Fig. 17 (b), it can be concluded that the reliability of DHAS decreases as the demand of the control surface increases.

3.3.1. Reliability analysis of DHAS under variable power demands

Since the power requirements of aileron actuation system are different in different stage of flight profiles, the power consumption at different flight stage is also various. Table 6 shows the power requirement at conventional flight profile of an aircraft, which includes taxing, taking off, climbing, cruising, descending and landing.

Table 6 shows the different power demand of each flight profile, in which power means the power requirement at different flight stages and proportion indicates the proportion of power at different stages to the power of whole flight profile. It shows that the more power the demand needs in the flight profile, the better the maneuverability is. We established the PTE-based reliability model at different stages for DHAS and calculated its reliability shown in Fig. 18.

Table 6. Power consumption under 6 flight profile and duration of an aileron [20]

Phase	Taxing	Taking off	Climbing	Cruising	Descending	Landing
Power (kW)	0.98	1.71	1.31	1.25	1.51	2.35
Proportion	0.03	0.02	0.04	0.85	0.04	0.02

Fig. 18 shows that the reliability diversification is sensitive to the large power requirement profiles such as taking off, descending and landing, so the power-based reliability of DHAS needs to be seriously considered in such operational condition. However, the power-based reliability has little change in some small power demand such as cruise and taxiing.

3.3.2. Reliability analysis of DHAS with the change of hydraulic pump's flow rate

Hydraulic pump is the key component of power supply system of DHAS. The pump maintains the constant pressure while changes the flow rate accordingly to meet the mission demand. The flow rate supply of a kind of hydraulic pump, during different flight stages, is listed in Table 7.

Applying the numerical procedure for hydraulic pump presented in Section 3.2.1, the DHAS reliability curve can be obtained shown in Fig. 19 with the changes of flow rate.

In Fig. 19, the reliability of DHAS exits a trough in the middle of power requirement and corresponding flow rate, which means that DHAS cannot drive the control surface according to the control command even when DHAS can provide partial power. With the power approaches the power requirement, the reliability increases and keeps at a high level for a long time. In such condition, aircraft can operate in satisfied performance under enough power supply. In addition, the reliability of DHAS keeps at a high level when the hydraulic pump changes its flow rate to meet the power supply needs according to the flight command. If we observe a two-dimensional view of the three-dimensional reliability curve, we can see that the different variation tendency with the change of pump flow rate or the change of power requirement. The ideal situation is the reliability keeping high with appropriate power requirement and appropriate power supply. We can get the optimal value with the reliability as a

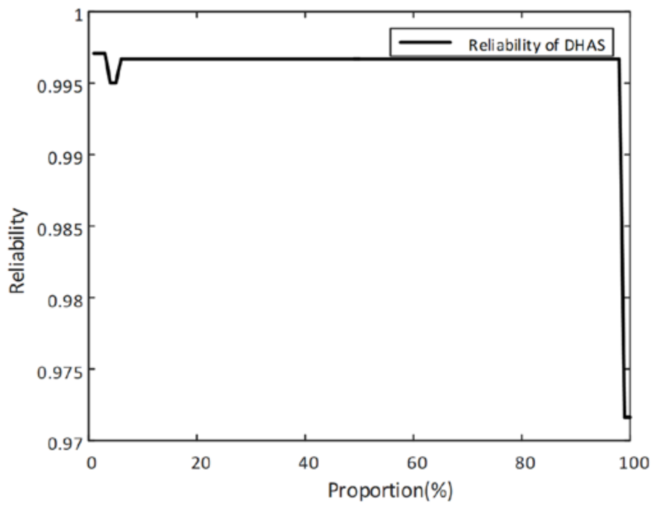


Fig. 18. Reliability of aileron driven by DHAS under flight profile, total $t = 100$h

Table 7. The flow rate of pump in different flight stage [20]

Flight stage	Pitch check	Roll check	Yaw check	Take off start	Stop taking off	Take off	Climbing	Cruise	Emergency decline
Flow rate (gpm)	6.33	11.7	9.47	12.27	21.84	4.45	4.45	4.55	7.04

Flight stage	Air collision	Decline	Turbulence	Go around	Spoiler unfoldinhg	Landing brake	Landing run	Stop landing	Slide
Flow rate (gpm)	7.21	4.55	7.9	10.45	19.74	6.65	21.94	4.55	6.15

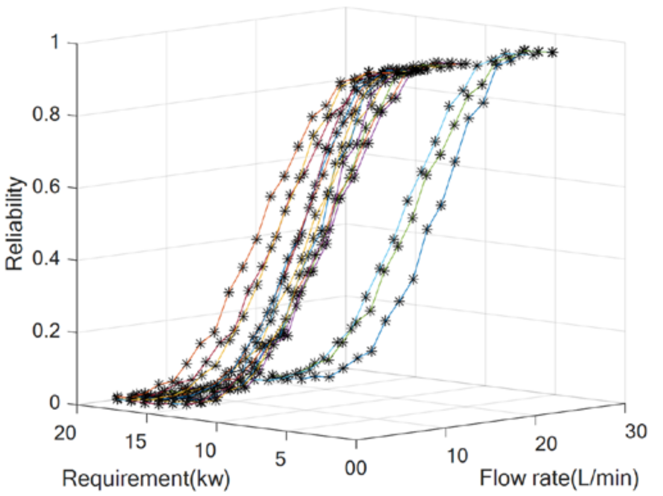


Fig. 19. Reliability of DHAS with the changing flow rate of pump under changing demand

objective function, the power requirement and pump flow rate supply as a restriction. It also indicates that the PTE-based reliability model of DHAS can provide integrated evaluation on DHAS task completion effectively considering reliability and power changes.

3.3.3. Reliability analysis of DHAS with different hydraulic pump degradation rates

In order to analyze the influence of hydraulic pump degradation to DHAS reliability in the GSPN model, we change the degrading failure rate λ_{P1} and calculate the reliability of DHAS under different power requirements in $t = 100\text{h}$ as shown in Fig. 20.

Fig. 20 gives the sensitivity analysis of PTE-based reliability of DHAS to degradation rate of hydraulic pump. Fig. 20. (a) indicates that the reliability of DHAS is less than 0.01% under terminal power demand of 1 kW while the degradation rate of pump changes from 0.1×10^{-4} / h to 5×10^{-4} / h . Fig. 20 (b) shows the variation of the PTE-based reliability of DHAS reaches 5.5% with terminal power demand ranging from 1 kW to 5 kW while the degradation rate changes from 0.1×10^{-5} / h to 5×10^{-5} / h . Fig. 20 (c) expresses the variation of the PTE-based reliability of DHAS reaches 18% under terminal power demand as 20 kW. So PTE-based reliability is sensitive to the terminal power requirements.

3.3.4. Comparison between PTE-based reliability and traditional reliability of DHAS

In order to verify the effectiveness of the proposed reliability analysis, we compare the PTE-based reliability model proposed in this paper with the traditional reliability model based on RBD for DHAS. The reliability curve under proposed method and the RBD with the parameters of Table 1 is shown in Fig. 21.

In Fig. 21, the blue line means the reliability based on RBD, the yellow line indicates the reliability under 2 kW power requirement, the red line is the reliability under 3 kW power demand and the purple line expresses the reliability under 5 kW power demand. It is apparently that the traditional reliability of DHAS is higher than PTE-based reliability if we don't consider the power requirement. However, DHAS is the fast response system that needs enough power supply to implement large maneuvering flight, so it is necessary to consider the dynamic performance related to the power supply. From the reliability curve of different power requirements in Fig. 21, it indicates that the PTE-reliability is closely related to the power demands. When the response time is not very high, the power demand is small and the PTE-based reliability decreases gradually. When the DHAS needs fast maneuvering, PTE-based reliability drops quickly. Hence, the PTE-based reliability is the appropriate reliability model to evaluate its reliability under satisfied dynamic performance.

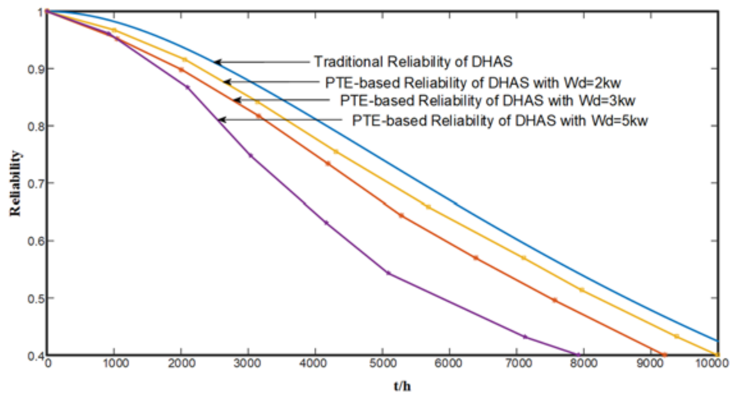
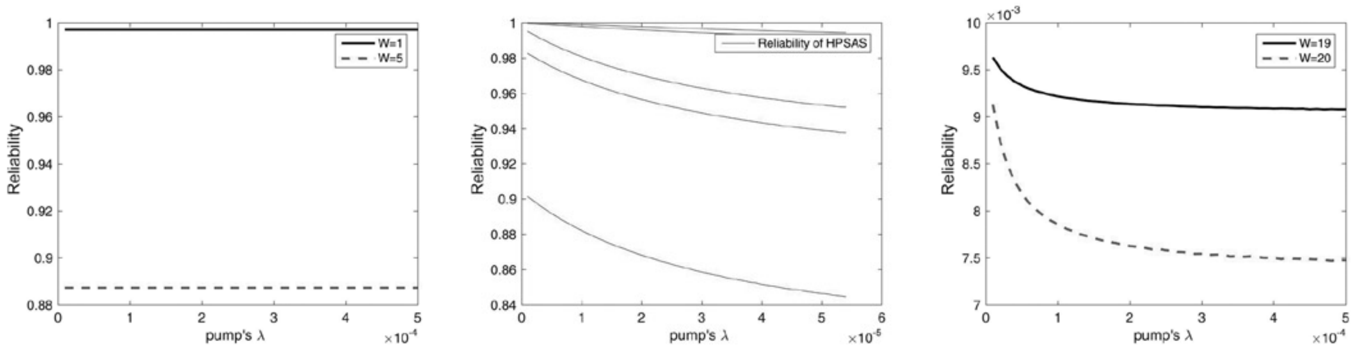


Fig. 21. Reliability comparison between PTE-based model and BRD model

4. Conclusion

This paper proposes a PTE-based reliability model considering the PTE and degradation process of component and system. In PTE-based reliability model, the power requirement, power capacity and



(a) $W^{(d)} = 1\text{kW}, 5\text{kW}$ *(b)* $W^{(d)} = 1\text{kW} \sim 5\text{kW}$ *(c)* $W^{(d)} = 19\text{kW}, 20\text{kW}$

Fig. 20. Reliability comparison of DHAS with changing physical degradation rate of pump

PTE are randomized to describe its reliability. Due to the inherernt degradation of PTS, this paper presents a stochastic power transimission model under multiple degradation states, utlizing the GSPN and UGF to calculate the power-based reliability with specific power requirement. DHAS is taken as a case to show how to carry out the PTE-reliability evaluation for dynamic performance related system. Based on the PTE-relability model of hydraulic pump, servo valve, hydraulic cylinder and control surface, this paper establishes the PTE-reliability model of DHAS and gives the comparsion between the trdational reliability model and the proposed method. The results indicate that the PTE-based reliability model is more suitable for the system requiring dynamic perfromace under enough power supply. PTE-based reliability model can characterize the essential reliability characteristics of DHAS accurately under multiple power demands and performance requirments.

Acknowledgement

The research work is supported by the National Natural Science Foundation of China (Grants No. 51875014,51620105010, 51575019).

References

1. Benabid R, Merrouche D, Bourenane A, and Alzbutas R. Reliability Assessment of Redundant Electrical Power Supply Systems using Fault Tree Analysis, Reliability Block Diagram, and Monte Carlo Simulation Methods. 2018 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM). IEEE 2018: 1-7, https://doi.org/10.1109/CISTEM.2018.8613431.
2. Bennett J W, Atkinson G J, Mecrow B C, et al. Fault-tolerant design considerations and control strategies for aerospace drives. IEEE Transactions on Industrial Electronics 2011; 59(5): 2049-2058, https://doi.org/10.1109/TIE.2011.2159356.
3. Cepin, M. Evaluation of the importance factors of the power plants within the power system reliability evaluation. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21(4): 631-637, https://doi.org/10.17531/ein.2019.4.12.
4. Chakraborty I, Trawick D R, Mavris D N, et al. A requirements-driven methodology for integrating subsystem architecture sizing and analysis into the conceptual aircraft design phase. 14th AIAA Aviation Technology, Integration, and Operations Conference 2014: 3012, https://doi.org/10.2514/6.2014-3012.
5. Cui X, Wang S, Li T, and Shi J. System Reliability Assessment Based on Energy Dissipation: Modeling and Application in Electro-Hydrostatic Actuation System. Energies 2019; 12(18): 3572, https://doi.org/10.3390/en12183572.
6. Dong W, Liu S, Tao L, Cao Y, and Fang Z. Reliability variation of multi-state components with inertial effect of deteriorating output performances. Reliability Engineering & System Safety 2019 186: 176-185, https://doi.org/10.1016/j.ress.2019.02.018.
7. Hespanha J P. Modeling and analysis of networked control systems using stochastic hybrid systems. Annual Reviews in Control 2014; 38(2):155-170, https://doi.org/10.1016/j.arcontrol.2014.09.001.
8. Hirsch W M, Meisner M, Boll C. Cannibalization in multicomponent systems and the theory of reliability. Naval Research Logistics Quarterly 1968; 15(3): 331-360, https://doi.org/10.1002/nav.3800150302.
9. IEC, Functional Safety of Electrical/Electronic/Programmable Electronic Safety Related Systems, IEC 61508, 2000.
10. Jafary B, Fiondella L. A universal generating function-based multi-state system performance model subject to correlated failures. Reliability Engineering & System Safety 2016; 152: 16-27, https://doi.org/10.1016/j.ress.2016.02.004.
11. Jia H, Levitin G, Ding Y, et al. Reliability analysis of standby systems with multi-state elements subject to constant transition rates. Quality and Reliability Engineering International 2019; 35(1): 318-328, https://doi.org/10.1002/qre.2401.
12. Lee S H. Reliability evaluation of a flow network. IEEE Transactions on Reliability 1980; 29(1): 24-26, https://doi.org/10.1109/TR.1980.5220695.
13. Levitin G. The universal generating function in reliability analysis and optimization. London: Springer, 2005.
14. Li S, Zhu Z C, Lu H, Shen G. A system reliability-based design optimization for the scraper chain of scraper conveyors with dependent fsilure modes. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21(3): 392-402, https://doi.org/10.17531/ein.2019.3.5.
15. Li X Y, Huang H Z, Li Y F, et al. Reliability assessment of multi-state phased mission system with non-repairable multi-state components. Applied Mathematical Modelling 2018; 61: 181-199, https://doi.org/10.1016/j.apm.2018.04.008.
16. Lin Y H, Li Y F, Zio E. Integrating random shocks into multi-state physics models of degradation processes for component reliability assessment. IEEE Transactions on Reliability 2014; 64(1): 154-166, https://doi.org/10.1109/TR.2014.2354874.
17. Lin Y K. A simple algorithm for reliability evaluation of a stochastic-flow network with node failure. Computers & Operations Research 2001; 28(13): 1277-1285, https://doi.org/10.1016/S0305-0548(00)00039-3.
18. Lisnianski A, Frenkel I, Ding Y. Multi-state system reliability analysis and optimization for engineers and industrial managers. Springer Science & Business Media 2010, https://doi.org/10.1007/978-1-84996-320-6.
19. Lisnianski A, Levitin G. Multi-state system reliability: assessment, optimization and applications. World Scientific Publishing Company 2003, https://doi.org/10.1142/5221.
20. Ma Z, Wang S, Zhang C, et al. A Load Sequence Design Method for Hydraulic Piston Pump Based on Time-Related Markov Matrix. IEEE Transactions on Reliability 2018; 67(3): 1237-1248, https://doi.org/10.1109/TR.2018.2830330.
21. Malinowski J. Reliability analysis of a flow network with a series-parallel-reducible structure. IEEE Transactions on Reliability 2015; 65(2): 851-859, https://doi.org/10.1109/TR.2015.2499962.
22. Peng H, Feng Q and 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.
23. Qin J L, Niu Y G, Li Z. A combined method for reliability analysis of multi-state system of minor-repairable components. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18(1): 80-88,https://doi.org/10.17531/ein.2016.1.11.
24. Raveendran V, Andresen M and Liserre M. Improving onboard converter reliability for more electric aircraft with lifetime-based control. IEEE Transactions on Industrial Electronics 2019; 66(7): 5787-5796, https://doi.org/10.1109/TIE.2018.2889626.
25. Rosero J A, Ortega J A, Aldabas E, et al. Moving towards a more electric aircraft. IEEE Aerospace and Electronic Systems Magazine 2007; 22(3): 3-9, https://doi.org/10.1109/MAES.2007.340500.
26. Steffen T, Schiller F, Blum M, and Dixon R. Analysing the reliability of actuation elements in series and parallel configurations for high-redundancy actuation. International Journal of Systems Science 2013; 44(8): 1504-1521, https://doi.org/10.1080/00207721.2012.659694.

27. Ushakov I A. Optimal standby problems and a universal generating function. Soviet journal of computer and systems sciences, 1987; 25(4): 79-82.
28. Wang K, Wang S, Shi J. A novel multi-state reliability assessment model for servo HA/EHA system via universal generating function. 2017 IEEE International Conference on Mechatronics and Automation (ICMA). IEEE 2017: 1918-1923, https://doi.org/10.1109/ICMA.2017.8016111.
29. Wang S, Cui X, Shi J, Tomovic M M, and Jiao Z. Modeling of reliability and performance assessment of a dissimilar redundancy actuation system with failure monitoring. Chinese Journal of Aeronautics 2016; 29(3): 799-813, https://doi.org/10.1016/j.cja.2015.10.002.
30. Wang W, Di Maio F, Zio E. Three-loop Monte Carlo simulation approach to Multi-State Physics Modeling for system reliability assessment. Reliability Engineering & System Safety 2017; 167: 276-289, https://doi.org/10.1016/j.ress.2017.06.003.
31. Wang Y S, Fang X, Zhang C H , Chen X, Lu J Z. Lifetime prediction of self-lubricating spherical plain bearings based on physics-of-failure model and accelerated degradation test. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18(4): 528-538, https://doi.org/10.17531/ein.2016.4.7.
32. Wu J, Yan S, Xie L, et al. Reliability apportionment approach for spacecraft solar array using fuzzy reasoning Petri net and fuzzy comprehensive evaluation. Acta Astronautica 2012; 76: 136-144, https://doi.org/10.1016/j.actaastro.2012.02.023.
33. Xu D, Feng Z, Sui S, and Lin Y. Reliability Assessment of Electro-hydraulic Actuator Control System Subject to Multi-Sources Degradation Processes. IEEE/ASME Transactions on Mechatronics 2019, https://doi.org/10.1109/TMECH.2019.2953333.
34. Yu H, Chu C, Châtelet Ė, et al. Reliability optimization of a redundant system with failure dependencies. Reliability Engineering & System Safety 2007; 92(12): 1627-1634, https://doi.org/10.1016/j.ress.2006.09.015.
35. Zhang J, Qun C and Xu B. Analysis of the cylinder block tilting inertia moment and its effect on the performance of high-speed electro-hydrostatic actuator pumps of aircraft. Chinese Journal of Aeronautics 2018; 31(1): 169-177, https://doi.org/10.1016/j.cja.2017.02.010.
36. Zio E. An introduction to the basics of reliability and risk analysis. World scientific 2007, https://doi.org/10.1142/6442.
37. Zio E. Some Challenges and Opportunities in Reliability Engineering. IEEE Transactions on Reliability 2016; 65(4):1769-1782, https://doi.org/10.1109/TR.2016.2591504.

Xiaoyu CUI
School of Materials Science and Mechanical Engineering
Beijing Technology and Business University
100048, Beijing, China

Tongyang LI
China Civil Aviation Engineering Consulting Co., Ltd
100621, Beijing, China

Shaoping WANG
School of Automation Science and Electrical Engineering
Beihang University
100191, Beijing, China

Jian SHI
School of Automation Science and Electrical Engineering
Beihang University
100191, Beijing, China

Zhonghai MA
Beijing Key Laboratory of Advanced Manufacturing Technology
Beijing University of Technology
100124, Beijing, China

E-mails: xiaoyucui@btbu.edu.cn, soredsun@buaa.edu.cn,
shaopingwang@vip.sina.com, shijian@buaa.edu.cn, zm@bjut.edu.cn

Article citation info:
WYMYSŁOWSKI A, JANKOWSKI K. Strength analysis of solder joints used in microelectronics packaging. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 297–305, http://dx.doi.org/10.17531/ein.2020.2.12.

Artur WYMYSŁOWSKI
Krystian JANKOWSKI

STRENGTH ANALYSIS OF SOLDER JOINTS USED IN MICROELECTRONICS PACKAGING

BADANIA WYTRZYMAŁOŚCI POŁĄCZEŃ LUTOWANYCH STOSOWANYCH W MONTAŻU W MIKROELEKTRONICE*

The aim of the research was the problem of damage accumulation for solder alloys used in microelectronics packaging due to creep and fatigue as a result of a combined profile of loading conditions. The selected failure modes affect the lifetime of contemporary electronic equipment. So far the research activities are focused on a single failure mode and the problem of their interaction is often omitted. Taking into account the failure modes interaction would allow more precise lifetime prediction of the contemporary electronic equipment and/or would allow for reduction of time required for reliability tests. Within the taken research framework the reliability analysis of solder joints was conducted for the Sn63Pb37 solder alloy using the Hot Bump Pull method. The results of the presented research contain: reliability tests, statistical analysis and the problem of a damage accumulation due to a combined profile of loading conditions.

Keywords: *microelectronics, solder alloys, reliability, damage accumulation.*

Celem badań był problem kumulacji uszkodzeń dla stopów lutowniczych stosowanych w montażu w mikroelektronice w wyniku zmęczenia i pełzania na skutek złożonego profilu obciążeń. Wybrane rodzaje uszkodzeń przyczyniają się do ograniczenia czasu życia współczesnych urządzeń elektronicznych. Aktualnie prowadzi się badania z wykorzystaniem jednego rodzaju uszkodzeń i często pomijany jest problem ich wzajemnej interakcji. Uwzględnienie problemu wzajemnej interakcji pozwoliłoby na bardziej precyzyjne prognozowanie bezawaryjnego czasu pracy współczesnych urządzeń elektronicznych i/lub przyspieszenie testów niezawodnościowych. W ramach zrealizowanych badań przeprowadzono analizę wytrzymałości połączeń lutowanych dla stopu lutowniczego Sn63Pb37 z wykorzystaniem metody Hot Bump Pull. Wyniki przedstawionych badań obejmują: analizę wytrzymałości, analizę statystyczną oraz problem kumulacji uszkodzeń w wyniku złożonego profilu obciążeń.

Słowa kluczowe: *mikroelektronika, stopy lutownicze, niezawodność, kumulacja uszkodzeń.*

1. Introduction

Electronics is one of the quickest developing discipline of contemporary science and engineering. Due to the constant pursuit for miniaturization and integration most of the electronic components are designed and manufactured in the so-called micro scale. For this reason, the special term microelectronics was established among professionals. Nowadays, microelectronic components are an integral part of every industrial or home use electronic device. Unfortunately, like other devices, also microelectronic components experience a limited lifetime. One of the basic problems regarding their reliability are connections. In microelectronics packaging [17] soldered, glued and bonded connections are used, of which solder joints are the most important [13, 15, 27]. Most damages of the to solder joints occur as a result of thermomechanical loading, and their direct cause is stress resulting from the mismatch of coefficients of thermal expansion of the joined materials [17, 35, 40]. It is estimated that about 65% of damages in microelectronics packaging is associated to thermomechanical problems [2, 38].

Reliability is defined as an object property that determines its proper operation under given environmental conditions, over a defined period of time. Mathematical description of the reliability allows for probability assessment of the object failure in the defined operating conditions. One of the traditional ways of reliability prediction of joints in electronics packaging is theoretical analysis based of the so-called bimaterial interface. Bimaterial interface refers to the mechanical connection of two materials with different thermomechanical properties. Strength analysis requires in this case knowledge on stress distribution in the joint and nearby region. There are two possible failure modes: crack and delamination. The analytical solution for the bimaterial interface was proposed for the first time in 1925 by Timoshenko [30]. The solution referred to the structure deformation and maximal stress in the interface region. While the analytical solution describing the stress distribution in the interface region was proposed in 1989 by Suhir [29]. Analytical description of the stress distribution in the bimaterial interface region is difficult and requires a number of simplification, e.g. geometry of the structure and linear material model. Therefore, in modern engineering applications, numerical methods based on simulations using the FEM analysis are preferred [39, 41]. The FEM method allows simulations of real interface geometry and taking into account nonlinear material models. Additionally, along with the strength, numerical methods allow an inclusion during analysis failure criteria in order to predict reliability. Failure criteria are closely connected with the failure mode and material type. In microelectronics failures are recognized through reliability tests while the corresponding failure criteria are determined experimentally. The above problem was published lately in 2018 in a form of the standard document IPC/JEDECC-9301 "Numerical Analysis Guidelines for Microelectronics Packaging Design and Reliability" [24].

Due to dynamic development of electronic industry and a rapid changes of microelectronic technologies the crucial issue is an advanced research concerning reliability analysis as a consequence of

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

constantly growing integration and miniaturization of the electronic components. Solder joints are a good example of the above trend, as their dimensions being smaller than 100 μm are comparable with the dimension of single grains. Taking into account both the solder joint dimensions and different loading conditions of electronic components it came out that reliability prediction methods used in macro scale are not appropriate for micro scale [36]. For this reason, it is necessary to develop, on the one hand, advanced measurement techniques, typical for microelectronic technologies, while on the other hand, advanced methods of predicting failures due to thermomechanical loads [25]. The above requires knowledge on material behavior as a function of stress, temperature and time, which is refereed as rheology. One of the typical problems concerning rheology is analysis of creep and fatigue phenomena [20]. It should be underlined that in many fields of engineering, both fatigue and creep are treated as the main failure modes. Many researchers also emphasize the important role of their interactions, especially at elevated temperatures [10, 34]. In case of solder joints, used in microelectronic packaging, failures due to material creep and / or fatigue caused by thermomechanical loadings play a vital role in order to predict reliability accuratly [4, 6, 16].

2. Packaging and reliability of solder joints in microelectronics.

The concept of microelectronic packaging is associated with a number of activities and technological stages leading to a functional device. The goal of packaging is to provide good electrical connection, appropriate mechanical properties and heat transport. In order to achieve that, there are used various joining techniques such as bonding, soldering, gluing. As a result of the above diversity packaging techniques were divided into levels [8, 31]:

- zero: carried out at the production level and concerns the mounting of components on the semiconductor wafer surface,
- first: concerns the problem of making connections on the silicon chip and its packaging in an enclosure,
- second: packaging of silicon integrated circuits and other electronic components both active and passive on the surface of printed circuit board PCB,
- third: assembly of electronic modules in the form of individual PCB and other components in order to achieve the functional blocks and finally into a complete electronic device.

All types of connections used on the mentioned packaging levels must meet the reliability requirements in order to ensure long-term and appropriate operation of the device. The current research was devoted to the analysis of the solder joint reliability used at the second packaging level [12].

In fact, the goal of the research was to run a series of experimental strength tests of solder joints used in microelectronic packaging due to a combined loading profile. This allowed for analysis of typical failure modes, as creep and fatigue and additionally their interaction phenomenon. According to authors, the used measuring technique and the obtained results will allow in the future more accurate prediction of the solder joints strength at micro scale, and thus improve the reliability of microelectronic components. The presented research refers to the solder joints made with the traditional solder alloy $Sn_{63}Pb_{37}$. It should be underlined, that the EU directive RoHS (Restriction of Hazardous Substances) limited, starting form 2006, the sell of new electronic equipment in the EU region that contain hazardous substances such as e.g. lead solder alloys. Nevertheless, the same directive, introduced a number of exceptions to that rule, in cases where replacement of a certain material would be difficult or impossible and additionally in the case of research and development application. Traditional lead solder alloy $Sn_{63}Pb_{37}$ is characterized by a unique thermomechanical material properties and allows for durable and consistent connection of components. For this reason it is still used in the case of such equipment as medical, military, etc. Additionally, due to high strength of solder joints made with this alloy, it is used for comparison reasons of reliability tests in scientific research. Besides, the cited directive does not restrict the use of the alloy in the case of amateur applications.

Unfortunately, reliability tests are long-lasting and costly basically due to the fact that the results require statistical analysis. The advantage of such a procedure is the ability of estimation of the durability time of the device, while the basic disadvantages are:

- statistical analysis requires running a high number of single tests, which leads to long-lasting experiments; it is assumed that in the case of microelectronics components the realiability tests last from a few over a dozen of months,
- analysis of a single failure mode leads to a wrong reliability prediction, because in real operating environmental conditions there coexists a number of different failure modes and their interactions.

Currently there are only a few research reports containing description and results of strength analysis of solder joints in micro scale due to an interaction of failure modes as a result of combined loading profile [4,6,11]. For this reason, the presented results fit into a current research trend devoted to the problem of simultaneous occurrence of different failure modes and including the scale of analysis. The key problem was the measurement of small (from an order of fraction and up to a couple of micrometers) displacements, that occur in the joint and its nearby region. The periodically changed force applied to the solder joint allowed observation of both elastic and inelastic displacements.

2.1. Solder alloys and solder joints in microelectronics

As it is been mentioned earlier, in reference to EU directive RoHS, currently used solder alloys in microelectronics are subjected to restrictions concerning the problem of reduction of hazardous substances that can affect the environment. Unfortunately, replacement of traditional lead solder alloys requires technological changes, e.g. higher soldering temperature, adaptation of assembly lines for soldering at higher temperatures, adaptation of electronic components for assembling at a higher temperature, introduction of new fluxes enabling adequate surface wetting of solder alloys requiring higher temperatures, etc. On the other hand it should be underlined that traditional solder alloy based on lead and tin ($Sn_{63}Pb_{37}$) posses very good thermomechanical properties. In case of the lead-free solder alloys mainly are used such metals as tin, silver and copper, which establishes abbreviation for such alloys as SAC (SnAgCu). As the percentage content of each metal can be different therefore there is variety of alloys, e.g. SAC305 ($Sn_{96.5}Ag_{3.0}Cu_{0.5}$) [3,9].

Mechanical behavior of solder joints depends on various factors: alloy microstructure, content of intermatallic compounds, the size of a joint or sample, the cooling rates after the joint is created or the aging process. Other important factors are: thermomechanical load profile, spread of thermomechanical properties of materials, etc. So far research studies of solder alloys and joints in microelectronics contributed to the elaboration of simplified models characterizing their mechanical behavior and analysis methods based on classical mechanics and lately numerical simulation techniques. The key research problems are mathematical material models of solder alloys, which describe their behavior as a result of thermomechanical loads. In that case both simple and combined models are used. There is also used another classification criterion, which is linearity: linear material models, e.g. elastic and non-linear material models, e.g. plastic, viscous [33, 35].

2.2. Reliability of solder joints in microelectronics

Accurate reliability prediction of solder joints in microelectronic packaging requires application of different experimental techniques and methods, numerical tools as well as taking into account various physical phenomena. It should be underlined that in the case of solder joints in microelectronics, reliability evaluation is usually carried out for a selected single failure mode i.e. the phenomenon of fatigue or creep [1,19,23]. It is worth mentioning that quite often during reliability analysis the problem of residual stresses is neglected. Residual stresses are the result of the performed technological processes. In order to minimize their influence, the technological processes are designed in such a way so the build-in stresses are relaxed, through slow cooling or additional heating process [40].

2.2.1. Fatigue phenomenon

The fatigue phenomenon of the solder alloy describes influence of a cyclic changing thermomechanical loading as a function of time on the solder joint reliability. As a consequence, deformation of a joint occurs, which is then followed by a complete or significant damage. A characteristic feature of the fatigue failure is that a damage can occur at stress level much lower than it is specified by the material strength. Fatigue strength is measured by counting load cycles to failure, which can be converted into time. There are various factors that affect the number of load cycles: the type of load, its value, sequence and duration. In addition, in the case of fatigue, two types of tests can be distinguished:

- low-cycle fatigue (generally<1000), it is associated with the use of high stress value, which leads to the significant inelastic deformations during each cycle, and thus the short lifetime of the tested material,
- high-cycle fatigue (generally>1000), it is associated with the use of small stress value, which leads to elastic deformations during each cycle, and thus to the increased lifetime of the tested material.

An example of a popular test, frequently used in microeletronics packaging, is the low-cycle analysis with an application of the Coffin-Manson empirical model, which relates a dissipation of an inelastic energy ΔW with the fatigue strength N_f:

$$\Delta \mathrm{W} = \mathrm{C} \cdot N_f^a \tag{1}$$

where coefficients C and a are material constants, which are estimated through experimental tests, while amount of inelastic energy ΔW is estimated basing on a hysteresis strain-stress in a selected fatigue cycle [22, 32].

2.2.2. Creep phenomenon

The creep phenomenon of solder alloys, often referred as a "cold flow", occurs in the case of constant or time dependent thermomechanical loads, which leads to permanent deformation, also known as creep deformation. Creep may or may not lead to the material failure. If the joint is unable to accumulate the energy resulting from the load, then it may be transferred to the nearby joint region or may lead to the irreversible deformation as e.g. crack. The creep phenomenon affects practically all materials, but in most cases this process is very slow in time and depends on such parameters as temperature and stress:

$$\varepsilon = f\left(t, T, \sigma\right) \tag{2}$$

where ε is a creep strain, t is a time, T is a temperature, σ is a stress. In mechanics, e.g. for metals and their alloys, the creep phenomenon is observed in the case when the stress value exceeds the yield point or below, the so-called diffusion creep, when the homologous temperature is greater than 0.4 [7]:

$$T_h = \frac{T_o}{T_t} > 0,4 \tag{3}$$

where T_h is a homologous temperature, T_o is an ambient temperature, T_t is a material melting temperature. The melting temperature for $Sn_{63}Pb_{37}$ is *456K*, thus the solder alloy homologous temperature, for room conditions, is higher than 0.6. In the case of bimaterial structures, as e.g. solder joints, the creep is activated by the stress, which is the result of an ambient temperature change and a difference of the coefficients of thermal expansion of the joined materials. The yield stress for the solder alloy $Sn_{63}Pb_{37}$ is around *40MPa*, while the stress value can be estimated using the following formula:

$$\sigma = \Delta\alpha\Delta T \frac{a}{h} \tag{4}$$

where σ is a stress, $\Delta\alpha$ is a difference of the coefficients of thermal expansion of the joined materials, ΔT is a temperature change, a is a distance from a neutral point of the joined materials, h is a thickness of the joint. The creep phenomenon leads to a change of soldered joint parameters due to physical phenomena occurring in the material, e.g. diffusion at grain boundaries, dislocations, cracks and voids [14]. It is also necessary to mention an additional factor having a significant impact on a creep phenomenon of the solder joints in electronics applications, which is an increase of a joint local temperature as a result of the electric current.

2.3. Problem of a damage accumulation.

One of the most crucial problems associated with testing and analyzing the strength of the solder joints in microelectronics packaging is an occurrence of a number of failure modes at the same time or their interaction. For this reason, it is necessary to take into account the problem referred as a damage accumulation. One of the most popular ways to take it into account in engineering applications is a method named by their authors as Palmgren-Miner's method [21]. The method is based on summing partial damages for given fatigue cycles, and allows for determination of the equivalent load amplitude, which is consistent in terms of the total damage with the case of variable load amplitude [28]:

$$\sum_{\mathrm{i}=0}^{k} \frac{n_i}{N_i} = 1 \tag{5}$$

where n_i is a number of cycles for a given load value, N_i is a fatigue limit for a given value of the load amplitude, while k is a number of selected values of load amplitudes. Unfortunately, this hypothesis assumes a lot of simplifications, which means that the result of reliability prediction is error biased. Additionally, the hypothesis refers rather to macroscopic scale and does not include phenomena occurring at the micro scale, which is characteristic for the microelectronics packaging [11]. In the case of simultaneous occurrence of two different failure modes, i.e. fatigue and creep phenomenon, the alternative is to use the linear superposition principle. The principle can be written in a form of the mathematical formula as a sum of fractional parts for both phenomena at a given load amplitude [18]:

$$\sum_{\mathrm{i}=1}^{k} \frac{n_i}{N_f} + \sum_{\mathrm{i}=1}^{k} \frac{\Delta \mathrm{t}_i}{T_c} = 1 \tag{6}$$

where Δt_i is a time interval for a given load amplitude at a constant level, while T_c is a time to object failure at this load amplitude. The first term of the equation describes fractional part for the fatigue cycle, and the second term is responsible for the creep phenomenon. Similarly, to Palmgren-Miner's hypothesis, it is assumed that failure occurs when the accumulated damage value equals *1*. In this case, the equation solution yields a straight line on the solution graph that represents the cumulative reliability of the object. Unfortunately, this model is based on certain simplifications, and thus the obtained results depend, among others on material properties. Moreover, the hypothesis assumes that time intervals Δt_i as a result of compressive or tensile stresses have a similar effect and do not take into account the problem of material hardening or softening. Nonlinear models of damage accumulation refer to those materials whose mechanical or physico-chemical properties change during the reliability tests, which requires additional tests. This is a reason why in the case of engineering application the linear model is preferred. Figure 1 presents the hypothesis of linear damage accumulation and the problem of material hardening and softening phenomena for non-linear models [12].

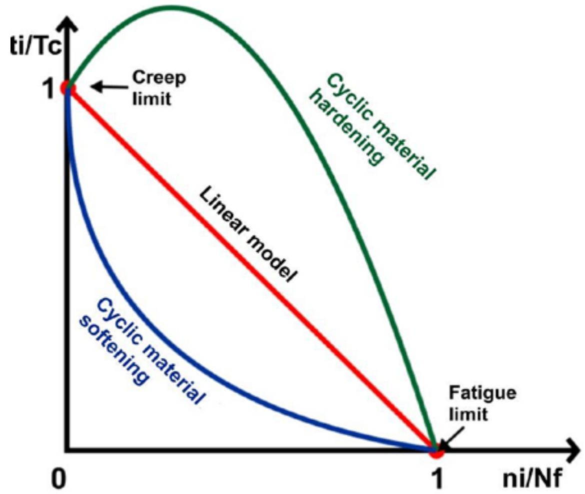


Fig. 1. Linear damage accumulation model and a corresponding hardening and softening phenomena for nonlinear material models

2.4. Statistical analysis

The reliability assessment is usually based on analysis of failures causes and then probability assessment of their occurrence in the given conditions. In order to achieve that, it is necessary to know typical loads, failure modes and the corresponding failure criteria. Unfortunately, the results of reliability tests are defined by random variables, and therefore a quantitative description of reliability requires the use of probabilistic characteristics and probability distributions. In the reliability science, in order to describe the object lifetime, sometimes referred as time-to-failure, probability distributions of failures as a function of time *f(t)* are used. In microelectronics packaging the durability time *t* of an object can described by the two-parameter Weibull distribution:

$$f(t)=\frac{\beta}{\lambda}\cdot\left(\frac{t}{\lambda}\right)^{\beta-1}\cdot e^{-\left(\frac{t}{\lambda}\right)^{\beta}} \tag{7}$$

where β is a shape parameter, λ is a scale parameter. Thus, the probability value that an object will fail by time *t* can be estimated by the cumulative distribution function *F(t)*:

$$F(t)=1-e^{-\left(\frac{t}{\lambda}\right)^{\beta}} \tag{8}$$

The purpose of reliability research is to estimate the values of probability distribution coefficients and to determine selected reliability parameters. In practice, it is convenient to use the concept of time-to-failure of an object corresponding to the numerical value of the scale parameter λ regardless of the value of the shape parameter β. Time-to-failure can be estimated by taking an assumption that $t=\lambda$, thus:

$$F(\mathrm{t}=\lambda)\simeq 0.632 \tag{9}$$

As an alternative solution that one can use is an associated reliability parameter, which is referred as an object strength. For example time-to-failure parameter is correlated with the fatigue strength as follows:

$$F\left(\bar{N}_f\right)\simeq 0{,}632 \tag{10}$$

In engineering practice the scale coefficient λ, corresponding to the fatigue strength, is estimated through graphical methods, which significantly reduces a number of experimental tests. This method allows additionally for:

- estimation of the failure criterion required by numerical prototyping, which means that extracted experimentally time-to-failure is adopted in numerical prototyping as a failure criterion, which in turn allows building empirical failure models as, e.g. the Coffin-Manson model,
- running the so-called accelerated tests, e.g. thermal; in this case, the shift along the abscissa axis on the Weibull plot, depending on the load amplitude, allows for a significant reduction of time required for performing complete reliability tests.

Figure 2 presents an example Weibull plot and a graphical method for experimental data analysis, including estimation of the fatigue strength and failure acceleration coefficient *n* [37].

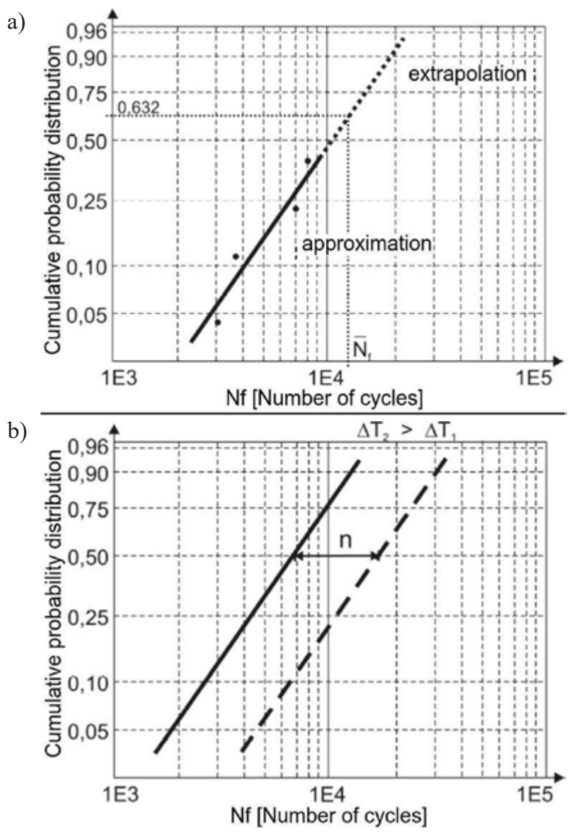


Fig. 2. Example of a graphical analysis method based on the Weibull plot: (a) approximation and extrapolation (b) accelerated thermal tests.

3. Description of the research and achieved results

One of the novel methods, which allows testing of solder joints due to the combined profile of loading conditions, is the Hot Bump Pull HBP method, designed by Nordson Dage company. This method is based on testing the strength of joints by melting a pin into the solder joint and thus allows the study of the phenomenon of creep and fatigue in a microscale. The method allows the use of several methods concerning strength tests: destructive, fatigue, creep and combined creep-fatigue. Within the presented research, a fatigue and creep tests were used, which allowed to develop an appropriate measurement technique and algorithm for reliability model identification of solder joints in the microscale due to fatigue and creep and additionally their interaction [5].

Different testing methods for solder joints are possible thanks to variety of cartridges, that can be exchanged according to the experimental requirements. Available cartridges allow testing wire connections, solder balls, etc, using different tests, such as: pulling, compression, shearing and bending. At the same time, the included software helps the user setting up various test parameters, as:

- maximal and minimal force value,
- dwell time for the maximal and minimal loading value,
- speed of fall and rise time of the combined loading profile,
- setting up and selection of the defined or predefined temperature profiles.

The presented preparing measuring technique was based on testing the strength of solder alloy balls, which corresponded to BGA technology, using a copper pin [12]. It should be underlined that, all the tests were run for $Sn_{63}Pb_{37}$ solder alloy This refereed to the first stage of research devoted to reliability analysis of solder joints used in the microelectronics packaging.

3.1. Method of making test samples, solder joints and carrying out reliability tests

Before carrying out reliability tests, it was necessary to design and manufacture appropriate test samples in the form of a solder joints between the solder and test pin. The figure 3 presents the following stages of preparing test samples:

- test samples in the form of copper squares etched on the glass-reinforced epoxy laminate FR4 and covered with a thin layer of gold,
- 2 solder balls used in BGA technology were placed on the surfaces covered with a thin layer of flux,
- test samples were heated up to the temperature exceeding the melting point of a solder alloy.

After setting up appropriate test samples and placing them on the sample holder, it was then possible to make solder joints. The figure 4 presents the following stages of making solder joints between the test sample and copper pin:

- stage I, after applying a small amount of flux to the test needle, a reference point and depth of sinking the pin was set - half of the average sample height,
- stage II, the pin was heated up and sink into the test sample within the so-called flow time, in order to make a solder joint,
- stage III, the prepared test sample with solder joint was slowly cooled down in order to avoid the formation of undesirable residual stresses.

In the first place the strength tests were made, which consisted of pulling the pin with a given force and speed until the connection was destroyed. In this way, the range of forces and velocities of the test pin were determined for the following strength tests. All the tests were performed at room temperature. Figure 5a presents the results of the performed tests. It can be concluded from the achieved results, that

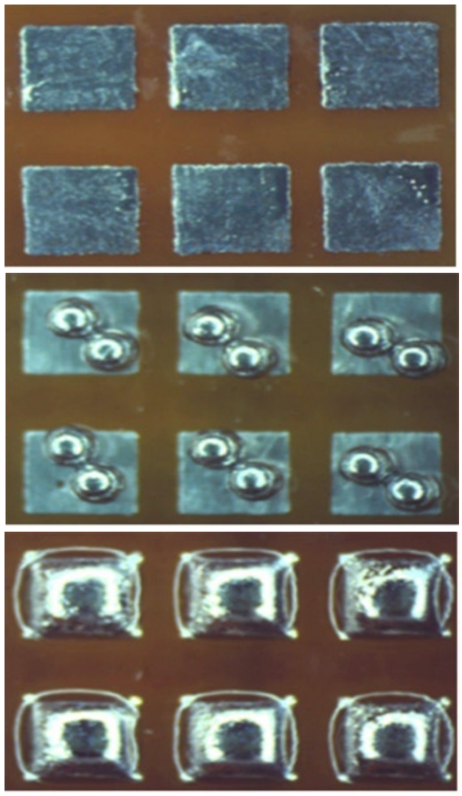


Fig. 3. The following stages of preparing the test samples

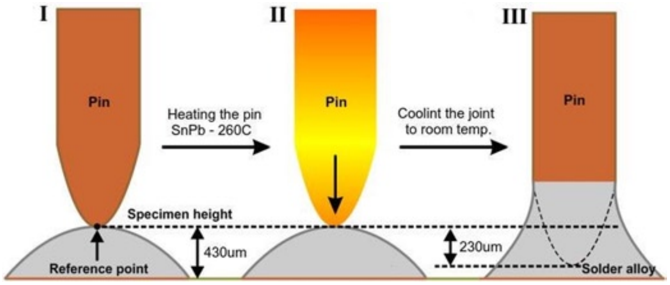
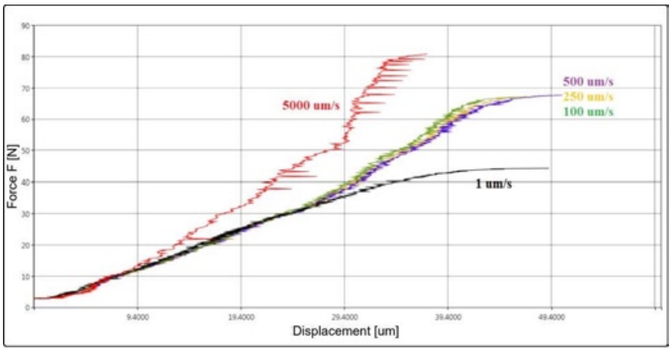


Fig. 4. The following stages of making the solder joint between the copper pin and test sample: (I) setting the reference point and sinking depth, (II) sinking the pin, (III) cooling down the test sample

the high speed of the pin movement, of the order of *5000μm/s*, leads to higher force values and thus cause the early joint failure, which in reverse means that the solder alloy behaves as a brittle material. On the other hand, the low speed of the pin movement, of the order of *1μm/s*, leads to creep phenomena, which is undesirable in the case of pure fatigue tests. Because of this, in order to run reliability tests, it was assumed to select the speed of the pin movement to *500μm/s*, which on one hand allowed avoiding the hardening phenomenon of the material due to the high speed pin movement but on the other hand allowed to reduce the overall testing time. The reduction of testing time is very crucial in research that requires statistical analysis, which is associated with the need to carry out many single tests.

In the second place the strength tests were done with the defined loading profile and test parameters, which is presented in the figure 5b. The profile consisted of two parts responsible for the both fatigue and creep phenomena. The creep part depended on dwell time Δt_i, which varied in the range of *1* to *80s*. The force amplitude varied in the range of *18* to *63N*. Additionally, an initial force value *F* was assumed *3N*.

a)



b)

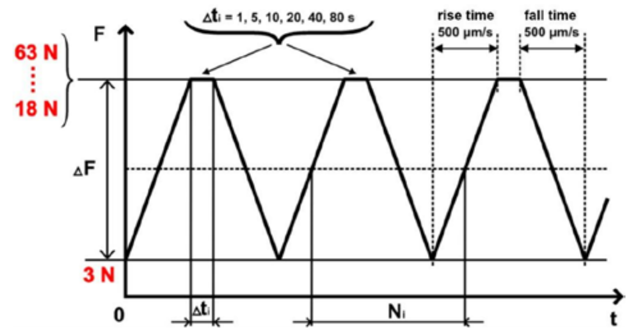
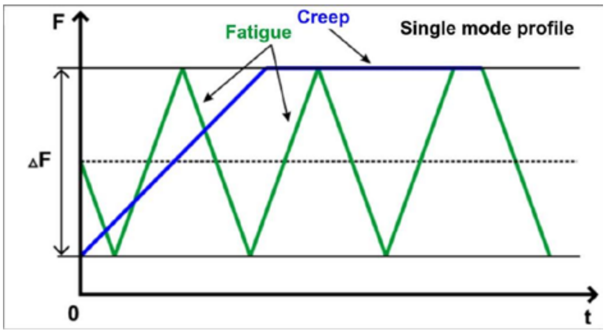


Fig. 5. Results of the reliability tests for different speed of the test pin movement (a) and the assumed profile of the loading for the reliability tests (b)

a)



b)

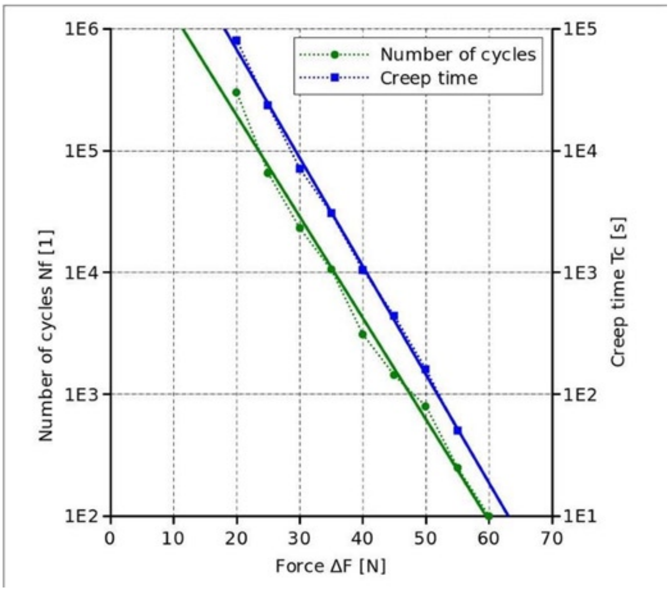


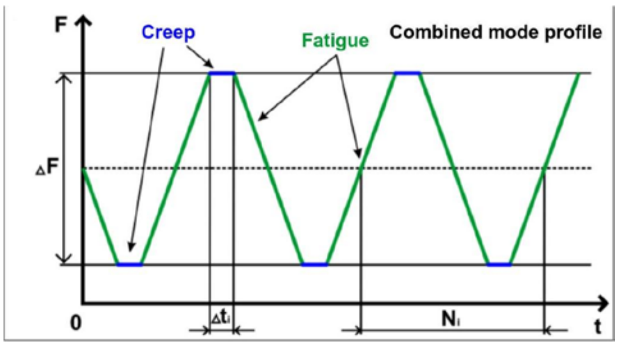
Fig. 6. Loading profile (a) and test results for a single failure mode, i.e. creep and fatigue (b)

3.2. Reliability tests

During the first stage of the research, strength tests were performed for a single selected failure mode, i.e. in succession for fatigue and creep. The tests allowed assessing the maximal number of fatigue cycles N and creep time t leading to the solder joint failure at a given force amplitude ΔF. The assumed combined loading profile and achieved results are given in the figure 6.

According to the achieved results, it can be concluded that both modes of failure, i.e. fatigue and creep, depend exponentially on the force amplitude ΔF. The results allowed estimation of test parameters for the creep-fatigue failure interaction, referring for dwell time values Δt_i. Figure 7a presents the combined load profile and the corresponding results for strength tests.

a)



b)

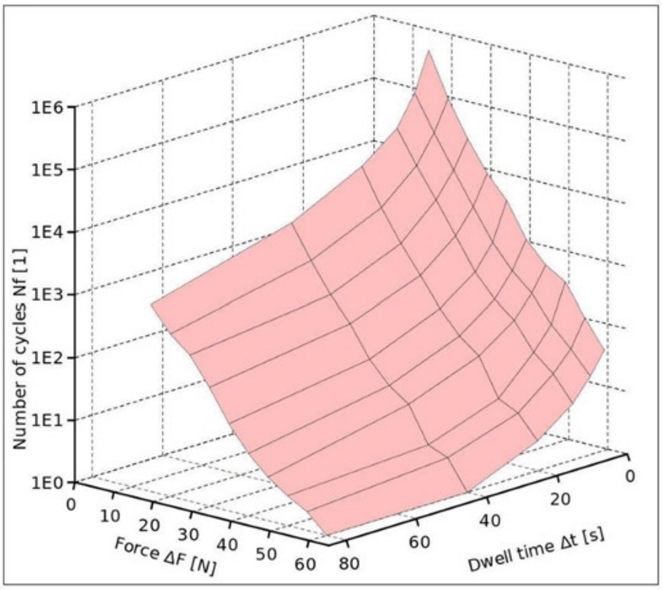


Fig. 7. Combined loading profile (a) and test results for the creep-fatigue failure interaction (b)

According to the achieved results (Fig. 7b), it can be concluded that the number of fatigue cycles to failure in the case of creep-fatigue tests depend not only on the force amplitude ΔF, but additionally on the interaction between both failure modes, which are creep and fatigue. It can be noticed that the creep participation reduces significantly number of cycles to failure N_f.

3.3. Statistical analysis.

Reliability analysis can be done with the help of random variables and thus, as it was mentioned earlier, requires statistical methods. For this purpose, two-parameter Weibull distribution is the most often used. Unfortunately, detailed analysis of the Weibull statistics requires

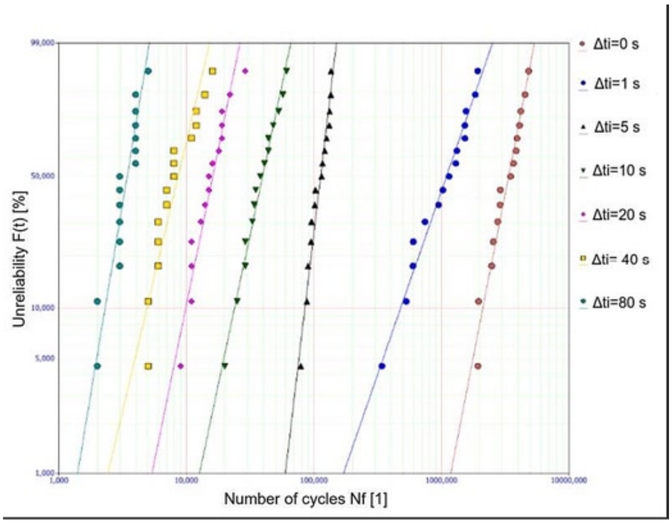
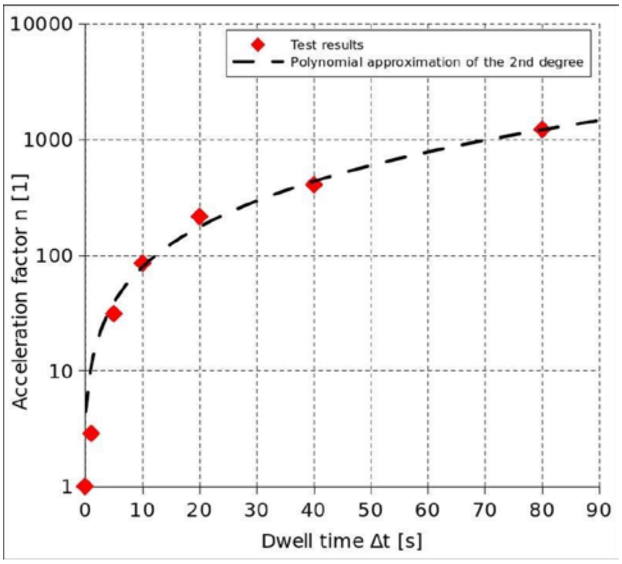


Fig. 8. Weibull statistical analysis for the performed reliability tests in the case of a single and combined failure mode for the force amplitude ΔF equal 40 N

a)



b)

	Ftigue - cycles [1]	Dwel time Δti [s]						Creep-time [s]
		1	5	10	20	40	80	
β	4,05	2,28	6,64	3,72	3,86	3,33	4,72	3,16
λ	3673	1283	118	43	17	9	3	271

Fig. 9. Results of the statistical analysis for the force amplitude ΔF equal 40 N: a) coefficients of the Weibull distribution i.e. the shape β and scale factor λ, b) values of the failure acceleration coefficient n.

many single tests. Figure 8 shows the test results for a selected value of the force amplitude ΔF equal *40 N* – the total number of single tests in the presented case was higher than *100* [26].

Based on the statistical analysis results, the coefficients of two-parameter Weibull distribution were determined, i.e. the shape β and scale factor λ. Results of the analysis are given in table 9a. The failure acceleration coefficient n, describing the reduction of a number of fatigue cycles N_f leading to failure as a result of dwell time Δt_i, was also estimated. The results are shown in the figure 9b.

3.4. Model of damage accumulation

As it was mentioned earlier, the research goal was reliability analysis with the use of creep-fatigue tests for solder joints in microscale, which are typical in microelectronics packaging. Figure 10 presents the damage model accumulation achieved on the basis of performed experimental tests and the following statistical analysis.

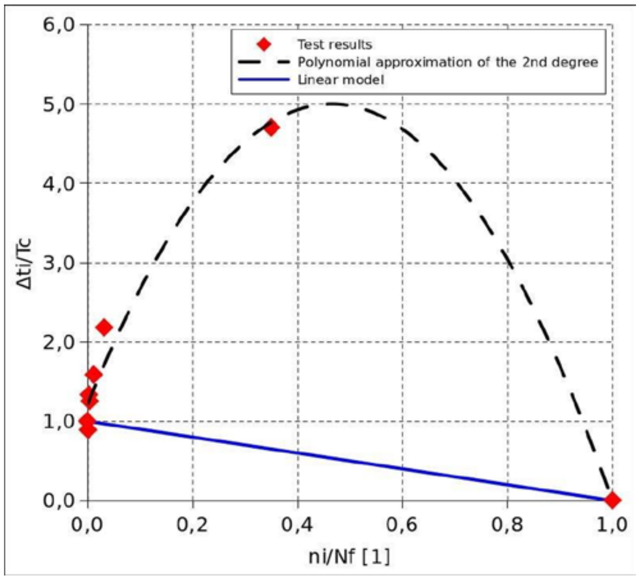


Fig. 10. Damage accumulation model achieved on the basis of the performed experimental tests and statistical analysis

Basing on the achieved results it can be concluded that solder alloy $Sn_{63}Pb_{37}$ analyzed with a combined loading profile in order to induce the failure interaction manifests cyclic hardening phenomenon, which leads to reduction of the joint strength. The material with every fatigue cycle becomes more brittle and thus susceptible to the formation of cracks that finally cause the solder joint failure. According to the result presented in the figure 10, it can be stated that the obtained damage accumulation model reveals non-linear behavior and significantly deviates from a theoretical curve for the linear model.

4. Summary

The manuscript contains results for the first stage research concerning strength analysis of solder joints used in microelectronics packaging due to a combined loading profile. The proposed loading profile allowed for analysis of typical failures modes of solder alloys, which is creep and fatigue and additionally interaction of the both. Within the performed research a combined creep-fatigue test was introduced based on the Hot Bump Pull HBP measuring technique developed by Nordson Dage company. The presented results contain strength analysis of solder joints for a traditional $Sn_{63}Pb_{37}$ solder alloy, done in room temperature, as follows:

- strength analysis of a single failure modes and their interaction,
- statistical analysis for a combined loading profile,
- comparison of the linear model of the damage accumulation with the model achieved on the basis of performed experimental strength tests.

The final conclusion, based on the presented results, can be formulated as follows:

- the proposed methodology of testing solder alloy strength makes it possible to take into account the phenomenon of failure mode interaction, i.e. creep and fatigue,

- the creep phenomenon for the tested solder alloy in the microscale plays an important role for the creep-fatigue tests even at room temperature, at which the strength tests were carried out,
- a popular among engineers the linear failure accumulation model is not suitable for predicting the strength of solder joints in the microelectronics concerning the combined creep-fatigue tests.

It should be underlined, that analysis of a single failure mode leads to wrong predictions, because the real working conditions of electronic components are characterized by existence of a couple of failure modes, which is due to the combined environmental loading profile. Taking into account the above, the next research should contain such problems as:

- application of a different damage accumulation model in case of solder joints in microelectronics for the combined creep-fatigue tests, i.e. one of non-linear models or relatively simple model for use in engineering practice refereed as double linear model [18],
- running comparative tests for lead-free solder alloys and for other ambient temperatures,
- developing a strength criterion, which would allow reliable prediction of the solder joints strength in microelectronics packaging as a result of fatigue and creep failure mode due to the combined loading profile.

In conclusion, it can be stated that the proposed methodology concerning reliability tests and the presented results of strength analysis of solder joints in microscale would allow for a more precise reliability prediction of the modern electronic components and / or an implementation of the accelerated reliability tests.

References

1. Arra M, Shangguan D, Ristolainen E. Characterization of mechanical performance of Sn/Ag/Cu solder joints with different component lead coatings. Soldering and Surface Mount Technology 2004; 16(1): 35-43, https://doi.org/10.1108/09540910410517031.
2. Bisschop J. Failure mechanisms in plastic package IC's. Proceedings of 2nd EuroSimE Conference 2002: 328-331.
3. Bukat K, Hackiewicz H. Lutowanie bezołowiowe, Warszawa: Wydawnictwo BTC, 2007.
4. Chicot D, Tilkin K, Jankowski K, Wymysłowski A. Reliability analysis of solder joints due to creep and fatigue in microelectronic packaging using microindentation technique, Microelectronics Reliability 2013; 53(5): 761-766, https://doi.org/10.1016/j.microrel.2013.01.008.
5. DAGE UK Hot Bump Pull/Hot Pin Pull, 2011.
6. Dasgupta A, Oyan C, Barker D, Pecht M. Solder Creep-Fatigue Analyses by an Energy Partitioning, Journal of Electronic Packaging 1991; 114(2): 152-160, https://doi.org/10.1115/1.2906412.
7. Dowling NE, Kampe SL, Kral MV. Mechanical Behavior of Materials, Pearson Prentice Hall, 2018.
8. Felba J. Montaż w Elektronice, Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 2010.
9. Ganesan S. Lead-free Electronics, John Wiley & Sons Inc., 2006, https://doi.org/10.1002/047000780X.
10. Holdsworth SR. Component assessment data requirements from creep-fatigue tests, ASTM International 2011; 1-14, https://doi.org/10.1520/JAI103583.
11. Jankowski K, Wymysłowski A. Life prediction of lead alloy based on multi-failure criteria. 18th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems 2017; 1-5, https://doi.org/10.1109/EuroSimE.2017.7926217.
12. Jankowski K, Wymysłowski A, Chicot D. Combined loading and failure analysis of lead-free solder joints due to creep and fatigue phenomena, Soldering & Surface Mount Technology 2014; 26(1): 1-6, https://doi.org/10.1108/SSMT-10-2013-0029.
13. Judd M, Brindley K. Soldering in Electronics Assembly, Elsevier Ltd., 1999, https://doi.org/10.1016/B978-075063545-5/50004-2.
14. Kim DH. Reliability Study of SnPb and SnAg Solder Joints in PBGA Packages, The University of Texas at Austin, 2007.
15. Kisiel R. Podstawy technologii montażu dla elektroników, Wydawnictwo BTC, 2012.
16. Lau JH. Solder Joint Reliability, Boston MA: Springer, 1991, https://doi.org/10.1007/978-1-4615-3910-0.
17. Lau JH. Thermal Stress and Strain in Microelectronic Packaging, US Springer, 1993, https://doi.org/10.1007/978-1-4684-7767-2.
18. Manson SS, Halford GR. Practical implementation of the double linear damage rule and damage curve approach for treating cumulative fatigue damage, International Journal of Fracture 1981; 17: 169-192, https://doi.org/10.1007/BF00053519.
19. Matkowski PK, Dudek R, Kreyssig K. Reliability testing of SnAgCu solder joints under combined loading. Micromaterials and Nanomaterials 2011; 13: 86-91.
20. Miller D, Priest RH, Ellison EG. A review of material response and life prediction techniques, Materials 2015; 8(11): 7757-7769, https://doi.org/10.3390/ma8115418.
21. Miner MA. Cumulative Damage in Fatigue, Journal of Applied Mechanics 1945; 12 (3): 159-164.
22. Nelson W. Accelerated Testing, New York: Wiley, 1990, https://doi.org/10.1002/9780470316795.
23. Nguyen J, et al. Backward compatibility study of lead free area array packages with tin-lead soldering process, IPC - Printed Circuits Expo and the Designers Summit 2006.
24. Numerical Analysis Guidelines for Microelectronics Packaging Design and Reliability, IPC/JEDEC-9301: 2018.
25. Qu J. Thermomechanical Reliability of Microelectronic Packaging. Comprehensive Structural Integrity 2007; 8: 219-239, https://doi.org/10.1016/B0-08-043749-4/08040-X.
26. ReliaSoft, https://www.reliasoft.com/, 2019.
27. Seraphim DP, Lasky R, Li CY. Principles of Electronic Packaging, New York: McGraw-Hill, 1989, https://doi.org/10.1115/1.3226524.
28. Stone D, Wilson R, Li CY. The Mechanisms of Damage Accumulation in Solders during Thermal Fatigue. Proccedings of 36th Electronic Components Conference 1986: 630-635.
29. Suhir E. Interfacial stresses in biomaterial thermostats, Journal of Applied Mechanics 1989; 56: 595-600, https://doi.org/10.1115/1.3176133.
30. Timoshenko S. Analysis of bi-metal thermostats, Journal of the Optical Society of America 1925: 233-255, https://doi.org/10.1364/JOSA.11.000233.
31. Tummala RR. Fundamentals of Microsystems Packaging, McGraw Hill, 2001.
32. Tummala RR, Rymaszewski EJ, Klopfenstein AG. Microelectronics Packaging Handbook, Springer Science+Business Media Dordrecht,

1997, https://doi.org/10.1007/978-1-4615-6037-1.
33. Van Driel WD. Virtual Thermo-Mechanical Prototyping of Microelectronics Devices, The Netherlands: PrintPartners Ipskamp, 2007
34. Wareing J. Mechanisms of high temperature fatigue and creep-fatigue failure in engineering materials, United Kingdom: Applied Science, 1983.
35. Wymysłowski A. Numeryczne metody projektowania termomechanicznego w montażu elektronicznym. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 2007.
36. Wymysłowski A, Dowhań Ł. Application of nanoindentation technique for investigation of elasto-plastic properties of the selected thin film materials, Microelectronics Reliability 2013; 53(3): 443-451, https://doi.org/10.1016/j.microrel.2012.10.009.
37. Zahavi E. Fatigue design - life expectancy of Machine Parts, A Solomon Press Book, 1996.
38. Zhang GQ. The challenges of virtual prototyping and qualification for future microelectronics, Microelectronics Reliability 2003; 43: 1777-1785, https://doi.org/10.1016/S0026-2714(03)00299-3.
39. Zhang GQ, Ernst LJ, de Saint Leger O. Benefiting from thermal and mechanical simulation in microelectronics, Dordrecht: Kluwer Academic Publishers, 2000, https://doi.org/10.1007/978-1-4757-3159-0.
40. Zhang GQ, van Driel WD, Fan XJ. Mechanics of Microelectronics, Dordrecht: Springer, 2006, https://doi.org/10.1007/1-4020-4935-8
41. Zienkiewicz OC, Taylor R, Zhu JZ. The Finite Element Method: Its Basis and Fundamentals, Elsevier Butterworth-Heinemann, 2005.

Artur WYMYSŁOWSKI
Krystian JANKOWSKI
Wroclaw University of Science and Technology
Faculty of Microsystem Electronics and Photonics
ul. Janiszewskiego 11/17, 50-372 Wrocław, Poland

E-mails: artur.wymyslowski@pwr.edu.pl,
krystian.jankowski@pwr.edu.pl

Article citation info:
PAULAUSKAS V, FILINA-DAWIDOWICZ L, PAULAUSKAS D. Ships speed limitations for reliable maintenance of the quay walls of navigation channels in ports. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 306–315, http://dx.doi.org/10.17531/ein.2020.2.13.

Vytautas PAULAUSKAS
Ludmiła FILINA-DAWIDOWICZ
Donatas PAULAUSKAS

SHIPS SPEED LIMITATIONS FOR RELIABLE MAINTENANCE OF THE QUAY WALLS OF NAVIGATION CHANNELS IN PORTS

OGRANICZENIA PRĘDKOŚCI STATKÓW DO NIEZAWODNEGO UTRZYMANIA ŚCIAN NABRZEŻY ZLOKALIZOWANYCH PRZY KANAŁACH NAWIGACYJNYCH W PORTACH

There is a number of ports where approach or inside navigation channels are located close to the quay walls. In difficult hydro-meteorological conditions appropriate speed of ship is needed to keep proper ship's steering while passing through channel. The ships that pass near the quay walls with high speed create high interaction forces on moored ships and negatively interact on their mooring equipment and quay walls. Ports should ensure relevant maintenance and reliability of quay walls and ships' mooring equipment. That is why investigation of the ships interaction forces during ship passing near the ships moored to quay walls is very important to find limitations of the passing ship speed depending on passing ship's parameters, distances and environmental conditions to provide reliable maintenance of navigation channel. In the article the conditions of dynamic forces caused by passing ships are investigated, including possible external forces influencing on moored ships, mooring equipment and quay walls. The methodology to assess the forces exerted on ship moored to quay wall by ship passing close to them is created. On the basis of the case study analysis results, the recommendations for the limitations to ships passing near the ships moored to quay walls were proposed that will allow providing relevant maintenance and reliability of navigation channel and its infrastructure.

Keywords: *dynamic forces, navigation safety in ports, quay walls maintenance and reliability, port infrastructure, ship mooring, quay walls mooring equipment.*

Istnieje wiele portów, w których tory podejściowe lub wewnętrzne kanały nawigacyjne znajdują się obok ścian nabrzeży. W trudnych warunkach hydrometeorologicznych należy utrzymywać odpowiednią prędkość statku, aby zapewnić jego prawidłowe sterowanie podczas przemieszczania się przez kanał. Statki, które przepływają z dużą prędkością w pobliżu ścian nabrzeży, wywierają duże siły na zacumowane jednostki i negatywnie oddziałują na urządzenia cumownicze statków i ściany nabrzeży. Porty powinny zapewniać należyte utrzymanie oraz niezawodność ścian nabrzeży i urządzeń cumowniczych statków cumowanych przy nabrzeżach. W związku z tym ważne jest zbadanie sił interakcji podczas przemieszczania się statku w pobliżu jednostek przycumowanych przy nabrzeżu, aby znaleźć ograniczenia prędkości przepływającego statku w zależności od jego parametrów, odległości i warunków środowiskowych, co pozwoli zapewnić niezawodną eksploatację kanału nawigacyjnego. W artykule zbadane są uwarunkowania sił dynamicznych wywieranych przez przepływające statki, w tym możliwe siły zewnętrzne, które wpływają na zacumowane jednostki, urządzenia cumownicze i ściany nabrzeży. Opracowano metodologię oszacowania sił wywieranych na statek przycumowany przy nabrzeżu przez przepływającą obok jednostkę. Na podstawie wyników analizy studium przypadku zaproponowano zalecenia dotyczące ograniczeń w odniesieniu do statków przepływających w pobliżu statków przycumowanych przy nabrzeżu, które pozwolą zapewnić należytą eksploatację i niezawodność kanału nawigacyjnego i jego infrastruktury.

Słowa kluczowe: *siły dynamiczne, bezpieczeństwo żeglugi w portach, eksploatacja i niezawodność ścian nabrzeży, infrastruktura portowa, cumowanie statku, urządzenia cumownicze ścian nabrzeży.*

1. Introduction

There are many ports in the world where approach and inside navigation channels (fairways) are located very close to quay walls where small and large ships may be moored [1]. Hydrodynamic interaction between the passing ships and ships moored to quay wall influence the moored ships, together with additional external forces that are caused by wind, waves, shallow water effect, inertia forces, created by movements of the moored ship and currents [2, 3, 5, 18, 39]. That is why it is important to provide relevant maintenance and reliability of the quay walls of navigation channel and to conduct calculations of generated external forces influence on quay wall and its equipment (quay wall bollards and fenders), as well as on moored ships mooring equipment (winches, bollards, mooring ropes, etc.) [4, 16].

Ships passing with high speed near the vessels moored to quay walls create high hydrodynamic interaction forces between ships and have negative influence on moored ships mooring equipment and quay walls [26, 41, 43]. Examination of the ship's interaction forces during ship passing near moored vessels is necessary in order to find limitation of the passing ship's speed that may depend on the passing ship's parameters, distances, hydro-meteorological and hydrological conditions [30, 36, 43]. In ports located at the mouth of the rivers or close to the seashore (Fig. 1), such as Southampton (UK), Livorno (Italy), Klaipeda (Lithuania) and others, the precaution measures must be taken to minimize the possible negative influence of the passing ships on moored ships and quay walls to avoid their damages [1, 28, 37, 40, 42].

In some ports passing ships speed limitations are introduced. Such restrictions depend mainly on ships size. It should be mentioned that this limitation approach is rather risky, because ships with different size create different hydrodynamic interaction forces on moored ships and do not always provide relevant maintenance of navigation channel infrastructure. The main solution used in ports to decrease the risk of the moored vessels and quay walls damages - is to reduce the passing ships speed. However, at the same time very low speed of these ships decreases their manoeuvrability, reliability of vessels and quay walls and finally increases the risk of the passing ships navigation and probability of collision with moored vessels or quay wall [3, 4, 8, 11].

Fig. 1. The examples of ports' where ships are mooring very close to navigation channel, where special attention should be given to relevant maintenance and reliability of the ships and quay walls [1]: a) Southampton port (UK), b) Livorno port (Italy), c) Klaipeda port (Lithuania)

For example, in Klaipeda port the regulations were set for Ro-Ro vessels, and other ships, stating that during passing near the ships that are moored to quay wall they had to decrease the speed up to 6 knots. Such low speed in worse weather conditions (strong wind, current and high waves) was not enough to steer the ship safely. Moreover, there were incidents when Ro-Ro vessels rubbed against tankers moored to the oil terminal jetties. After the occurrence of a few accidents the decision was taken to increase permitted speed of passing ships in port up to 8 knots. The implementation of decision created new problems, because this speed did not take into account moored and passing ships parameters, as well as environmental conditions, such as wind velocity and its direction, currents speed, wave's parameters, shallow water influence etc. [3, 17, 26, 41, 42].

The conducted analysis of available literature sources revealed that there are a lot of studies and research articles describing ships interaction in open and restricted waters. There is a number of research works and articles showing the ways to assess the critical areas for a ship facing collision encounter, risk assessment approaches based on ships safety patterns, especially during LNG and other terminals studies and design. However, most of those sources are concentrated mainly on anti-collision tasks and do not take into account ship to ship complex influences occurring, for example, in narrow and shallow channels and acting on moored ships hydro-meteorological and hydrological conditions. Moreover, available calculation and evaluation methods is usually hard to apply for solving practical tasks, such as evaluation of influences of passing ships on the vessels moored to quay walls close to navigation channels and guaranty reliability of the moored ships and quay walls. The main findings of available literature analysis can be formulated as follows:

- there is a lack of complexity in available approaches used to analyse interaction between passing ships on vessels moored to quay walls close to narrow channels, because inertia forces created by moored ship movement alongside quay wall are not taken into account,
- there is a lack of clear and complex methods allowing to calculate the forces or energies distribution between fenders and mooring ropes of the moored ship and quay wall, as well as mooring ropes pretension forces.

The identified gap motivated the authors to undertake the research on moored ships and quay wall safety and reliability in situation when the forces created by ships passing very close to moored ones occur together with other external forces like wind, waves, current, shallow water effect. It also justifies the need to develop the complex calculation method that could be used in practice to assess the influence of passing ships on moored ones and quay walls.

The article aims to investigate the ship to ship interaction in mentioned conditions, develop methodology to assess the forces exerted on ship moored to quay wall by ship passing close to them and find the solution to minimize the risk of damage to ships moored to quay walls when other ships are passing close to them. Moreover, it aims to find optimal passing ship speed considering different environmental conditions (e.g. hydro-meteorological and hydrological) that will allow increasing the reliability of navigation channel operation.

2. State of the art analysis

The situation when ships are moored to the quay wall very close to the narrow navigation channel is observed in different ports [1, 2, 3, 42, 43]. Such situations may take place when port is located at the mouth of the river and has narrow fairway, as well as when it is situated near the seashore, but has limited infrastructure. Port location and layout may influence the volume of investments needed to widen channels. Such investments could be very expensive, time consuming or even not feasible due to the ground and other conditions [12]. In such cases the ships movement should be planned and organized in a safe and efficient way using existing infrastructure.

Based on observations made it could be stated that ships may be moored to one or both sides of the navigation channel. For example, in Livorno port in Italy (Fig. 2) ships are moored to both sides of the channel and large vessels are passing very close to the moored ones, sometimes even on the distance less than width of the passing ship [1].

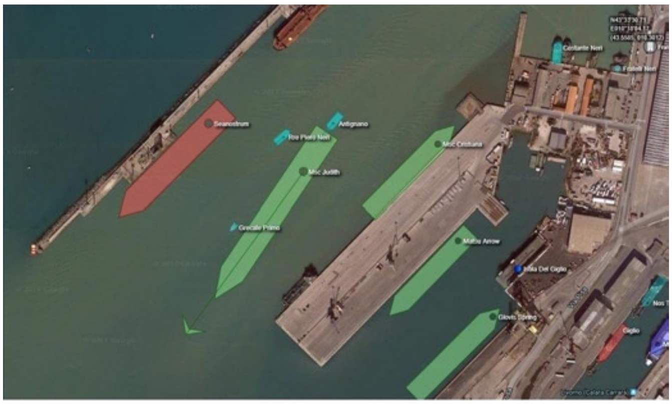


Fig. 2. Livorno port with ships moored to both sides of navigation channel [1]

In some ports the passing ship should make manoeuvres, e.g. turn to large angle (Fig. 3), due to infrastructure limitations, and pass very close to the moored ships. It is possible to perform such operation just with tugs assistance. At the same time, it happens that passing ships cannot decrease speed, because external forces caused by wind, current and waves, as well additional shallow water influence the decrease of the passing ship manoeuvrability [3, 18, 22, 29, 36, 42]. In these cases the probability that passing ship will move out of navigation channel increases and internal possibilities, like thrusters, or external assistance, like tugs, sometime are not able to keep passing large ships in channels and provide relevant reliability level.

In some ports the specific terminals, like container terminal in Bremerhaven port, are located very close to the navigation channel through which MEGA container vessels (with length about 400 m) pass very close to the other moored ships during their mooring and unmooring operations (Fig. 4). Very high hydrodynamic interaction forces created by passing MEGA container vessels have negative influence on moored ships and quay walls mooring equipment [1, 2, 5, 28]. Sometimes large passing ships create very high propeller screw flow that has negative impact on moored ships as well [7, 38, 42].

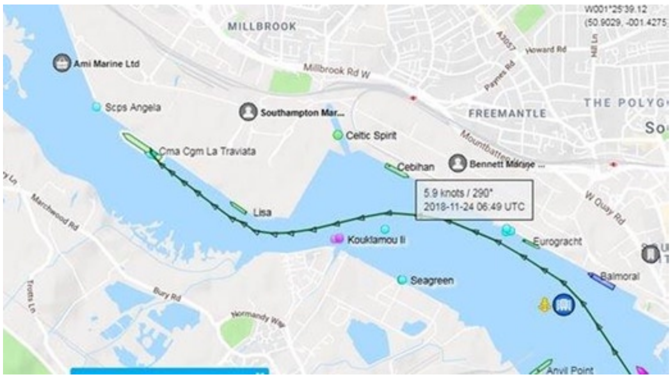


Fig. 3. Container vessel "CMA CGM La Traviata" (L= 334 m, B = 43 m, T = 13 m) passing through narrow channel close to the moored ships in Southampton port (UK) [1]

The analysis of the available literature revealed that currently a lot of attention is paid to the aspects of traffic and navigation safety. Researches investigate i.a. interaction between the elements of transport systems [29], the impact of traffic conditions on transport safety [23], ship manoeuvring [33], navigation safety [9, 24, 42], as well as interactions between ships [34, 36] in navigation channels etc. Ships accidents are analysed in order to find the ways to prevent their occurrence [8, 13, 15, 21, 22]. These accidents are generally collision situations that may take place in different sailing areas. However, mentioned research works do not take into account ships mooring to

Fig. 4. Container vessel "Maribo Maersk" (L = 399,2 m, B = 59 m, T = 13,8 m) in Bremerhaven port (Germany) [1]

quay walls and quay walls reliability in case of different forces are acting simultaneously.

In order to prevent extraordinary situations appearance risk assessment methods are proposed [10, 11, 30, 32, 37, 38]. These methods aim to use mathematical models to evaluate risk level and possible losses of its influence. Separate literature positions analyse the risk-based target reliability indices for quay walls [19, 20, 31] that deals with their maintenance in ports and influence the quays design models [12, 16]. It should be mentioned that risk of quay walls damage influenced by different forces on quay walls and moored ships was not presented sufficiently.

Moreover, ship to ship interaction problems are considered in the literature. The available positions describe the selected aspects of ships interaction in selected conditions. Kadri and Weihs [14] modelled hydrodynamic interactions between two slender bodies of revolution moving in close proximity, in an unbounded, inviscid, and incompressible fluid. Von Graefe et al. [35] introduced the nonlinear steady flow method that accounts for the nonlinear free-surface conditions, ship wave, and dynamic trim and sinkage. Hydrodynamic interactions between two ships advancing in waves were also considered by Chen and Fang [5].

Yuan et al. [41] analysed interaction between two ships with side-by-side arrangement by using 3-D Rankine source panel code, investigating the influence of the distance between the vessels. Yuan et al. [39, 40] also investigated hydrodynamic interaction between two ships travelling or being stationary in shallow water. Bhautoo et al. [3] considered Port of Brisbane conditions to assess moored vessel interaction induced by passing ships. The developed model included a 2D numerical hydrodynamic and wave model, validated against recorded tidal elevations and wave conditions. Weiss et al. [36] validated the criteria for dimensioning pretension suitable for mooring lines of Post-Panamax class ships docked at a port channel. Hydrodynamic interaction phenomena investigations during the ship overtaking manoeuvre for marine related simulators with the use of CFD methods was analysed by Nikushchenko and Zubova [27], as well as Wnęk et al. [37].

There are several positions that consider ship and bank interaction [6, 26]. Time-domain numerical method based on a three-dimensional potential flow solver was developed by Nam and Park [25] to investigate the passing ship problem with a moored barge alongside quay. Potential flows around the passing ship and the moored barge alongside a quay was solved by using a classical finite element method. Lataire and Vantorre [17] described hydrodynamic interaction between ships in restricted Waterways. In turn, Lee et al. [18] investigated hydrodynamic interaction forces between two large vessels, moving each other in curved narrow channel, this topic was also analyzed by Zalewski and Montewka [42].

Comparing the research undertaken in the presented article to similar research works it should be highlighted that reviewed research articles and other available sources do not contain the complex view on interactions between ships sailing through the navigation channels and moored ships. These positions also do not take into account created movement of moored ships and insufficient pretension of the mooring ropes result in inertia forces of the moored ship. These forces influence is analysed separately. Moreover, forces or energies distribution between fenders of the quay wall and mooring ropes of the moored ship, mooring ropes pretension forces evaluation to avoid or minimize moored ship movement near the quay wall were not explained sufficiently.

The examples presented above, as well as results of conducted literature analysis, confirm the need to develop heuristic methodology that will allow analysing and calculating hydrodynamic interaction of the passing ships influence on ships moored to the quay walls and determine the speed limitations of ships moving along the navigation channel.

3. Hydrodynamic interaction between passing and moored ships together with theoretical basis of possible external forces

3.1. Research methodology

In order to develop methodology, first of all, the available literature analysis was conducted that allowed to review the state of the art in the area of ships maintenance, reliability, hydrodynamics and models used to assess the interactions between the ships etc. Necessary data was collected based on literature sources and observations of ships movement in port area and experimental data received from ports (Fig. 5).

It was stated that methodology used in the research should take into account ships particular geometrical and sailing parameters, hydro-meteorological and hydrological conditions in a certain place. Main ships parameters have to be considered, such as: passing ship's displacement, displacement of ship moored to quay wall, width of a passing ship, draft of passing ship, moored ship air projection on diametric square, moored ship air projection on middle square, area of the moored ship's hull in water, moored ship length between perpendiculars, moored ship average draft, moored ship block coefficient, moored ship width, moored ship's speed near the quay wall (in case if ship moves near the quay wall), mass of the ship moored to quay wall, mooring scheme of the moored ship.

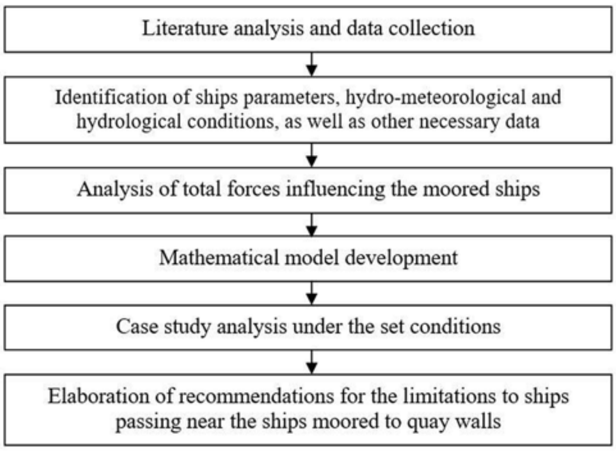


Fig. 5. The algorithm of the research methodology

Hydro-meteorological and hydrological conditions that have to be taken into account in the proposed method are: wind velocity, wind course angle (the angle to a quay wall), current velocity, current course angle to moored ship, waves course angle on moored ship, waves high, speed of movement of water particles in the waves.

Moreover, additional data, such as: navigation channel width (evaluated distance between passing and moored ships), channel depth and depth near the quay wall, necessary for conducting the research should be collected and analysed. Furthermore, the relevant coefficients, received by theoretical and experimental investigations, should be considered.

In order to assess the total forces exerted on ship moored to quay wall by vessels passing close to them and additional external forces acting on moored ship the appropriate separate forces were identified and analysed. Those forces include: dynamic interaction force between passing and moored ships, current created force on moored ship, aerodynamic forces created on moored ship, forces on moored ship created by waves, shallow water influence on moored ship.

Then, the mathematical model was developed to calculate the forces acting on moored ship in case other ship passing near this vessel, as well as wind, current, waves, shallow water and additional inertia forces. This model takes into consideration implementation of following steps:

- collection and analysis of data mentioned above,
- planning possible distances between passing and moored ship,
- calculation of hydrodynamic interaction between ships based on collected needed data, distances between passing and moored ships and passing ship's speed,
- calculation of particular external forces acting on moored ship,
- calculation and analysis of total forces acting on moored ship,
- evaluation of forces or energies distribution between fenders and mooring ropes based on ship's mooring scheme,
- calculations of mooring ropes pretension of the moored ship,
- drawing the conclusions and recommendations for the specific conditions.

Boundary conditions of methodology and model are as follows: minimum distance between moored and passing ships should be not less than 0,25 B_1 in case passing ship's speed is more than 4 knots; maximum distance between passing and moored ships should be less than 3 B_1 , current speed - not more than 4 knots from any direction, wind velocity cannot be more than 18 m/s, waves high cannot be more than 1,5 m. In case of distance between moored and passing ships is more than 3 B_1 , boundary conditions based on possible real conditions in port areas or interaction influence on moored ship do not have to be taken into account because, for example, the hydrodynamic interaction on the moored ship is very low.

The proposed methodology was verified on the basis of case study analysis. The ship movement in Klaipeda port was analysed in detail and calculations based on real data were carried out. On the basis of the archived results recommendations for limitations to ships passing near the vessels moored to quay walls were proposed.

3.2. Mathematical model

The limitations of ships passing near the quay walls should be based on the balance between hydrodynamic interaction between ships and their manoeuvrability in case of external forces occurrence, like wind, currents, waves and shallow water effect. Very often in narrow navigation channels the quay walls influence the possible ships speed, because hydrodynamic interaction take place [17, 23, 31]. In case when current, wind, waves and hydrodynamic interaction forces are acting in one direction, it is very difficult to compensate all this external forces by ship's and quay wall mooring facilities. In general, moored ship mooring lines (ropes) pretension must be equal or bigger then the created external forces on longitudinal and transverse forces.

Total forces that can act on ships and quay wall mooring equipment (bollards and fenders) in the longitudinal direction by passing ship together with other external forces could be calculated as follows:

$$F = F_{inter} + F_c \cdot \cos q_c + F_a \cdot \cos q_a + F_w \cos q_w + F_s\,, \quad (1)$$

where: F - total forces that can act on ships and quay wall mooring equipment; F_i – particular forces: F_{inter} - dynamic interaction force between passing and moored ships; F_c - current created force on moored ship; q_c - current course angle to moored ship; F_a - aerodynamic forces created on moored ship; q_a - wind course angle on moored ship; F_w - forces on moored ship created by waves; q_w - waves course angle on moored ship; F_s - shallow water influence on moored ship.

Hydrodynamic interactions between ships are based on pressure distribution around the sailing ship [16]. Revising the case of impact of ship passing near a vessel moored to a quay wall reveals that the passing ship moves and turns the moored ship (Fig. 6) [11, 16].

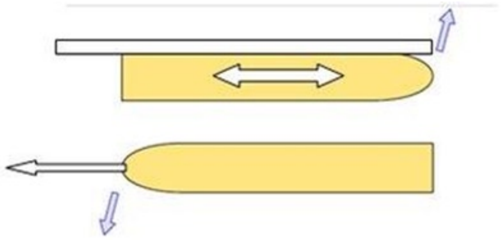


Fig. 6. The influence of passing ship on a moored vessel, where: purple arrow – forces that turn the ship, white arrow – ship's moving direction

The hydrodynamic interaction force (F_{inter}) that acts on moored ship depends on ship's dimensions and speed, distance between the ships and the depths of the channel. Basing on theoretical and experimental investigations, it could be estimated as follows:

$$F_{inter} = k_1 \cdot D_1 \left(1 + \frac{D_1}{D_2}\right) \cdot v_1^2 \cdot \frac{k_2 B_1}{S^2}\left(1 - \frac{H - T_1}{H}\right), \quad (2)$$

where: k_1 - coefficient, for calculations could be taken from 0,10 up to 0.20 depending on the ship's block coefficient, expressed in Fig. 7, based on theoretical and experimental studies [2]; D_1 - passing ship's displacement; D_2 - displacement of ship moored to quay wall; v_1 - passing ship's speed; k_2 - coefficient depending on passing ship's speed, based on theoretical and experimental investigations [2] expressed in Fig. 8; B_1 - width of a passing ship; T_1 - draft of passing ship; S - distance between the ship moored to a quay wall and the passing ship; H - channel depth.

Wind and current acting on a ship moored to a quay wall can be calculated using the established ship theory methods [7, 16, 17]. Aerodynamic (wind) force (F_a) could be calculated as follows:

$$F_a = C_a \cdot \frac{\rho_1}{2} \cdot (S_{2x} \cdot \sin q_a + S_{2y} \cdot \cos q_a) \cdot v_a^2\,, \quad (3)$$

where: C_a - aerodynamic coefficient, which in average is about 1,07 (specific data could be taken from aerodynamic tube testing); ρ_1 - air density, for the calculations could be taken as 1,25 kg/m³; S_{2x} - moored ship air projection on diametric square; S_{2y} - moored ship air projection on middle square; v_a - wind velocity; q_a - wind course angle (the angle to a quay wall).

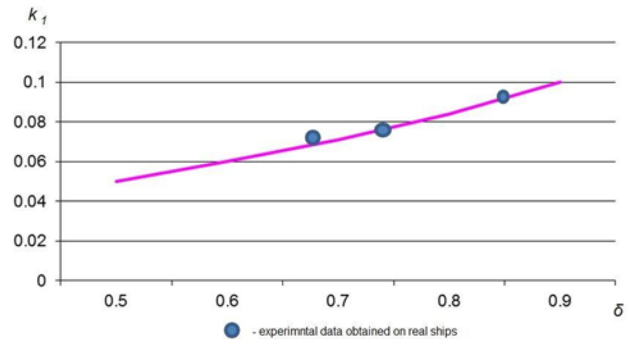


Fig. 7. Dependence of coefficient k_1 on the ship's block coefficient

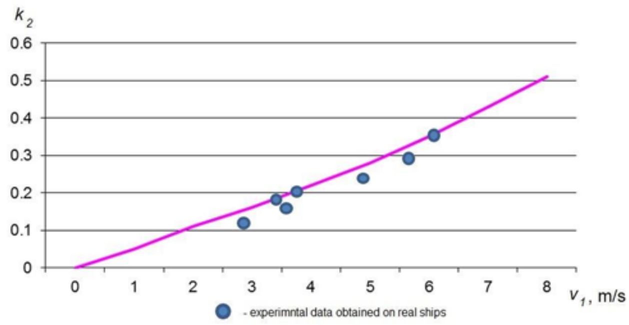


Fig. 8. Dependence of coefficient k_2 on the passing ship's speed

In turn, force caused by current (F_c) could be calculated as follows [2]:

$$F_c = C_y' \cdot \frac{\rho}{2} \cdot \Omega \cdot v_c^2\,, \quad (4)$$

where: C_y' - hydrodynamic coefficient of the ship and in case low current velocity it could be taken in average as 0,15; ρ - water density; v_c - current velocity; Ω - area of the ship's hull in water, can be calculate as follows [22]:

$$\Omega = 1{,}05 L_2 (1{,}7 T_2 + \delta_2 B_2)\,, \quad (5)$$

where: L_2 - moored ship length between perpendiculars; T_2 - moored ship average draft; δ_2 - moored ship block coefficient; B_2 - moored ship width.

Forces on moored ship created by waves (F_w) could be calculated as it is presented below:

$$F_w = C_w \frac{\rho}{2} L_2 \cdot h_w \cdot v_w^2\,, \quad (6)$$

where: C_w - wave acting on ship's hull coefficient, could be taken as hydrodynamic coefficient of the ship and in case of low current (water) velocity it could be taken in average as 0,15; h_w - waves high; v_w - speed of movement of water particles in the waves, could be taken as 0,6 v_a.

Shallow water effect on moored ship (F_s) could be calculated as follows:

$$F_s = k'_{R11} \cdot L_2 \cdot B_2 \cdot \frac{T_2}{H_2} \cdot v_2^2 , \quad (7)$$

where: k'_{R11} - shallow water resistance coefficient, depending on the ratio T/H, based on theoretical and experimental studies [2], expressed in Fig. 9; B_2 - moored ship width; T_2 - moored ship average draft; H_2 - depth near the quay wall; v_2 - moored ship's speed near the quay wall (in case if ship moves near the quay wall).

Finally, total forces acting on a ship moored to a quay wall could be calculated as follows:

$$\vec{F} = \vec{F}_{inter} + \vec{F}_a + \vec{F}_c + \vec{F}_w + \vec{F}_s . \quad (8)$$

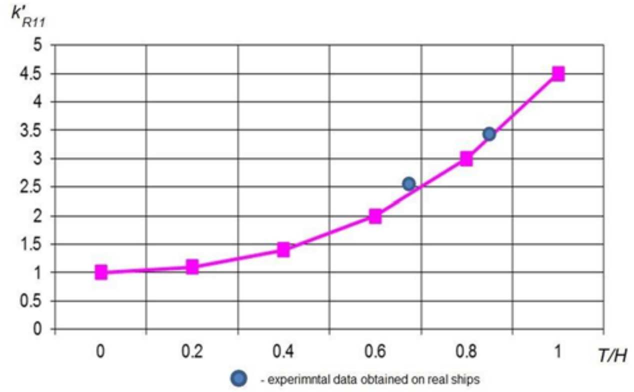


Fig. 9. Dependence of ship's resistence coefficient k'_{R11} on ship's draft and depth T/H

The total force $\vec{F}$ must be absorbed by quay wall fenders and bollards. This force will influence the reliability of the quay walls and moored ships. In case of correct mooring scheme, about 50 % of the total force is absorbed by quay wall fenders and about 50 % should be absorbed by ship's mooring ropes and quay wall bollards [2]. Moored ship will be stable near quay wall just in case if ships mooring ropes pretension will be not less than external forces acting on mooring ropes and quay wall mooring bollards: longitudinal and transverse forces. In case if there is not enough ship mooring rope pretension, the moored ship starts move along the quay wall and creates inertia forces (F_{in}), that sometimes could be higher as any other forces. The inertia forces could be calculated as it is shown below [2]:

$$F_{in} = m'_2 \cdot a , \quad (9)$$

where: m'_2 - the mass of the ship moored to quay wall; a - acceleration that could be calculated as follows:

$$a = \frac{v'_2}{2 \cdot T_P} , \quad (10)$$

where: v'_2 - maximum possible ship movement speed near the quay wall that could be calculated on the basis of ship movement distance near the quay wall and the movement period. The movement distance near the quay wall for the moored to quay wall ship could be taken up to 25 – 30 % of the shortest mooring rope depending on pretension. Ship movement period (T_P) near a quay wall could be equal to $L_1 / 2 - L_1 / 3$.

Finally, inertia forces (F_{in}) could be calculated as it is presented below:

$$F_{in} = k_P \cdot l_r \cdot \frac{m'_2}{2 \cdot T_P^2} , \quad (11)$$

where: k_P - mooring rope pretension coefficient, based on theoretical and experimental studies [2] (Fig. 10), its values could change between 0,0001 and 0,25 (0,25 assumes that there is no mooring rope pretension at all).

Mooring ropes pretension in one direction could be calculated as bow direction that is compensated by astern long mooring ropes and bow springs, as astern direction, that is compensated by bow long mooring ropes and astern springs, and could be calculate as follows:

$$T_{pretb} = \sum_{i=1}^{n_{amr}} T_{amr} \cos\frac{\sum_{i=1}^{n_{amr}} \alpha_{amr}}{n_{amr}} \cdot \sin\frac{\sum_{i=1}^{n_{amr}} \beta_{amr}}{n_{amr}} + \sum_{j=1}^{n_{bsr}} T_{bsr} \cos\frac{\sum_{j=1}^{n_{bsr}} \alpha_{bsr}}{n_{bsr}} \cdot \sin\frac{\sum_{j=1}^{n_{bsr}} \beta_{bsr}}{n_{bsr}} , \quad (12)$$

$$T_{preta} = \sum_{y=1}^{n_{bmr}} T_{bmr} \cos\frac{\sum_{y=1}^{n_{bmr}} \alpha_{bmr}}{n_{bmr}} \cdot \sin\frac{\sum_{y=1}^{n_{bmr}} \beta_{bmr}}{n_{bmr}} + \sum_{z=1}^{n_{asr}} T_{asr} \cos\frac{\sum_{z=1}^{n_{asr}} \alpha_{asr}}{n_{asr}} \cdot \sin\frac{\sum_{z=1}^{n_{asr}} \beta_{asr}}{n_{asr}} , \quad (13)$$

where: T_{pretb} - sum of the pretension forces on ship's bow direction; $\sum_{i=1}^{n_{amr}} T_{amr}$ - total pretension astern long mooring ropes force, $i = 1,...,n_{amr}$; $\sum_{i=1}^{n_{amr}} \alpha_{amr}$ - sum of the horizontal aster long mooring ropes angles, n_{amr} - number of astern long mooring ropes; $\sum_{i=1}^{n_{amr}} \beta_{amr}$ - sum of the vertical astern long mooring ropes angles, $j = 1,...n_{bsr}$; $\sum_{j=1}^{n_{bsr}} T_{bsr}$ - sum of the pretension forces of the bow springs; $\sum_{j=1}^{n_{bsr}} \alpha_{bsr}$ - sum of the horizontal bow springs angles, *j=1, …,* n_{bsr} ; $\sum_{j=1}^{n_{bsr}} \beta_{bsr}$ - sum of the vertical bow springs angles, *j=1, …,* n_{bsr} ; n_{bsr} - number of bow springs; T_{preta} - sum of the pretension forces on ship's astern direction; $\sum_{y=1}^{n_{bmr}} T_{bmr}$ - total pretension bow long mooring ropes force, y=1, …, n_{bmr} ; $\sum_{y=1}^{n_{bmr}} \alpha_{bmr}$ - sum of the horizontal bow long mooring ropes angles, y=1, …, n_{bmr} ; $\sum_{y=1}^{n_{bmr}} \beta_{bmr}$ - sum of the vertical bow long mooring ropes angles, y=1, …, n_{bmr} ; n_{bmr} - number of bow long mooring ropes; $\sum_{z=1}^{n_{asr}} T_{asr}$ - sum of the pretension forces of the astern springs, z=1, …, n_{asr} ; $\sum_{z=1}^{n_{asr}} \alpha_{asr}$ - sum of the horizontal astern springs angles, z=1, …, n_{asr} ; $\sum_{z=1}^{n_{asr}} \beta_{asr}$ - sum of the vertical astern springs angles, z=1, …, n_{asr} ; n_{asr} - number of astern springs.

The theoretical basis presented above could be used for the calculation of the real forces, acting on ship moored to a quay wall close to the navigation channel crossed by other vessels, as well as for design

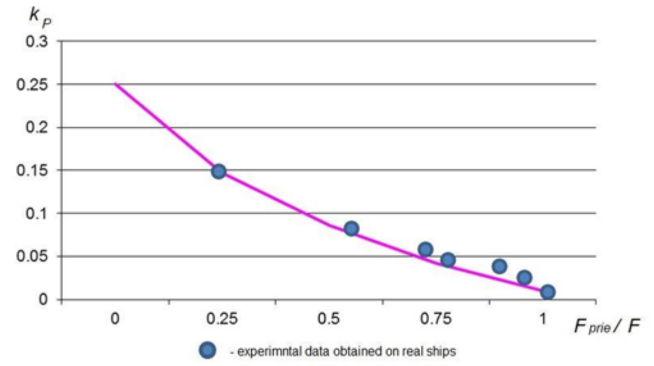


Fig. 10. Dependence of mooring rope pretension coefficient k_P on the mooring rope pretension and total external forces acting on moored ship F_{prie} / F

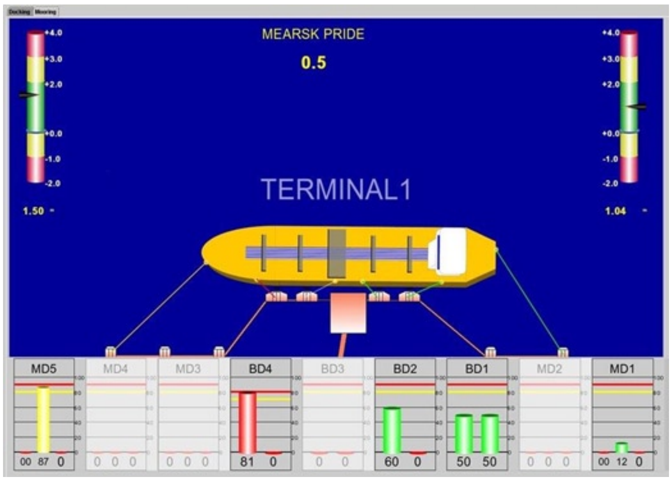


Fig. 11. Laser "Dockmaster 3" measuring system, implemented on quay walls No. 1 and No. 2 in Klaipeda port ring rope pretension and total external forces acting on moored ship F_{prie} / F

of the quay walls and its elements, such as fenders, mooring bollards and others [9, 30].

While conducting calculations it is important to take into consideration the fact that ship mooring ropes pretension depends on the ship parameters and they should be equal to periodical forces, which are created by periodical influence, such as wind, waves, or hydrodynamic interaction of passing ships. In case pretension forces of the mooring ropes are equal to the periodic forces, the increase of the length of mooring ropes could be minimized and inertia forces will decrease or even could be close to 0.

4. Case study of limitations for the passing ship's speed and distance to vessel moored to the quay wall to ensure quay walls maintenance and reliability

In order to investigate the limitations for passing ship's speed and distance between a passing ship and a vessel moored to a quay wall, taking into account external forces caused by wind, current, waves and shallow water effect, a case study of Klaipeda port was considered. Such conditions are typical for the ships moored to quay wall in oil terminal of this port.

Experimental studies were carried out in Klaipeda port during different types of ships (Handy size, PANAMAX, POST PANAMAX tankers and bulk vessels) sailing close to moored vessels. The laser measurement system "Dockmaster 3" was implemented on oil terminal quay walls No.1 and No. 2 (Fig. 11) to measure mooring ropes tension and ships movement. Accuracy of "Dockmaster 3" system while measuring ship's position was about +/- 2-3 mm, ships movement speed was measured with accuracy +/- 0,1 knot (0,05 m/s), ropes tension was assessed with accuracy +/- 0,2 kN. Additionally, during the experiments video cameras were implemented on a quay wall close to a moored ship, which recorded moored ship movement while other vessels passed through navigation channel (Fig. 12). Passing ships movement parameters and distances between the ship moored to the quay wall and passing vessels were measured by Differential GPS and additionally checked by AIS.

Figure 13 presents the situation when Standard LNG tanker (with the capacity of 150000 m³) is passing near the Klaipeda port oil terminal, where Handysize tanker (with the capacity about 30000 t) is moored to the quay wall. The distance between passing and moored to quay wall vessels is usually about 100 m up to 80 m. In turn, POST PANAMAX (DWT 110000) tanker passing near the Klaipeda port oil terminal, where Handysize tanker is moored to the quay wall, is shown in Fig. 14.

Fig. 12. Moored ship movement fixed by video cameras (example)

Fig. 13. Standard LNG tanker passing near the Klaipeda port oil terminal where Handysize tanker is moored to the quay wall

The calculations for the specific case study were conducted under the set assumptions. Moored ship's parameters (Handysize tanker in ballast) are as follows: length (L_2) - 170 m; width (B_2) - 27 m; draft

Fig. 14. POST PANAMAX (DWT 110000) tanker passing near the Klaipeda port oil terminal where Handysize tanker is moored to the quay wall

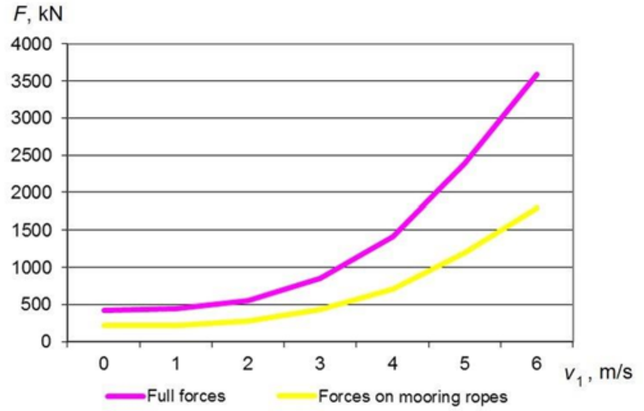


Fig. 15. Dependance of total forces (F) acting on moored vessel and forces acting on mooring ropes on speed (v_1) under the set conditions

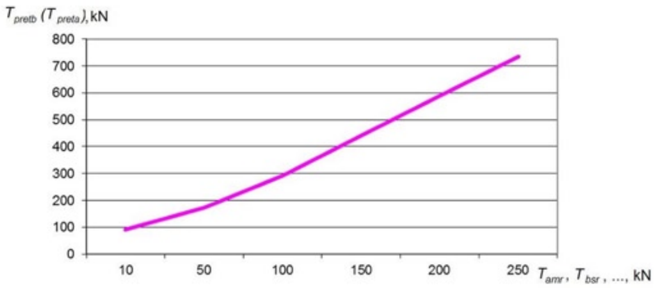


Fig. 16. Dependance of total mooring rope pretension forces in one direction (bow or astern) on each mooring rope pretention

(T_2) - 7,0 m; water displacement of the moored vessel (D_2) - 20800 t, moored ship air projection on diametric square (S_{2x}) - 2600 m²; moored ship air projection on middle square (S_{2y}) - 400 m². For ship's mooring the ropes scheme 3 + 2 + 2 is used, that means there are 3 long mooring ropes on bow and astern, 2 breast mooring ropes and 2 springs. Total bow and astern long mooring ropes horizontal angle ($\sum_{i=1}^{n_{amr}} \alpha_{amr}; \sum_{j=1}^{n_{bsr}} \alpha_{bsr}$) is 15° each, total vertical angle or the bow and astern long mooring ropes angle ($\sum_{i=1}^{n_{amr}} \beta_{amr}; \sum_{y=1}^{n_{bmr}} \beta_{bmr}$) - 80° each, bow and astern springs horizontal angle ($\sum_{j=1}^{n_{bsr}} \alpha_{bsr}; \sum_{z=1}^{n_{asr}} \alpha_{asr}$) - 20° each, vertical angle ($\sum_{j=1}^{n_{bsr}} \beta_{bsr}; \sum_{z=1}^{n_{asr}} \beta_{asr}$) - 70° each. The depths near the quay wall (H_2) is 14 m. Moreover, the external factors influence was also assumed: wind blows 30° to quay wall; wind velocity is 12 m/s; current along quay wall is 0,5 m/s; waves influence - 30° to quay wall, waves height is 1 m.

Passing through the navigation channel Standard LNG tanker has the distance (S) 100 m to moored Handysize tanker. LNG tanker length is 290 m, width (B_1) - 49 m, draft (T_1) – 12 m; displacement (D_1) - 125000 t, block coefficient (δ) - 0,75, ship's speed (v_1) is from 6 up to 8 knots (from 3,1 up to 4,1 m/s), depth in area (H) - 14,5 m.

Total forces acting on moored ship should be divided into forces, that are taken by mooring ropes and resistance forces created between moored ship's hull and fenders. In case of goods pretension mooring ropes, about 50 % of forces are taken by mooring ropes and about 50 % of forces belong to resistance between ship hull and fenders. Considering the mentioned conditions, the total forces (F) that have arisen by influence of wind, waves, current, passing LNG Standard tanker on the distance of 100 m to a moored ship depending on its speed are presented in Fig. 15. Total mooring ropes pretension forces in one direction, depending on pretension of every mooring rope, is presented in Fig. 16.

On the basis of the conducted calculations it was possible to compare and analyse the achieved results. In case LNG Standard tanker is passing with the speed range from 6 knots up to 8 knots (from 3,1 m/s up to 4,1 m/s) on the distance of 100 m from Handysize tanker moored to quay wall, as it was explained in the case study, LNG tanker interaction forces together with external ones (wind, waves, current and shallow water effect) that act on moored tanker mooring ropes, are in the range of 430 kN up to 720 kN. In order to prevent the movement of moored Handysize tanker along the quay wall, moored tanker mooring ropes should be pretensioned from 150 up to 250 kN each to provide relevant maintenence and reliability of the moored ships and quay walls.

5. Conclusion

The organization of safe navigation in the port area should be one of the port's priority tasks. Therefore, complex methods should be applied to assess totality of forces affecting the ships moored to the quay walls of navigation channels while other ships are passing near these vessels. Mooring ropes pretension of the ships moored to quay wall close to navigation channels have to be considered and calculated in order to assure safe maintenance of quay walls and ships. Limitations for ships passing near the quay walls should be established for the particular conditions that may occur in port. Considering these aspects, the relevant methodology and mathematical model were elaborated and the case study was analysed.

It should be highlighted that implementation of methodology presented in this article allows to ensure relevant maintenance of quay walls and ships moored to quay walls in navigation channels in ports. Proposed methodology could be applied to different ports, where wind, current, waves and shallow water effects may take place. The received findings may be interesting and useful for seaports with navigation channels and may allow them to calculate i.a. passing ships speed depending on the distance between ships, conduct moored ships mooring ropes pretension evaluations and other calculations. The research bridges the gap in the area of reliability of the navigation channels operation in ports where ships are moored close to quay walls.

Preliminary calculations carried out using the developed methodology may form the basis for the decision-making about the speed of ships movement in navigation channels where the vessels are moored near the quay walls. Proper preparation of ships mooring to quay walls and ships passing near moored ships can ensure moored vessels and quay walls safety and reliability. Comparing the calculations results under certain conditions with real forces acting on moored ship, good correlation of achieved values may be noted. That shows the possibility to use presented methodology for solving the practical tasks.

Presented methodology has also its drawbacks; it does not take into consideration the wide range of the coefficients used to calculate the forces acting on moored ships. This problem, as well as problem of measuring the individual component forces will be considered in our further research.

Acknowledgement

This article is based on the research conducted within the Interreg SBSR project SB Transport LOOPS co-financed by the European Union from the European Regional Development Fund.

References

1. AIS. Ships automatic identification system, 2019, www.marinetraffic/ais.
2. Barzdziukas R. el at. Shipping engineering (in Lithuanian). Klaipeda University publish house, 2019: 544 p.
3. Bhautoo P, Mortensen S, Hibberd W, Harkin A, Kirkegaard J, Morley B. Moored vessel interaction induced by passing ships at the Port of Brisbane. Australasian Coasts & Ports Conference 2015, 15-18 September 2015, Auckland, New Zealand, pp. 1-9.
4. Biehl F, Lehmann E. Collisions of ships with offshore wind turbines: Calculation and risk evaluation. In: Köller J., Köppel J., Peters W. (eds.) Offshore Wind Energy. Berlin: Springer, 2006: 281-304, https://doi.org/10.1007/978-3-540-34677-7_17.
5. Chen G-R, Fang M-C. Hydrodynamic interactions between two ships advancing in waves. Ocean Engineering 2001; 28(8): 1053-1078, https://doi.org/10.1016/S0029-8018(00)00042-1.
6. Chun-Ki L, Sam-Goo L. Investigation of ship manoeuvring with hydrodynamic effects between ship and bank. Journal of Mechanical Science and Technology 2008; 22(6): 1230-1236, https://doi.org/10.1007/s12206-008-0309-9.
7. EAU 2012. Recommendations of the Committee for Waterfront Structures - Harbours and Waterways. Ernst & Sohn 2012; 620 p.
8. Erol S, Başar E. The analysis of ship accident occurred in Turkish search and rescue area by using decision tree. Maritime policy and management 2015; 42(4): 377-388, https://doi.org/10.1080/03088839.2013.870357.
9. E-Sea Fix navigation system. Denmark: Marimatech, 2012; 120 p.
10. Gill A. Optimization of the technical object maintenance system taking account of risk analysis results. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19 (3): 420-431, https://doi.org/10.17531/ein.2017.3.13.
11. Gucma L, Gucma M, Perkovic M, Vidmar P. Simulation methods for risk assessment in LNG terminal design. In: Rizzuto, Soares, Guedes (eds.). Sustainable Maritime Transportation and Exploitation of Sea Resources. London: Taylor & Francis Group, 2012; 755-761, https://doi.org/10.1201/b11810-113.
12. Gucma S, Gucma M: Optimization of LNG terminal parameters for a wide range of gas tanker sizes: the case of the port of Świnoujście. Archives of Transport 2019; 50(2): 91-100, https://doi.org/10.5604/01.3001.0013.5696.
13. Huang, Y., Chen, L., Chen, P., Negenborn, R.R., van Gelder, P.H.A.J.M. Ship collision avoidance methods: State-of-the-art. Safety Science 2020; 121: 451-473, https://doi.org/10.1016/j.ssci.2019.09.018.
14. Kadri U, Weihs D. Higher order hydrodynamic interaction between two slender bodies in potential flow. Journal of Marine Science and Technology 2015; 20: 249-256, https://doi.org/10.1007/s00773-014-0275-0.
15. Krata P, Montewka J. Assessment of a critical area for a give-way ship in a collision encounter. Archives of Transport 2015; 34(2): 51-60, https://doi.org/10.5604/08669546.1169212.
16. Kuancheng H, Suprayogi, Ariantini. A continuous berth template design model with multiple wharfs. Maritime policy and management 2016; 43(6): 763-775, https://doi.org/10.1080/03088839.2016.1169449.
17. Lataire E, Vantorre M. Hydrodynamic interaction between ships and restricted Waterways. International journal of maritime engineering 2017; 151, https://doi.org/10.3940/rina.ijme.2017.a1.391.
18. Lee C.-K., Moon S.-B., Jeong T.-G. The investigation of ship manoeuvring with hydrodynamic effects between ships in curved narrow channel. International Journal of Naval Architecture and Ocean Engineering 2016; 8: 102-109, https://doi.org/10.1016/j.ijnaoe.2016.01.002.
19. Lendering KT, Jonkman SN, Peters DJ. Risk approach to land reclamation: Feasibility of a polder terminal. Safety, Reliability and Risk Analysis: Beyond the Horizon - Proceedings of the European Safety and Reliability Conference, Amsterdam, The Netherlands, ESREL 2013; 2507-2514, https://doi.org/10.1201/b15938-376.
20. Lendering KT, Jonkman SN, Van Gelder PHAJM, Peters DJ. Risk-based optimization of land reclamation. Reliability Engineering and System Safety 2015; 144, 5380: 193-203, https://doi.org/10.1016/j.ress.2015.07.025.
21. Leveson N. A new accident model for engineering safer systems. Safety Science 2004; 42(4): 237-270, https://doi.org/10.1016/S0925-7535(03)00047-X.
22. Lisowski J. Sensitivity of the game control of ship in collision situations. Polish Maritime Research 2015; 22 No. 4(88): 27-34, https://doi.org/10.1515/pomr-2015-0067.
23. Luty W. Simulation-based analysis of the impact of vehicle mass on stopping distance. Eksploatacja i Niezawodność - Maintenance and Reliability 2018; 20 (2): 182-189, https://doi.org/10.17531/ein.2018.2.03.
24. Mironiuk W. Model-based investigations on dynamic ship heels in relation to maritime transport safety, Archives of Transport 2015; 33(1): 69-80, https://doi.org/10.5604/08669546.1160928.
25. Nam B., Park J. Numerical simulation for a passing ship and a moored barge alongside quay. International Journal of Naval Architecture and Ocean Engineering 2018; 10: 566 - 582, https://doi.org/10.1016/j.ijnaoe.2017.10.008.
26. Nazrul Islam Md, Rafiqul Islam M, Sadiqul Baree Md. Passing ship effects on a moored ship: a numerical study. Proceedings of MARTEC The International Conference on Marine Technology, BUET, Dhaka, Bangladesh, 11-12 December 2010: 201-207.
27. Nikushchenko D V, Zubova A A. Hydrodynamic interaction phenomena investigations during the ship overtaking maneuver for marine related simulators with the use of CFD methods. International Conference on Ship Maneuverability and Maritime Simulation (MARSIM 2015), Newcastle University, United Kingdom, 8-11 September 2015, paper 3-4-3: 13.
28. Pallotta G, Vespe M, Bryan K. Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. Entropy 2013; 15(6): 2218-2245, https://doi.org/10.3390/e15062218.

29. Podvezko V, Sivilevičius H. The use of AHP and rank correlation methods for determining the significance of the interaction between the elements of a transport system having a strong influence on traffic safety. Transport 2013; 28(4): 389-403, https://doi.org/10.3846/16484142.2013.866980.
30. Quy N M, Vrijling J K, Gelder P, Groenveld R. Modeling risk and simulation-based optimization of channel depths at Campha coal port. IASTED Asian conference, Beijing, 8-10 October 2007: 192-198.
31. Roubos A A, Steenbergen R D J M, Schweckendiek T, Jonkman S N. Risk-based target reliability indices for quay walls. Structural Safety 2018, 75: 89-109, https://doi.org/10.1016/j.strusafe.2018.06.005.
32. Semenov JN. Risk management in maritime economy. Volume I. Safety management of transport vessels and oceanotechnical facilities (in Polish). Technical University of Szczecin publishing house, Szczecin, 2003.
33. Skjetne R. Ship maneuvering: The past, the present and the future. Sea Technology 2003; 44(3): 33-37, https://doi.org/10.1177/003932070303300104.
34. Tuck E, Newman J. Hydrodynamic interactions between ships. Symposium on Naval Hydrodynamics, 10th, Proceeding, Pap and Discuss, Cambridge, Mass, June 24-28, 1974, 47 p.
35. Von Graefe A, Shigunov V, el Moctar O. Rankine source method for ship-ship interaction problems. Journal of Offshore Mechanics and Arctic Engineering 2015; 137, https://doi.org/10.1115/1.4029316.
36. Weiss J, de Araujo M, Pereira A, Carmignotto M, Candara M. Hydrodynamic interactions between ships in navigation channels. 11th International Marine Design Conference 2012, At University of Strathclyde, Glasgow, 2012: Proceedings 3: 56-74.
37. Wnęk AD, Sutulo S, Guedes Soares C. CFD analysis of ship-to-ship hydrodynamic interaction. Journal of Marine Science and Application 2018; 17(1): 21-37, https://doi.org/10.1007/s11804-018-0010-z.
38. Yang Z, Adolf KY, Wang NJ. A new risk quantification approach in port facility security assessment. Transportation Research. Part A 2014; 59: 72-90, https://doi.org/10.1016/j.tra.2013.10.025.
39. Yuan Z-M, Incecik A, Dai S, Alexander D, Ji C-Y, Zhang X. Hydrodynamic interaction between two ships travelling or stationary in shallow waters. Ocean Engineering 2015; 108: 620-635, https://doi.org/10.1016/j.oceaneng.2015.08.058.
40. Yuan Z-M, Incecik A, He S. Hydrodynamic interaction between two ships arranged side by side in shallow water. Proceedings of the ASME 2014 33rd International Conference on Ocean, Offshore and Arctic Engineering, OMAE 2014, At San Francisco, California, USA, 2014; 8A:1-9, https://doi.org/10.1115/OMAE2014-23325.
41. Yuan Z-M, Ji C-Y, Incecik A, Zhao W, Day A. Theoretical and numerical estimation of ship-to-ship hydrodynamic interaction effects. Ocean Engineering 2016; 121: 239-253, https://doi.org/10.1016/j.oceaneng.2016.05.032.
42. Zalewski P, Montewka J. Navigation safety assessment in an entrance channel, based on real experiments. Proceedings of the 12th International Congress of the International Maritime Association of the Mediterranean (IMAM 2007), Varna, Bulgaria, 2007: 1113-1120.
43. Zwijnsvoorde T, Vantorre M, Ides S. Container ships moored at the port of Antwerp: modelling response to passing vessels. 34th PIANC World Congress, Panama, 2018: 476-493.

Vytautas PAULAUSKAS
Klaipeda Shipping Research Centre
V. Berbomo str. 7-5, LT-92219 Klaipeda, Lithuania

Ludmiła FILINA-DAWIDOWICZ
Faculty of Maritime Technology and Transport
West Pomeranian University of Technology, Szczecin
Ave. Piastów 41, 71-065 Szczecin, Poland,

Donatas PAULAUSKAS
Maritime Engineering Department
Klaipeda University
H. Manto g. 84, LT-92294 Klaipeda, Lithuania

E-mails: vytautaskltc@gmail.com, ludmila.filina@zut.edu.pl, paulauskasd75@gmail.com

Article citation info:
PUCHALSKI A, KOMORSKA I, ŚLĘZAK M, NIEWCZAS A. Synthesis of naturalistic vehicle driving cycles using the Markov Chain Monte Carlo method. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 316–322, http://dx.doi.org/10.17531/ein.2020.2.14.

Andrzej PUCHALSKI
Iwona KOMORSKA
Marcin ŚLĘZAK
Andrzej NIEWCZAS

SYNTHESIS OF NATURALISTIC VEHICLE DRIVING CYCLES USING THE MARKOV CHAIN MONTE CARLO METHOD

SYNTEZA EKSPLOATACYJNYCH CYKLI JEZDNYCH SAMOCHODÓW PRZY WYKORZYSTANIU METODY MONTE CARLO Z ZASTOSOWANIEM ŁAŃCUCHÓW MARKOWA*

Simulation methods commonly used throughout the design and verification process of various types of motor vehicles require development of naturalistic driving cycles. Optimization of parameters, testing and gradual increase in the degree of autonomy of vehicles is not possible based on standard driving cycles. Ensuring representativeness of synthesized time series based on collected databases requires algorithms using techniques based on stochastic and statistical models. A synthesis technique combining the MCMC method and multifractal analysis has been proposed and verified. The method allows simple determination of the speed profile compared to classic frequency analysis.

Keywords: *naturalistic vehicle driving cycles, synthesis of driving cycles, Markov models, Monte Carlo simulation.*

Metody symulacyjne powszechnie stosowane w całym procesie projektowania i weryfikacji różnych typów pojazdów mechanicznych wymagają opracowania eksploatacyjnych cykli jezdnych. Optymalizacja parametrów, testowanie i stopniowe zwiększanie stopnia autonomiczności pojazdów nie jest możliwe na bazie standardowych cykli jezdnych. Zapewnienie reprezentatywności syntezowanych szeregów czasowych na podstawie zgromadzonych baz danych wymaga algorytmów wykorzystujących techniki bazujące na modelach stochastycznych i statystycznych. Zaproponowano i zweryfikowano technikę syntezy łączącą metodę Monte Carlo wykorzystującą łańcuch Markowa (MCMC) oraz analizę multifraktalną. Metoda umożliwia proste wyznaczenie profilu prędkości jazdy w porównaniu do klasycznej analizy częstotliwościowej.

Słowa kluczowe: *eksploatacyjne cykle jezdne samochodu, synteza cykli jezdnych, modele Markowa, symulacja Monte Carlo.*

1. Introduction

Contemporary methods of designing and testing mechanical vehicles are based on simulation techniques that require the use of precise models of vertical and horizontal dynamics and sequences of random events occurring in road traffic conditions. This problem is relevant to the optimization of vehicles with internal combustion engines, electric vehicles (EV) and hybrid electric vehicles (HEV). To select the most appropriate drive system architecture for a particular vehicle class and driving cycle, it is necessary to optimize the size of components according to their cost functions, such as the lowest CO_2 emissions, the lowest weight, fuel savings or any combination of these attributes in the architecture [1, 7, 9 and 18].

Regardless of the simulation technique used: quasi-static using a "Backward-facing" vehicle model or a dynamic simulation with a "Forward-facing" model, understanding of the representative driving cycle is essential. In the first case, for an open-loop system, the time series of speed is imposed on the input of the vehicle model in order to calculate rpm and torque on the wheels. In a closed-loop vehicle model, on the other hand, the driving cycle is a setpoint for the driver block, which generates a suitable engine torque. The time and cost constraints associated with the design and testing of various possible vehicle architectures require methods of driving cycle synthesis that can meet the modelling and simulation requirements of automotive engineers throughout the R&D process. It is not possible to optimize the parameters and gradually increase the autonomy of the vehicles based on standard driving cycles, and such optimization cannot prevent "cycle beating". To ensure that the synthesized time series based on the collected databases are representative, it is necessary to use algorithms adopting techniques based on stochastic and statistical models [6, 19]. To define the equivalence criteria, the synthesis process is concluded with a verification of the results, i.e. each generated cycle, through statistical analysis in the time or frequency domain. A combination of multiple criteria is frequently used [2, 4].

The methods of driving cycles construction require quantization of traffic parameters. Depending on their function (emissions estimation, fuel consumption estimation or traffic engineering, etc.), the defined states can be synthesized for micro-trips, segments, heterogeneous classes or modal cycles [17]. Micro-trips are driving models between stops including periods of inactivity. Traffic signals and overloads contribute to "stop-go" driving patterns, and result in increased fuel

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

consumption. Micro-trips are a good representation of fuel consumption and emissions. Segments model situations for various types of roads and driving conditions classified, for instance, by the Level of Service (LOS). They may start and end with different driving parameters, which is why speeds and accelerations have to be accordingly adjusted when combining segments in the course of cycle synthesis. Driving cycles based on heterogeneous traffic classes determined through a statistical breakdown of data are constructed as kinematic sequences using probabilistic methods and analysis of probability distributions. This method is not aimed at testing emissions and fuel consumption. Modal cycles represent recorded parameters of vehicle traffic for specific acceleration intervals, constant speed or idle periods. In procedures using the theory of stochastic processes to analyse the equations of vehicle dynamics, represented by speed and acceleration, the major trend in recent research focuses on methods based on the Markov chains theory [8, 11]. There have also been attempts at using 3D Markov models in the synthesis of driving cycles, which incorporate the roadway slope [20]. Methods based on multi-dimensional Markov chains enable a realistic assessment of fuel consumption and CO_2 emissions, even after time compression of the synthesized time series [5]. However, such simulations involve a high time cost.

This paper proposes a method for synthesizing naturalistic driving cycles in which the information about instantaneous values of acceleration can be replaced by the degree of multifractality assessed using formalism based on wavelet leaders. This helped reduce the number of Markov chain dimensions in the simulation process. The process was illustrated on the example of the Markov chain Monte Carlo (MCMC) algorithm for the vehicle speed signal. The input for the algorithm was recorded during a series of experiments in real conditions. Statistical factors and mean tractive force (MTF) were used to select and classify road traffic models equivalent to the real conditions.

2. Wavelet leaders multifractal formalism in MCMC technique

Each real driving cycle can be regarded as a sequence of random transitions between defined m-states of the vehicle occurring in road traffic. The frequency of specific states depends on the technical parameters of the car, the intensity of road traffic and the driver's behaviour. By determining the probability of remaining in the current state or transition into a different state, we can describe the examined phenomenon in the form of a transition probability matrix (TPM) (1):

$$P = \begin{bmatrix} P_{11} & \cdots & P_{1m} \\ \vdots & \ddots & \vdots \\ P_{m1} & \cdots & P_{mm} \end{bmatrix} \in R^{mxm}, \tag{1}$$

where the entry P_{ij} (2) is the probability of transition from and to state j when $j \neq i$ or remaining in state i when $j = i$. The probability P_{ij} can be calculated using the following equation:

$$P_{ij} = \frac{N_{ij}}{\sum_j N_{ij}}, \tag{2}$$

where N_{ij} is the number of transitions from and to state j. The sum of the values of entries in each row is equal to one. The random process $\{X_n\}_{n\in N}$ is referred to as the Markov chain if for any $n \in N$ the following equation is true: $\mathrm{P}\{X_{n+1}|X_n\} = \mathrm{P}\{X_{n+1}|X_0, X_1, \ldots, X_n\}$. It is assumed that the TPM is stationary, which implies that the Markov chain is homogeneous. Therefore, for Markov chains, the conditional distribution of the random variable X_{n+1} depends only on the currently known value of X_n. Thus, considering the current driving state, the future state can be determined using Monte Carlo simulation based on the transition probability matrix. It is possible to generate a driving cycle of any duration, which may be used to identify a cycle with the required duration, for the assumed equivalence criteria.

The synthesis of a driving cycle using the MCMC method, where – in addition to the speed signal – also consider other parameters are taken into consideration, requires a multi-dimensional description of the defined vehicle states, which significantly complicates the determination of the transition probability matrix and extends the implementation time of the algorithm. If the second parameter is acceleration, which is not measured directly in most real driving cycles, it becomes necessary to differentiate the speed signal in order to acquire information about motion dynamics. Where this is the case, the standard 1-second sampling period for the time series of speed does not guarantee a sufficient precision of the acceleration signal.

Papers where road traffic was analysed based on recorded vehicle speed signals indicate the multifractal properties of the dynamics of such traffic [3, 16]. Multifractality can also be observed both in real and standard driving cycles [14]. Our research proposes to eliminate the acceleration signal from the multi-dimensional description of vehicle states using information about driving dynamics represented by multifractal parameters of the speed signal. The iteration in the Monte Carlo simulation was performed for a specific time, with a requirement concerning driving dynamics. The multifractal analysis, which is based on estimated scaling exponents of the signal, is a popular statistical tool used to assess empirical data. In the case of time series, mathematical formalism was initially based on increments of their value, measured as Hölder point exponents h of time function $x(t)$ at point t_0, determined by the supremum of all exponents that, for constant $C > 0$, meet the following condition: $\left|x(t) - P_n(t-t_0)\right| \leq C\left|t-t_0\right|^h$, where $P_n(t-t_0)$ is a polynomial of degree $n < h$ [13, 15, 16]. The result of the algorithm is the multifractal spectrum $D(h)$, i.e. a function describing the fractal dimensions of points with the same Hölder exponent.

The multifractal formalism in the time and frequency domain that is used in the research makes it possible to estimate multifractal parameters using wavelet leaders, which are representatives of local Hölder exponents of the signal. The algorithm is characterized by low computing costs, numerical stability and high versatility with respect to real signals. For coefficients (3) of the discrete wavelet transform (DWT) of function $x(t)$ and basic wavelet with a compact support $\psi_0(t)$:

$$d_x(j,k) = \int_R x(t) 2^{-j} \psi_0\left(2^{-j}t - k\right) dt, \tag{3}$$

wavelet leaders (4) for the collection of largest coefficients $d_x\left(j',k'\right) \equiv d_{\lambda'}$ in the neighbourhood of 3λ are defined in any scale by the following equation:

$$L_x(j,k) = sup_{\lambda' \in 3\lambda}\left|d_{\lambda'}\right|, \tag{4}$$

where j,k are integers and $3\lambda := 3\lambda_{j,k} = \lambda_{j,k-1} \cup \lambda_{j,k} \cup \lambda_{j,k+1}$ and $3\lambda := \lambda_{j,k} = \left[k2^j, (k+1)2^j\right]$.

It can be demonstrated [10] that Hölder exponents are scaling exponents of wavelet leaders: $L_x(j,k) \sim 2^{jh}$. Also, the structural function (5) defined for wavelet leaders is described by a power law where the exponent is a multifractal scaling exponent $\zeta(q): R \to R$.

$$Z_L(q,j) = \frac{1}{n_j}\sum_{k=1}^{n_j} L_x(j,k)^q = \mathbb{E}L_x(j,k)^q \sim 2^{j\zeta(q)}, \tag{5}$$

where q is the order of the structural function, and n_j is the number of intervals of the multi-resolution analysis.

The function generated using the Legendre transformation of the multifractal scaling exponent $\zeta(q)$, under mild regularity conditions, is the upper limit of the multifractal spectrum (6) of the investigated signal:

$$D(h) \le min_{q\neq 0}\left[1 + qh - \zeta(q)\right] \tag{6}$$

Coefficients of the Taylor expansion of the exponent $\zeta(q)$ – log-cumulants c_p – are an alternative description of the parameters of the multifractal spectrum of the analysed signal:

$$\zeta(q) = \lim_{j\to 0}\frac{\log_2 Z_L(q,j)}{j} = \sum_{p=1}^{\infty} c_p \frac{q^p}{p!} = c_1 q + c_2\frac{q^2}{2} + c_3\frac{q^3}{6} + \dots \tag{7}$$

In particular: the coefficient c_1 describes the position of the maximum of the spectrum, and coefficients c_2 and c_3 describe its degree of multifractality, i.e. the width of the spectrum and its asymmetry, respectively. The dynamic properties of the systems are successfully described based on the parameters of the multifractal spectra of the representative time series [12]. Approximation of $\zeta(q)$ (7), i.e. also of the multifractal spectrum $D(h)$, using coefficients c_p significantly simplifies the algorithms for a comparative analysis of the investigated systems.

3. Simulation tests of driving dynamics and wavelet leaders of speed signal

The relationship between the time series of acceleration and parameters of the multifractal spectrum of speed has been illustrated on the example of the synthetic signal $v(n)$ of vehicle speed (Fig. 1a). The signal was resampled to achieve signals with acceleration 2, 4 and 8 times higher. Due to the resampling, the histograms are not identical, but they are comparable. The signals that had been shortened were repeated 2, 4 and 8 times, respectively, to obtain signs with the same number of samples.

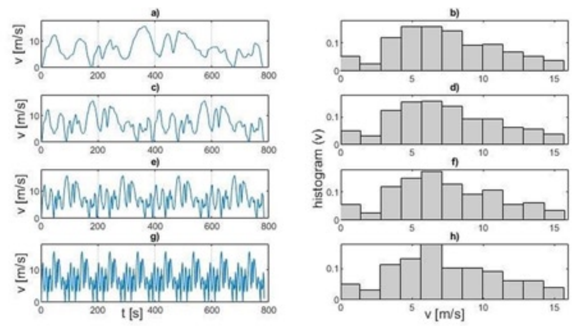


Fig. 1. Synthetic signal of vehicle speed (a) and its histograms (b) in the specified time; the same run in a two times shorter time (c) with comparable amplitude distribution (d); the same run in four times shorter time (e) with comparable amplitude distribution (f); the same run in eight times shorter time (g) with comparable amplitude distribution (h);

For the speed signals in figure 1, accelerations were determined (using differentiation – Fig. 2), as well as multifractal spectra (Fig. 3).

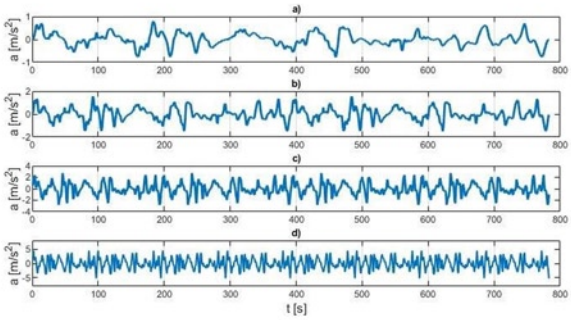


Fig. 2. Accelerations of signals in Fig. 1 a), c), e) and g)

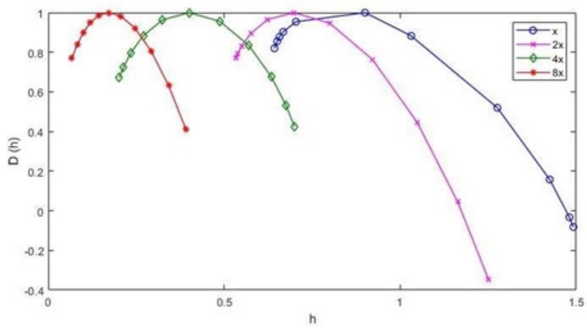


Fig. 3. Multifractal spectra of vibration signals in Fig. 1 a), c), e), g)

An analysis of the singularity spectra demonstrates that the position of their maxima and width depends on the accelerations of the simulated signals. The log-cumulants of synthetic signals (Table 1) and relationships of log-cumulants and acceleration (Fig. 4) were determined. The first and second log-cumulant, describing the position of the maximum of the multifractal spectrum and its width, respectively, were proposed as the synthetic parameters for the assessment of driving dynamics using the multifractal spectrum.

Table 1. Values of log-cumulants of synthetic signals

Log-cumulant	Signal			
	x	2x	4x	8x
c_1	0.8127	0.6976	0.4003	0.1723
c_2	−0.1640	−0.1187	−0.0841	−0.0307
c_3	0.1220	0.0282	0.0127	0.0055

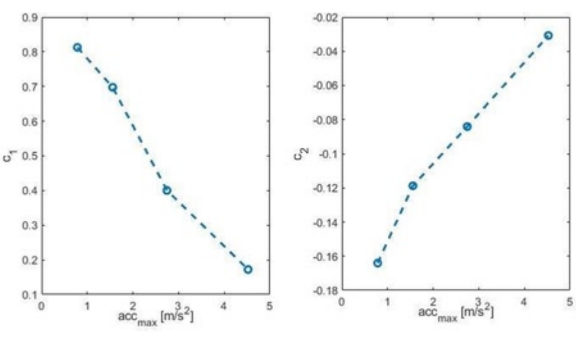


Fig. 4. Relationship between log-cumulants and maximum acceleration, a) first log-cumulant, b) second log-cumulant

4. Implementation of the algorithm for the synthesis of driving cycles and analysis of research results

An algorithm was proposed to generate naturalistic driving cycles using first-order Markov chains and multifractal formalism based on an analysis of wavelet leaders (Fig. 5).

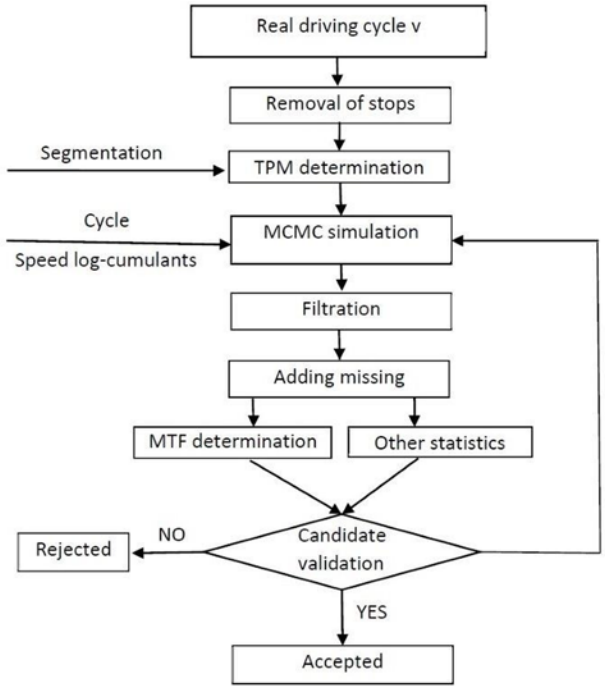


Fig. 5. Block diagram of the MCMC algorithm for synthesizing driving cycles

The paper presents results of tests and analysis of car traffic in actual road conditions, represented by urban driving in a large agglomeration (Fig. 6). The analysis was carried out based on the time series of vehicle speed recorded with a sampling period of 1 s. The research has been described in the paper [14]. Due to the large share of "zero speeds" (idle periods) in the test, amounting to approx. 25%, fragments corresponding to stops were removed from the recorded time series (Fig. 7), which enabled the segmentation and determination of the transition probability matrix (TPM) and testing of driving dynamics through an analysis of log-cumulants. The recorded speeds were divided into 20 even intervals corresponding to increasing speeds other than zero as well as a 21st interval corresponding to stops. Speed resolution is approx. 0.9 m/s. The authors attempted to achieve a fairly good car speed resolution while avoiding intervals with a very low (or zero) probability of occurrence.

Statistical analyses were conducted in the R environment, and the multifractal analysis was carried out in Matlab.

The transition probability matrix (TPM) calculated based on the reference signal of the cycle had the size of 21x21 (Fig. 8).

A simulation of 100 cycles was carried out in accordance with the Metropolis-Hastings algorithm. Three sample cycles – candidates No. 1, 2 and 3 (Figs. 9a–c) were selected to illustrate the results of the algorithm. The main statistics of the speed signal (maximum, minimum, mean and standard deviation) of the sample cycles are similar to the statistics of the reference signal. There was also a fourth cycle shown – candidate No. 4 – generated for verification purposes based on the distribution of speed amplitudes (Fig. 9d).

The first two log-cumulants determined for each cycle (Fig. 10) are the best match of dynamics in comparison with the reference signal for candidate No. 1.

The conformity of probability density distributions to the distribution of the reference cycle has also been verified (Fig. 11a–d). The

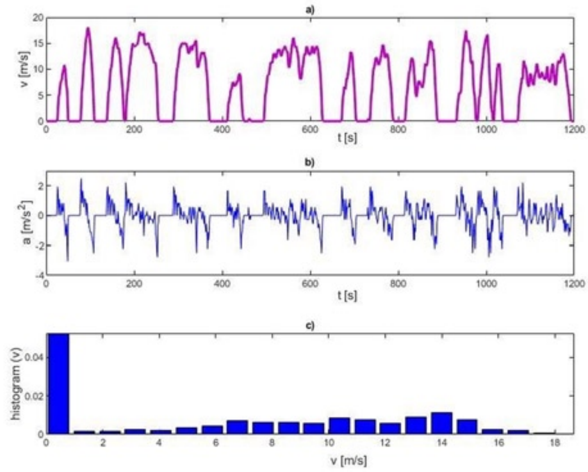


Fig. 6. Real cycle representative for the conducted research – 20-minute fragment, a) speed signal, b) accelerations calculated based on speed, c) car speed histogram

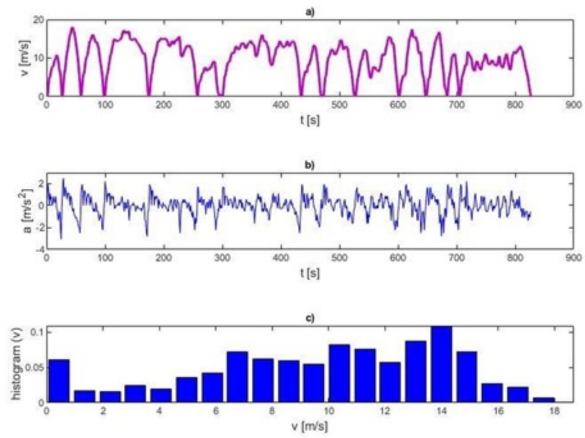


Fig. 7. Typical real cycle after removing stops, a) speed signal (reference signal), b) accelerations calculated based on speed, c) car speed histogram

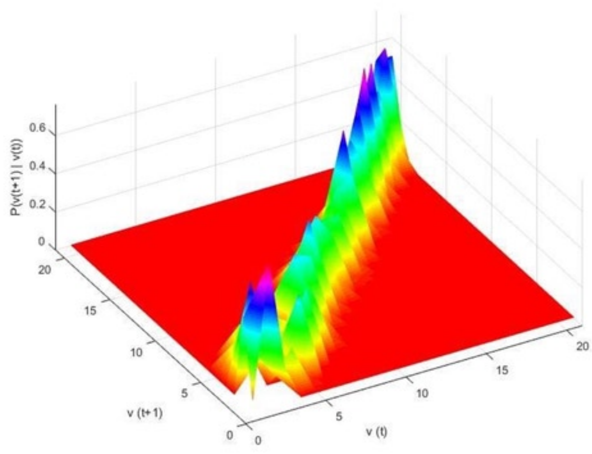


Fig. 8. Transition probability matrix (TPM) for the cycle shown in Fig. 7a

distribution of the reference cycle has been approximated with an empirical function. The chi-squared test or the Kolmogorov–Smirnov test can be used to check the goodness of fit of empirical data to the

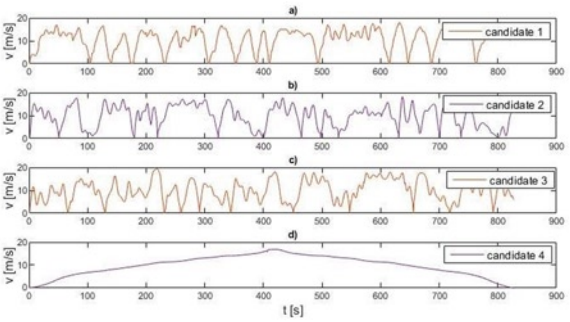


Fig. 9. Sample simulated driving cycles

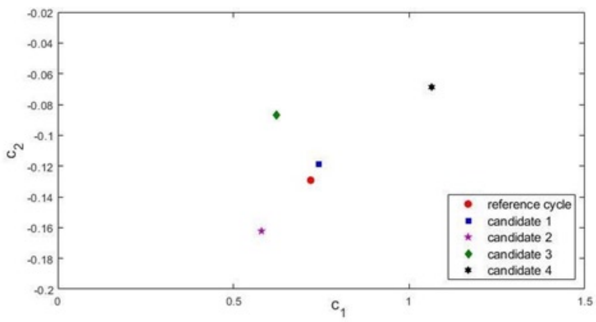


Fig. 10. Scatter plot of log-cumulants determined for the investigated cycles

approximated probability distribution function, but these tests reject the null hypothesis for the investigated duration of the driving cycle. For the test to confirm the null hypothesis, the duration of the driving cycle would have to be significantly extended, which is not possible. In such a situation, it is best to estimate the goodness of fit of the theoretical distribution to the observed distribution through a visual comparison. This was done using the probability-probability plot (Fig. 11e). Apart from the cycle of candidate No. 4, the best fit is demonstrated by candidate No. 1.

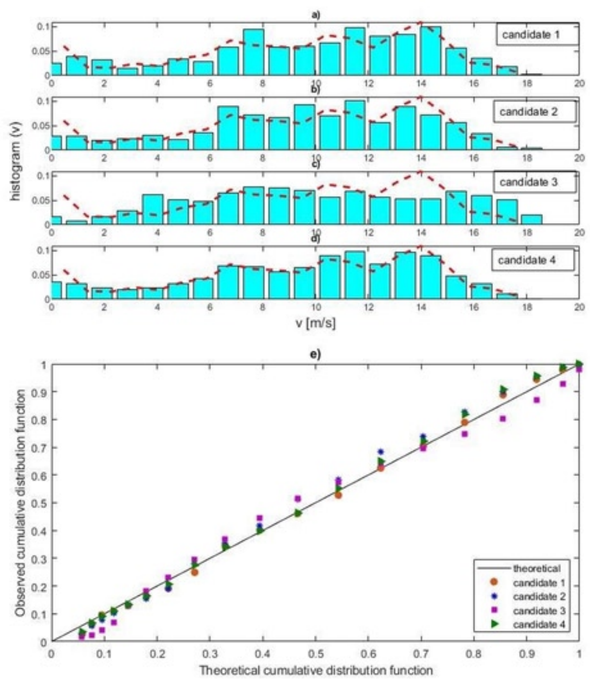


Fig. 11. Histograms of the speed signal for the simulated cycles compared with the reference distribution (red dashed line) a)–d) e) probability-probability plot

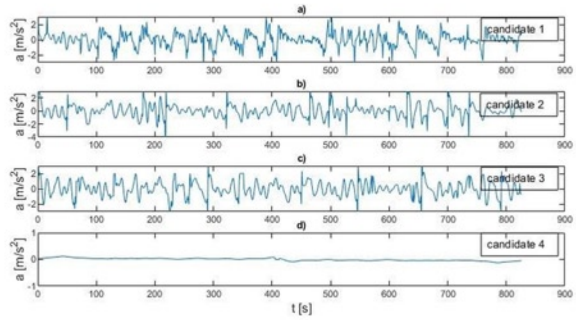


Fig. 12. Accelerations in the simulated cycles calculated based on speed

The accelerations of candidate No. 4 (Fig. 12), which has a speed amplitude distribution that perfectly matches the reference distribution, are entirely different – almost constant. The accelerations of the remaining candidates can only be assessed in terms of their minimum and maximum values.

In the method of modal cycles and speed-based segmentation, which was adopted in this paper, the time series produced using the Markov model are stepped, which means that they had to be smoothed in the next step. The method of local quadratic regression smoothing was selected from among the various methods to smooth the series. Once the iteration and filtration process was completed, the stop periods were added to the time series, and a search was started for the most representative cycles out of all of the cycles produced by the synthesis, for the selected equivalence factor.

In the course of the study, the results of the algorithm for the synthesis of equivalent driving cycles were analysed according to selected statistical parameters and the criterion of the mean tractive force (MTF) (8), i.e. the tractive energy of the vehicle (Table 2) transmitted through the wheels:

$$\bar{F}_{trac} = \frac{1}{x_{total}} \int_{t\in\tau_{trac}} F(t)v(t)dt \,. \tag{8}$$

where: total tractive force $F(t)$ is the sum of the forces of aerodynamic resistance F_{air}, rolling resistance F_{roll} and inertia of the vehicle F_{iner}, $v(t)$ and $a(t)$ are speed and acceleration, respectively, for a driving cycle of the duration x_{total} and τ_{trac} represents the time intervals during which $F(t) > 0$.

In the calculations of the MTF coefficient, the most significant element is the force of inertia, which is proportional to acceleration. The best fit to the real cycle according to the MTF criterion is represented by candidate No. 1. The primary parameters considered in the course of cycle verification are listed in Table 3. Minimum and maximum values of speed and acceleration, subject to initial verification, were omitted.

All synthesized driving cycles have the correct mean value and standard deviation. The selection also cannot be performed based on the distributions of probability density. Candidates No. 1 and No. 4 show the best fit of speed amplitudes probability distribution.

If we assume a discrepancy of the MTF coefficient of the generated cycle with the reference cycle of over 10% to be an unacceptable

Table 2. Vehicle parameters

F_{air} [N]	Aerodynamic resistance	$0.4v^2(t)$
F_{roll} [N]	Rolling resistance	383
F_{iner} [N]	Inertia	$1300a(t)$

Table 3. Summary of selected values characteristic to the investigated cycles

Driving cycle	Fit of the distribution	Mean value	Standard deviation	Log-cumulant 1	Log-cumulant 2	MTF
	[-]	[m/s]	[m/s]	[-]	[-]	[N]
Reference cycle		9.7	4.4	0.72	-0.13	689
Candidate 1	+	10.0	4.6	0.74	-0.12	743
Candidate 2	+/-	9.7	4.4	0.58	-0.16	820
Candidate 3	+/-	10.4	4.5	0.62	-0.10	796
Candidate 4	+	9.7	4.6	1.06	-0.07	450

in terms of equivalence to real driving conditions, the log-cumulants tested in the phase of synthesis of candidate cycles and the MTF used to verify their equivalence suggest candidate No. 1.

5. Conclusions

The presented research results provide a new perspective on statistical-random methods for synthesizing real vehicle driving cycles. It was demonstrated that driving dynamics represented by acceleration could be reproduced using multifractal parameters of speed signals. The use of wavelet leaders for driving dynamics testing made it possible to carry out cycle synthesis, which took into consideration speed and acceleration, using Monte Carlo simulation with a single-dimensional Markov chain. The algorithm for the synthesis of equivalent driving cycles was verified using the criterion of mean tractive force (MTF).

The database used so far included data from tests of vehicles with internal combustion engines. The authors' future research will include an analysis of driving cycle prediction and road traffic modelling for the purpose of drive system control and electricity management in electric vehicles. The expected results will be useful in designing the infrastructure of charging stations for electric cars.

References

1. Brady J, O'Mahony M. Development of a driving cycle to evaluate the energy economy of electric vehicles in urban areas. Applied Energy 2016; 177: 165-178, https://doi.org/10.1016/j.apenergy.2016.05.094.
2. Chłopek Z. The synthesis of driving tests according to the criterions of similarity of frequency characteristics. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18(4): 572-577, https://doi.org/10.17531/ein.2016.4.12.
3. Dai M, Zhang Ch, Zhang D. Multifractal and singularity analysis of highway volume data. Physica A 2014; 407: 332-340, https://doi.org/10.1016/j.physa.2014.04.005.
4. Esser A, Zeller M, Foulard S, Rinderknecht S. Synthesis of Representative and Multidimensional Driving Cycles. SAE International 2018; 2018-01-0095.
5. Hereijgers K, Silvas E, Hofman T, Steinbuch M. Effect of using Synthetized Driving Cycles on Vehicle Fuel Consumption. IFAC papers 2017; 50-1: 7505-7510, https://doi.org/10.1016/j.ifacol.2017.08.1183.
6. Lee T-K, Filipi ZS. Synthesis of real-world driving cycles using stochastic process and statistical methodology. International Journal of Vehicle Design 2011; 57(1): 17-36, https://doi.org/10.1504/IJVD.2011.043590.
7. Liu Z, Ivanco A, Filipi Z. Naturalistic drive cycle synthesis for pickup trucks. Journal of Safety Research 2015; 54: 109-115, https://doi.org/10.1016/j.jsr.2015.06.005.
8. Man Z, Shuming S, Nan L, Bingjian. High-efficiency Driving Cycle Generation Using a Markov Chain Evolution Algorithm. IEEE Transactions on Vehicular Technology. December 2018, https://doi.org/10.1109/TVT.2018.2887063.
9. Mayakuntla SK, Vermab A. A novel methodology for construction of driving cycles for Indian cities. Transportation Research Part D 2018; 65: 725-735, https://doi.org/10.1016/j.trd.2018.10.013.
10. Mukli P, Nagy Z, Eke A. Multifractal formalism by enforcing the universal behavior of scaling functions. Physica A 2015; 417: 150-167, https://doi.org/10.1016/j.physa.2014.09.002.
11. Nyberg P, Frisk E, Nielsen L. Generation of Equivalent Driving Cycles Using Markov Chains and Mean Tractive Force Components. Proc. of IFAC, 2014; 8787-8792, https://doi.org/10.3182/20140824-6-ZA-1003.02239.
12. Puchalski A, Komorska I. Binomial multifractal features of worldwide harmonized light duty vehicles test cycle. JVE Vibroengineering Procedia 2017; 13: 175-179, https://doi.org/10.21595/vp.2017.19074.
13. Puchalski A, Komorska I. Multifractal Nature of Diesel Engine Rattle Noise in Vehicle. Archives of Acoustics 2017; 42(3): 469-474, https://doi.org/10.1515/aoa-2017-0049.
14. Puchalski A, Ślęzak M, Komorska I, Wiśniowski P. Multifractal analysis vehicle's in-use speed profile for application in driving cycles. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (2): 177-181, https://doi.org/10.17531/ein.2018.2.02.
15. Puchalski A, Komorska I. A generalised entropy in multifractal time signals analysis of mechanical vibration. JVE Journal of Vibroegineering 2018; 20 (4): 1667-1675, https://doi.org/10.21595/jve.2018.18871.
16. Shang P, Lu Y, Kamae S. Detecting long-range correlation of traffic time series with multifractal detrended fluctuation analysis. Chaos, Solitions and Fractals 2008; 36: 82-90, https://doi.org/10.1016/j.chaos.2006.06.019.
17. Silvas E, Hereijgers K, Peng H, Hofman T., Steinbuch M. Synthesis of Realistic Driving Cycles with High Accuracy and Computational Speed, Including Slope Information. IEEE Transaction Vehicular Technology 2016; 65(6): 4118-4128, https://doi.org/10.1109/TVT.2016.2546338.
18. Tutuianu M, Marotta A et al. Development of a World-wide Worldwide harmonized Light duty driving Test Cycle (WLTC) and a possible pathway for its introduction in the European legislation. Transportation Research Part D Transport & Environment 2015; 40(4), 61-75, https://doi.org/10.1016/j.trd.2015.07.011.

19. Yi Y, Pengjian S. Multiscale multifractal detrended cross-correlation analysis of traffic flow. Nonlinear Dynamics 2015; 81:1329-1347, https://doi.org/10.1007/s11071-015-2072-7.
20. Zhao B, Lv Ch, Hofman T. Driving Cycle Aware Energy Management of Hybrid Electric Vehicles Using a Three Dimensional Markov Chain Model. Automotive Innovation 2019; 2: 146-156, https://doi.org/10.1007/s42154-019-00059-z.

Andrzej PUCHALSKI
Iwona KOMORSKA
Department of Mechanical Engineering
Kazimierz Pulaski University of Technology in Radom
Malczewskiego 29., 26-600 Radom, Poland

Marcin ŚLĘZAK
Andrzej NIEWCZAS
Motor Transport Institute
Jagiellońska 80, 03-301 Warszawa, Poland

E-mails: andrzej.puchalski@uthrad.pl, iwona.komorska@uthrad.pl, marcin.slezak@its.waw.pl, andrzej.niewczas@its.waw.pl

Article citation info:
KOSUCKI A, STAWIŃSKI Ł, MALENTA P, ZACZYŃSKI J, SKOWROŃSKA J. Energy consumption and energy efficiency improvement of overhead crane's mechanisms. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 323–330, http://dx.doi.org/10.17531/ein.2020.2.15.

Andrzej KOSUCKI
Łukasz STAWIŃSKI
Piotr MALENTA
Jakub ZACZYŃSKI
Justyna SKOWROŃSKA

ENERGY CONSUMPTION AND ENERGY EFFICIENCY IMPROVEMENT OF OVERHEAD CRANE'S MECHANISMS

ENERGOCHŁONNOŚĆ I POPRAWA EFEKTYWNOŚCI ENERGETYCZNEJ MECHANIZMÓW SUWNICY POMOSTOWEJ

The article presents the numerical investigation of the overhead crane's energy consumption. The analysis is based on the hybrid model of the crane consisting of numerical model of drive mechanisms as bridge, trolley, hoist and also experimentally measured power consumption of each control unit. The numerical model was verified experimentally on the real crane. The investigation focuses on analyzing the energy consumption of the overhead crane in relation both to the travelled distance and also for the lifting and lowering heights of a suspended payload. Particular attention was paid on the cases straightly related to the hoist, as a main factor of improvements in the energetic efficiency of the overhead crane. Energy consumption was investigated for a variety of magnitudes of transported mass.

Keywords: *energy consumption, energy efficiency, crane, hoist.*

W artykule przedstawiono badania symulacyjne energochłonności pracy suwnicy pomostowej. Podstawą badań jest model hybrydowy bazujący na numerycznych modelach mechanizmów mostu, wózka i wciągarki oraz na eksperymentalnie zmierzonym zapotrzebowaniu mocy dla układu sterowania. Model numeryczny został zweryfikowany na rzeczywistej suwnicy. W pracy przedstawiono analizę energochłonności mechanizmów jazdy w zależności od pokonanej drogi, jak również mechanizmu podnoszenia w zależności od wysokości podnoszenia i opuszczania ładunku. Zwrócono szczególną uwagę na mechanizm wciągarki, jako na główny czynnik poprawy efektywności energetycznej. Energochłonność została zbadana dla różnych mas transportowanego ładunku.

Słowa kluczowe: *energochłonność, efektywność energetyczna, suwnica, wciągarka.*

1. Introduction

Overhead cranes are dominant pieces of handling equipment, which can be found in the majority of industrial applications. Thus, the cranes and their mechanism are a subject of interest among researchers [19, 21]. In all manufacturing and logistics systems cranes are supplied with electrical energy, which is obtained mainly by burning fossil fuels [10]. Because of this, the air pollution and greenhouse gases emission causing the greenhouse effect becomes a major problem. Hence, main stream of improvements in overhead crane operation and design are focused on reduction of energy consumption and its optimization.

The main area of the energy flow in an overhead crane is the hoist. Lifted payload buffers significant amount of the potential energy. While the payload is being lowered, the energy of gravity ineffectually dissipates as heat energy in brake resistor or brake. The European Council of October 2014 made a commitment [6] to reduce overall greenhouse gas emissions of the Union by at least 40% below 1990 levels by 2030. Many countries struggle with reducing the CO_2 emission and need to buy more carbon units and make larger contribution towards emission reductions in order to achieve desirable levels. Hence, the electrical energy obtained from coal fired power plants becomes more expensive. Regarding to a lot of industrial applications of an old – type overhead crane solutions, this phenomena raises a problem in the economical point of view. Potential savings could significantly lower the costs of crane maintenance in the field.

Currently, many articles and publications focus on the problem of cranes energy efficiency. An example would be work [2], in which the study of lifting mechanisms was discussed, among others in terms of energy overload. In addition, many articles describe various methods of reducing energy consumption for overhead cranes.

Supercapacitors are a novel energy storage device based on the principle of double layer – electrolyte capacity. They have many merits such as a long lifetime, high efficiency and fast dynamic response [4]. Hybrid supercapacitors – based propulsion systems are considered in [1, 4, 17, 11, 16]. Paper [11] faces the problem of energy loses in AC powered overhead crane by proposing the application of an auxiliary energy storing device to hoisting plants. Scientists did the investigation by numerical simulations based on the mathematical models. Simulations report that introducing hybrid storage system allows to save one third of necessary energy. Paper [4] also takes under consideration the energy – saving system also based on the supercapacitors. Authors face the problem of voltage equalization strategy in rubber tires gantry crane (RTG). It was noticed that supercapacitors used in energy storage device can be exposed to an over – voltage, which leads to shortening a life time of the system and also causes problems with the reliability. An active voltage equalization circuit based on reversible converter is proposed. Theoretical simulations have shown that the active control unit solves the problem of partial over – voltage in supercapacitors groups and can be applied in RTGs energy saving systems. The authors of paper [5] evaluate a hybrid configuration and improved management system for the 65 tones RTG driven by diesel generators. Supercapacitors, in this case constituting Energy Storage System (ESS), were connected to a DC-bus and controlled based on the voltage of the DC – bus. The undoubted advantage of the above project was the simulation of a hybrid system in the real working cycle of the RTG crane. The improvement leads to 20% decrease of fuel consumption and double the total energy efficiency of the system.

The article [17] discusses the issue of ultracapacitors in the context of the practical application of NZEB (Nearly Zero-Energy Buildings) systems. Such systems use almost only locally produced energy (often from renewable sources). The use of ultracapacitors allows them to store excess energy generated.

Parallel to the energy saving systems, scientists have investigated other field of energy savings. The article [18] focuses on the energetic efficiency of hoisting mechanisms in industrial applications. A typical duty cycle of the hoist is periodically intermittent. Simultaneously there are pauses during the work. Cranes, the majority of time, work at partial load. When a sequential run with constant load takes place, the hoisting mechanism has to be driven by motors designed for S3 duty cycle [12]. Authors of paper [18] investigate the influence of motor's rotor material on the energetic efficiency for a few stack lengths. The conclusion is that because of low operating hours and major share of partial loads, increasing efficiency does not provide notable energy or cost savings.

Paper [22] considers energy efficiency in a proposed control approach. To minimize energy consumption the authors employ planning an optimal trajectory, tracked in real time. The study focuses only on horizontal transportation considered as a linear movement of the trolley with swing of the suspended payload. Two steps are key to minimizing energy consumption: planning and tracking the trolley's trajectory (so called Model Predictive Control MPC). An air resistance was also and the load was treated as point mass. The investigation notes that thanks to the elaborated method energy savings can be obtained, but the authors do not give a quantity of potentially saved energy.

Increasing energy efficiency by reducing energy consumption can also be achieved through nonlinear crane control schemes during load's skew rotation [7]. The use of an electromechanical clutch allows the engine to be disconnected from the load, which can only rotate under certain circumstances due to inertia forces. In the above work, ways of solving the control problem of such a system were proposed: with and without switching.

Energy efficiency can also be ensured by properly configured and calculated machine construction. In article [20] the finite element method was used to design an energy – efficient crane construction, and the previously developed transmission design was used as input. Therefore, the presented research on energy – saving crane design can be divided into three parts: energy-saving metal construction, energy-saving transmission design and energy-saving electrical system design. The result of the research is the improvement of a crane's parameters affecting its energy efficiency, such as crane weight, while maintaining all basic requirements related to safety.

A lot of papers concerning energy efficiency deal with the issue of the use of flywheels. Due to the fact that when the crane load is lowered, the energy from the engine is wasted on generating heat - the possibility of using flywheels that could store energy so that it does not dissipate was investigated [8, 9]. Thanks to that, it is possible to achieve a lower fuel consumption for diesel engines in RTG cranes.

In article [1] the advantages of flywheels over the previously mentioned supercapacitors were considered. Features of supercapacitors such as high efficiency, longer life span for flywheels are even more noticeable. The above work also discusses one way of controlling an energy storage system based on a flywheel.

Paper [2] lists other advantages of the flywheel used in overhead cranes for increased energy efficiency. First of all, the ability to work under overload is greater than in the case of supercapacitors. In addition, there is no risk of explosion with flywheels. An important aspect is also their lower price.

The literature studies lead to the conclusion that there is lack of investigations assessing the magnitude of energy consumption for cranes. Hence, the authors of this article decided to address this problem. A simulation of overhead crane's energy consumption was carried out in this paper.

The basis for the simulation studies is a hybrid model of crane drives and control systems consisting of experimentally verified mathematical models of the bridge, the hoist and the trolley and actual data covering energy consumption of the control and energy supply systems. The calculations refer to the typical overhead crane duty cycle including the possibility of energy restoration while lowering of the payload and braking of the whole system.

A new approach presented in this paper is developed by engaging numerical simulations with an experimentally measured amount of the energy consumed by the drives control system and also energy necessary for releasing the brakes. This approach gives a comprehensive view of energetic relations between each of mechanism in cranes, what provides the possibility of the energy recuperation or more rationally energy management. As an example can be the system presented in this publication, enabling both the supply of more drive systems from one intermediate circuit (DCLink) with the possibility of energy return to the power grid, or the exchange of energy between drives connected to DCLink.

The paper is organized in the following order. Chapter 2 defines the energy consumption of the mechanism. Chapter 3 describes a hybrid model engaged to assess the energy consumption of each mechanism. The experimental stand where the measurements were carried out is presented in chapter 4. Results of the investigation are provided in chapter 5. Finally, the conclusions drawn from the investigation and also discussion can be found in chapter 6.

2. Definition of the energy consumption

Driving system based on electric motor characterized by certain power consumption in each part of duty cycle. Fluctuations in power consumption depend on movement resistance of the mechanism and velocity of the movement. Both of these parameters are variable while the movement. Therefore, calculating an energy consumption can be done according to the following formula:

$$E = \int_0^{t_k} F \cdot v \cdot dt \tag{1}$$

where:

F - driving force [N],
v - instantaneous velocity [m/s].

Usually it is impossible to directly assess the driving force. An example could be driving mechanisms of the bridge and the hoist. Hence, estimation of the energy consumption can be calculated basing on magnitudes related directly to the electric motors. Mentioned magnitudes are both mechanical torque on the motor's shaft (or motor's current) and also shaft's angular velocity. Thanks to those data there is a possibility of assessing the instantaneous power as follows:

$$N = M_s \cdot \omega_s \tag{2}$$

where:

M_s - motor's torque [Nm],
ω_s - angular velocity of motor's shaft [rad/s].

Power is derivative of the work (and the energy also) in a definition of time. Hence, to estimate the energy consumption the integration of the motor's power function was conducted according to the following formula:

$$E = \int_0^{t_k} N \cdot dt = \int_0^{t_k} M_s \cdot \omega_s \cdot dt \tag{3}$$

Necessary data for above calculations can be obtained by measurement of parameters directly on the electric motors. Therefore, there is need to install appropriate sensors to obtain necessary data. The other way to assess the energy consumption and also energy efficiency is a computer simulation using verified numerical model of the crane, where costs in comparison to laboratory tests are negligible. This article presents the investigation of crane, where the maximum loading and maximum range are 50 kN and 10 m, respectively.

3. Hybrid model of crane's mechanisms

The hybrid model of the overhead crane is based both on the numerical simulations of drive mechanisms and on the measurement of the values of energy consumed by the drives control system.

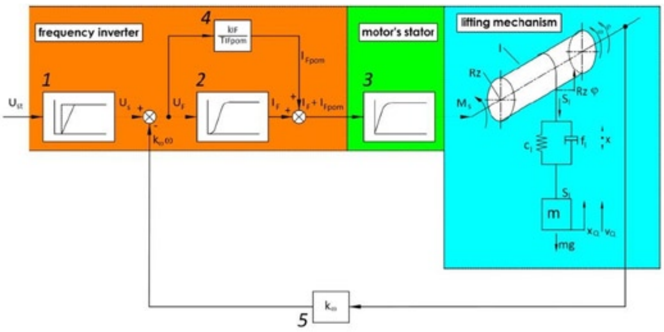


Figure 3.1. The model of the hoist and the inverter – motor's stator [13]

Numerical model of the lifting mechanism fed by inverter is widely discussed in the paper [14]. Figure 3.1 shows model of the hoist driven by an asynchronous motor fed by an inverter, developed basing on the model described in [13]. The dual mass model of mechanical part is defined by typical dynamic equations and kinetic relations. Drive system inverter – stator is described using typical equations of dynamic element, where: element 1 – integrating controller (installed in the inverter to prevent the overstoring oh the entry signal U_{st}, element 2 – second order delay element, element 3 – first order delay element and element 4 – integrator. Element 5 is an angular velocity feedback of the motor shaft. Magnitude of velocity is transformed to the voltage.

Basing on dependencies developed in [13] both systems were described in space of state variables as follows:

a) hosting mechanism:

$$\frac{d\omega}{dt} = \frac{1}{I_z} \cdot M_s - \frac{c_l}{I_z} \cdot x - \frac{R_z f_l}{I_z} \cdot \omega + \frac{f_l}{I_z} \cdot v_Q \tag{4}$$

$$\frac{dx}{dt} = R_z \cdot \omega + v_Q \tag{5}$$

$$\frac{dv_Q}{dt} = \frac{c_l}{m} \cdot x - \frac{R_z \cdot f_l}{m} \cdot \omega + \frac{f_l}{m} \cdot v_Q - g \tag{6}$$

a) inverter – stator system:

$$\frac{dU_s}{dt} = \frac{1}{T_c} \cdot U_{sz} \tag{7}$$

$$\frac{dI_F^{'}}{dt} = \frac{k_{IF} \cdot k_{wzm}}{T_{F_1}^2} \cdot \left(U_s - k_\omega \cdot \omega\right) - \frac{T_{F_2}}{T_{F_1}^2} \cdot I_F^{'} - \frac{1}{T_{F_1}^2} \cdot I_F \tag{8}$$

$$\frac{dI_F}{dt} = I_F^{'} \tag{9}$$

$$\frac{dM_s}{dt} = \frac{k_{M_s}}{T_{M_s}} \cdot \left(I_F + I_{F_{pom}}\right) - \frac{1}{T_{M_s}} \cdot M_s \tag{10}$$

$$\frac{dI_{F_{pom}}}{dt} = \frac{k_{I_F}}{T_{I_{F_{pom}}}} \cdot \left(U_s - k_\omega \cdot \omega\right) \tag{11}$$

where the state variables are:

x - elongation of the wire ropes
v_Q - velocity of the payload
ω - angular velocity of the motor
I_F - current of the stator
I_F' - additional variable - current
I_{Fpom} - stator supply current,
M_s - torque of the stator,
U_s - voltage of the control system

parameters are defined as follows:

I_z - moment of inertia reduced to the motor shaft
m - mass of the payload
R_z - effective radius of the mechanism
c_l - equivalent rigidity of the ropes
f_l - damping of the ropes
T_{F1},T_{F2} - time constants of element 2,
k_{IF} - conversion factor of element 2 and 4,
k_{wzm} - gain factor,
T_{IFpom} - time constant for element 4,
k_ω - factor of angular velocity conversion,
k_{Ms} - gain factor for element 3,
T_{Ms} - time constant for element 3.

Verified model of driving mechanisms for bridge and trolley is described in details in the paper [15]. Figure 3. 2 shows the model of overhead crane treated as a rigid body moving by planar motion with a trolley and hanged payload.

Thanks to mathematical transformations the following description of the overhead crane bridge and the trolley movements using the space of state variables notation was obtained:

$$\frac{dv_w}{dt} = \frac{\frac{M_{sw}}{R_{zw}}}{\left(m_w + \frac{I_{ow}}{R_{zw}^2}\right)} - \frac{W_w}{\left(m_w + \frac{I_{ow}}{R_{zw}^2}\right)} - \frac{H_w}{\left(m_w + \frac{I_{ow}}{R_{zw}^2}\right)} \tag{12}$$

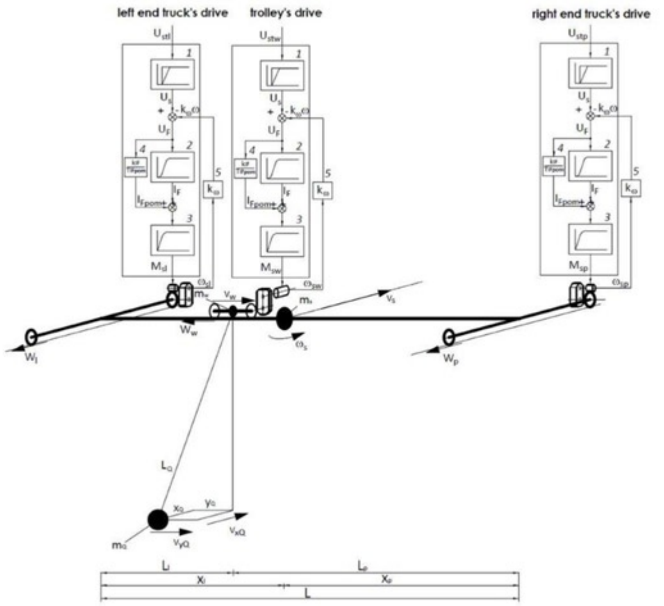


Fig. 3.2. The model of overhead crane drive system with a suspended payload

$$\frac{dx_w}{dt} = v_w \tag{13}$$

$$\frac{dv_s}{dt} = \left(\frac{1}{R_z\left(m_s + \frac{2 \cdot I}{R_{zr}^2}\right)} + \frac{R_{zr} - x_l}{R_z \cdot I_{zast}} \cdot R_{zr} \right) \cdot M_{sl} + \left(\frac{1}{R_z\left(m_s + \frac{2 \cdot I}{R_{zr}^2}\right)} + \frac{R_{zr} + x_p}{R_z \cdot I_{zast}} \cdot R_{zr} \right) \cdot M_{sp} + \tag{14}$$

$$\left(\frac{R_{zr} \cdot (x_l - L_l)}{I_{zast}} - \frac{1}{m_s + \frac{2 \cdot I}{R_{zr}^2}} \right) \cdot H + \left(\frac{R_{zr} \cdot x_l - R_{zr}^2}{I_{zast}} - \frac{1}{m_s + \frac{2 \cdot I}{R_{zr}^2}} \right) \cdot W_l + \tag{15}$$

$$- \left(\frac{R_{zr} \cdot x_p + R_{zr}^2}{I_{zast}} - \frac{1}{m_s + \frac{2 \cdot I}{R_{zr}^2}} \right) \cdot W_p \tag{16}$$

$$\frac{d\omega_s}{dt} = \frac{R_{zr} - x_p}{R_z \cdot I_{zast}} \cdot M_{sl} + \frac{R_{zr} + x_p}{R_z \cdot I_{zast}} \cdot M_{sp} + \frac{R_{zr} + x_l - L_l}{I_{zast}} \cdot H + \frac{R_{zr} + x_l}{I_{zast}} \cdot W_l + \frac{R_{zr} - x_p}{I_{zast}} \cdot W_p \tag{17}$$

where:

v_w - velocity of the trolley,
x_w - position of the trolley
M_{sw} - torque of the trolley's motor
R_{zw} - effective radius for trolley's mechanism,
W_w - friction of the trolley,
H_w - horizontal force caused by the payload oscillations, parallel to the axis of the trolley
m_w - mass of the trolley,
I_{ow} - moment of inertia for all parts of the trolley reduced to the motor's shaft
M_{sp}, M_{sl} - driver torques of both sides of the bridge,
W_p,W_l - friction of both sides of the bridge,
H - horizontal force caused by the oscillations of the payload
R_z - effective radius of end truck
L - range of the bridge,
L_p, L_l - position of the trolley related to both left and right end truck
x_p, x_l - position of the center of the mass related to both left and right end truck
I - moment of inertia of all rotating parts reduced to the motor's shaft
I_{zast} - effective moment of inertia.

Both end trucks and the trolley are driven by induction asynchronous motor described similarly as for the lifting mechanism. There are additional subscripts *l* and *p* for left and right end truck, and *w* for the trolley.

$$\frac{dx_s}{dt} = v_s \tag{18}$$

$$\frac{d\varphi_s}{dt} = \omega_s \tag{19}$$

$$\frac{dU_s}{dt} = k \tag{20}$$

$$dI_f' = \frac{k_{IF} \cdot k_{wzm}}{T_{F_1}^2} \cdot \left(U_s - k_\omega \cdot \omega_s\right) - \frac{T_{F_2}}{T_{F_1}^2} \cdot I_F' - \frac{1}{T_{F_1}^2} \cdot I_F \tag{21}$$

$$\frac{dI_f'}{dt} = I_F' \tag{22}$$

For presented mechanisms of the bridge and the trolley the following state variables were assumed:

v_w - velocity of the trolley,
x_w - position of the trolley
v_s - velocity of the bridge's center of the mass
ω_s - angular velocity of the bridge related to the center of the
x_s - position of the bridge's center of the mass,
ϕ_s - angular position of the bridge related to the center of the mass
$I_{F(l, p, w)}$ - current of the stator,
$I_F'{}_{(l, p, w)}$ - additional variable - current,
$I_{Fpom(l, p, w)}$ - additional current,
$M_{s(l, p, w)}$ - torque of the motor,
$\omega_{s(l, p, w)}$ - angular velocity of the motor's shaft,
$U_{s(l, p, w)}$ - control voltage

With assumption of no slip between the wheel and the trail velocities of motors were defined as follows:

$$\frac{dI_{F_{pom}}}{dt} = \frac{k_{IF}}{T_{I_{fpom}}} \cdot \left(U_s - k_\omega \cdot \omega_s\right) \tag{23}$$

$$\frac{dM_s}{dt} = \frac{k_{M_s}}{T_{M_s}} \cdot \left(I_F + I_{Fpom}\right) - \frac{1}{T_{M_s}} \cdot M_s \tag{24}$$

$$v_l = \frac{v_s - x_l \cdot \omega_s}{R_z} \tag{25}$$

$$v_p = \frac{v_s - x_p \cdot \omega_s}{R_z} \tag{26}$$

A constant value of energy consumed by the drive control system was measured using an intermediary system plugged between the power grid and the crane. The system, shown in the Figure 3.3 , consists of the three – phase current transformer and the programmable transducer. These two devices are governed by the software installed in the notebook.

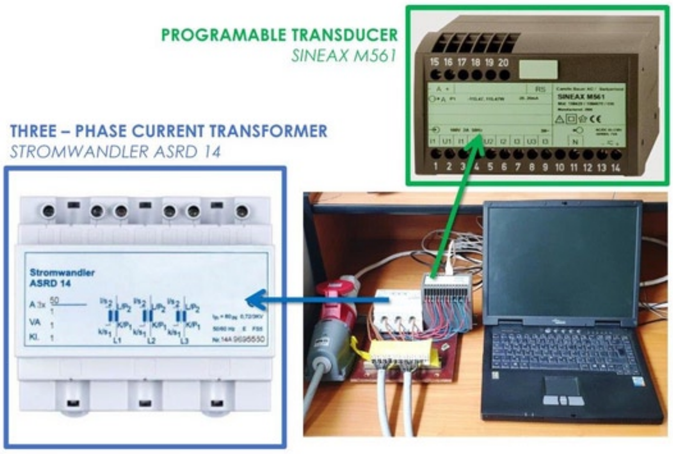


Fig. 3.3. Intermediary measuring system plugged between the crane's drive control and the power grid

4. Experimental stand

The experimental stand, shown in the Figure 4.1., considered in this article, is based on the overhead crane, where main parts of the crane as bridge, trolley and hoist, are an inverter – fed mechanisms. All inverters are supplied by the DC Link, where the voltage fluctuates from 560 to 780 V. To return the energy, which is gained from change the potential energy during lowering of a payload or during braking another mechanisms, to the grid, all drives are supplied using the regenerative power supply unit. The system is governed by the PLC controller, which allows control all crane mechanisms and also monitors some parameters of system's operation using software installed on the portable computer. Measurements of the actual energy demand of the crane is provided by using a power measuring system, plugged between the overhead crane system and the power grid.

5. Results of duty cycles investigation

Numerical models of the crane mechanisms and drives allow for a wide inspection of the overhead crane for example by investigating the payload trajectory while taking into account existing obstacles and also for analyzing the energy consumption.

Maximum range of the overhead crane displacement in the direction of movement of bridge and the trolley are 20 m and 8 m, respectively. The trajectory of the payload was determined by the application calculating safe path for the payload in the hall. The application is widely described in paper [15].

Using the energy measuring system, which is depicted in the Figure 3.3, the power consumption was investigated in a few cases of the overhead crane's operation. Results of this investigation can be observed in the Figure 5.1. The measurement cycle was divided into phases of standby of the crane and also phases when each mechanism

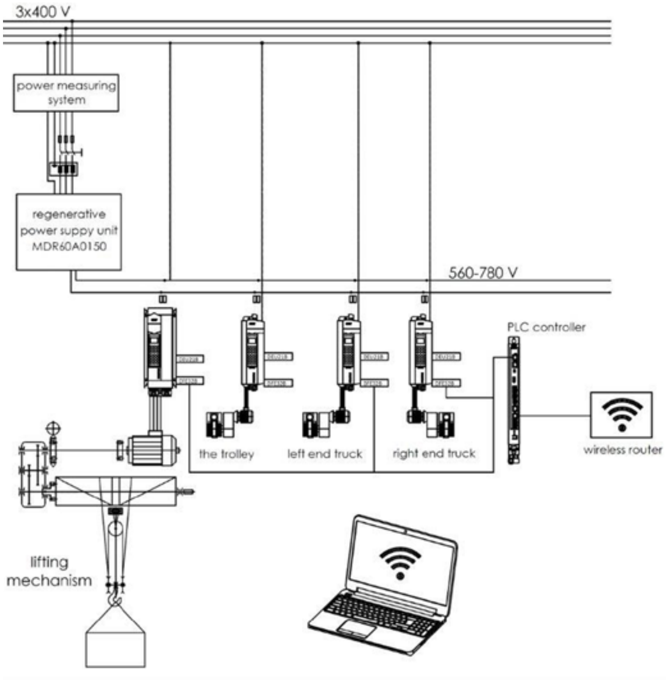


Fig. 4.1. The experimental stand based on the overhead crane

is activated. This means that the inverter is turned into *run allowed* mode with a simultaneous release of each brake. Green solid line shows actual power consumption of the crane, which was recorded during the operation. Violet solid line represents an average value of the actual power for individual phases of operation. Red solid line represents a power demand necessary for release the brakes of each mechanism taken from the catalog data.

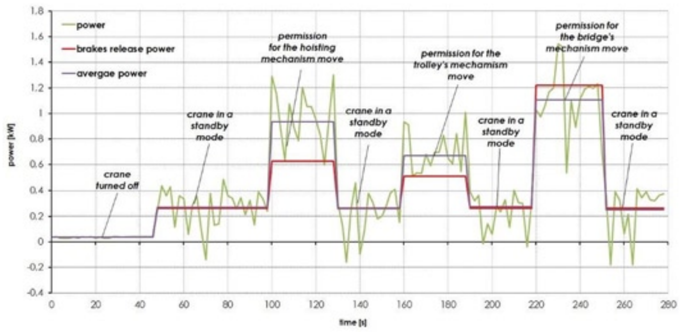


Fig. 5.1. Power demand of the overhead crane.

Based on the numerical investigation, and also on the inspection with real energy consumed, an energetic demand of the overhead crane for a few types of motion while the duty cycle was estimated. Results of the simulation presented in the Figure 5.2 shows the energy consumption related to the travelled distance for bridge and trolley mechanism. The investigation was carried out for motions with an empty hook and also with 5 ton weight payload. Dashed lines show a theoretical energetic demand of the overhead crane as a result of numerical simulation. The solid lines (with *_C* suffix) show a theoretical energetic demand corrected by the average value of the energy consumed by the control unit and brakes.

Thanks to the consideration of the hybrid model, it can be concluded that the energetic demand of the overhead crane increases significantly in comparison to theoretical simulations. The biggest change in the energy consumption can be observed in the case of the trolley's move with an empty hook, where energetic demand raises from 6 kJ to 18,5 kJ.

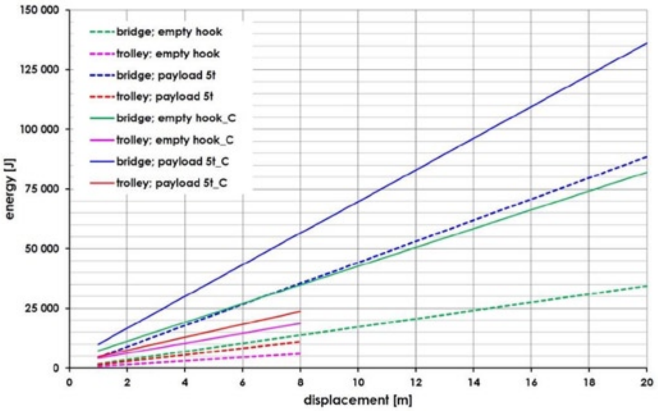


Fig. 5.2. Energy consumption of the overhead crane for the bridge and trolley mechanisms related to the travelled distance for different loading conditions

The presented above research allows estimate the real energy consumption of mechanisms responsible for horizontal displacement of payload. It is visible that the energy necessary for the operation of the control system and brake release can exceeds the energy consumed directly by drives. Performing research on energy consumption of drives equipped with motors with brakes fed by inverters should consider additional power demand which in real systems exist and are necessary to its operation and this applies not only to overhead cranes.

Next step of analysis was an investigation of the hoisting mechanism. It is the main part of an overhead crane's energetic optimization because of the huge with respect to other mechanisms amount of energy demand . Charts presenting energy consumption in the function of hoisting/lowering distance are shown in the Figure 5. 3. The investigation also includes energy consumption of the drive control system and brake release system.

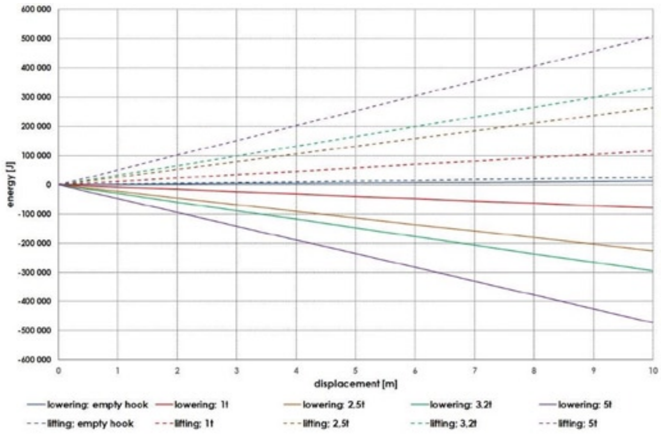


Fig. 5.3. Energy consumption of winch mechanism in relation to the travelled distance (lifting and lowering) for different loading conditions.

Analyzing results of the experiment it is easy to observe that the energy consumption changes for different loading conditions. he energy consumption during hoisting rises up with increasing payload mass. The energy necessary to lift an empty hook is 10 kJ and increases to the magnitude over 500 kJ while lifting the 5 ton mass payload. Negative values of the energy prove that there is a possibility to recover the energy of gravity stored in the payload, which now is dissipated as a heat into resistors installed in the crane. The graph does not show symmetry about the horizontal axis due to the additional energy necessary for the operation of control systems.

The value of recoverable energy Er in relation to the energy put in hoisting a payload Eh was determined. This proportion changes with the change of mass mQ of the payload what is visible in Figure 5.4. The changes range is from 0.5 for an empty hook to -0.93 for a nominal payload mass. Negative values mean, that energy recuperation, proportional to energy consumption for hoisting the particular payload, is possible. Positive values appeared when for properly operation of the system, energy must be supplied. In presented figure, the limit point , when recovery becomes possible, is 7% of the weight of the payload.

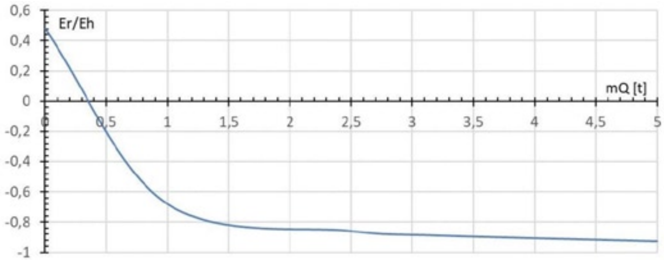


Fig. 5.4. The possibility of energy recuperation during lowering with respect to energy consumption during hoisting.

Due to the fact of buffering of significant amount of the energy in a suspended payload, the investigation was carried out for a typical duty cycle, which was compared to the similar one with possibility of energy recovery. Figures 5. 5 and 5.6 depict both of duty cycles.

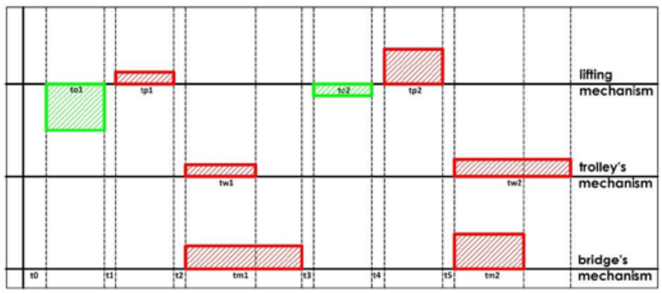


Fig. 5.5. Crane's duty cycle with possibility of restoring the energy

Presented typical duty cycle consists of lowering the payload from highest transportation level to the ground, next the hoisting the empty hook at the same level. Then simultaneous movement of trolley and bridge to the loading point and lowering empty hook. Next hoisting full load at transportation level and then return to the unloading point.

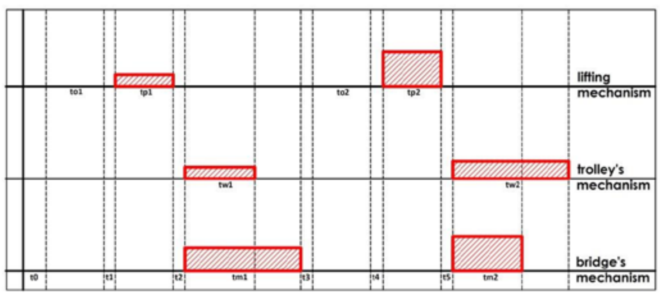


Fig. 5.6. Crane's duty cycle without possibility of restoring the energy

Comparison of energy consumption for both cycles is presented in the Figure 5. 7. Negative values show the possibility of energy restoration and potential savings as a consequence.

Analyzing the data in the Figure 5. 7 it could be observed that significant amount of the energy can be restored in the hoisting mechanism during the lowering of the suspended payload. The energetic demand in a standard cycle is about 610 kJ of the electrical energy.

Energy consumption of the system with the possibility of its restoration is about 360 kJ which gives a potential energy savings of 40%.

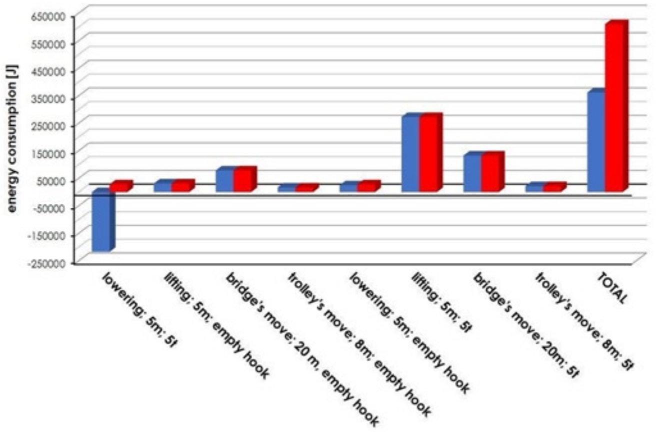


Fig. 5.7. Comparison of two types of duty cycle both with the energy saving system (blue) and without (red)

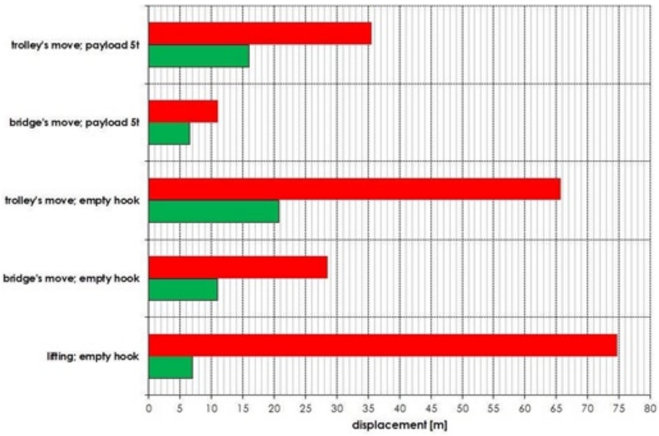


Fig. 5.8. Potential savings after lowering the 5 tons payload from the 5 meters height.

In order to illustrate the possibilities of using the energy of the hoisting mechanism, the possible travel lengths of individual mechanisms of an overhead crane using the same amount of energy from lowering 5 t on a road 1 m are presented in Figure 5.8.

Red bars shows movement possibilities of particular mechanisms according only to the numerical analysis. The green ones include also values of the energy consumed by the drives and brakes control systems.

6. Conclusion

This paper presents the analysis of overhead crane mechanisms operation for energy consumption. This analysis is based on the hybrid model developed using kinetic and dynamic relationships and also typical equations of dynamic elements and experimental data of the control system's energy consumption. The maximum range of crane travel was investigated in typical duty cycle for different loading conditions. It can be concluded that for the nominal payload 40% of the energy in overhead crane's duty cycle can be recovered.

This phenomenon can be observed mainly in the hoisting mechanism, but also during deceleration phases of traveling mechanisms. The analysis shows that this mechanism can provide significant energy savings during the lowering of the suspended payload even considering the energy consumed by control system and brakes. A significant share of the energy during the crane's operation is consumed to the release of electromagnetic brake based on the coils. In case of travelling or traversing mechanism energy consumption level of this system is several dozen percent of the energetic demand of the drive. Energetic demand of inverters and control system also cannot be neglected. It can be concluded that the investigation, which including also the energetic consumption of the control system and brakes, gives the full view of the crane's energy consumption.

Results of this investigation are important especially in the context of sustainable production. This seems to be important also from the economic and ecologic point of view. Increasing the energy efficiency of devices (including transport) is one of the directions of development of an environmentally friendly economy.

Results provided in this article can be considered in the context of implementing an energy storing or energy returning system to the grid. Magnitude of energy which is able to put to use again can be approximated as significant percentage of the energy needed in the whole duty cycle. In case of returning the energy back to the power grid can be consumed by other devices. On the other hand, engaging the energy storage, the energy can be consumed to finish the duty cycle during the loss of the electrical power network supply.

Acknowledgement

The fourth and fifth authors would like to thank for the financial support from the project co – financed by the European Union under the European Social Found as a part of the Operational Program: Knowledge, Education, Development, project No. POWR.03.02.00 – 00 – I042/16 – 00.

References

1. Ahmad NBb, Su C, Zhaoxia X, Vasquez JC, Guerrero JM. Modeling and Controls of Flywheel Energy Storage Systems for Energy Harvesting from Harbor Electrical Cranes. In Proceedings of the 2018 IEEE Industry Applications Society Annual Meeting (IAS) 2018: 1-8, https://doi.org/10.1109/IAS.2018.8544652.
2. Ahmad NBb, Su C, Zhaoxia X, Vasquez JC, Guerrero JM, Liao C. Energy Harvesting From Harbor Cranes With Flywheel Energy Storage Systems. IEEE Transactions on Industry Applications 2019; 55 (4): 3354 - 3364, https://doi.org/10.1109/TIA.2019.2910495.
3. Chang C, Yang J, Cao P, Niu Q, Zhu Z. Study on super capacitor energy saving system for rubber-tyred gantry crane. In Proceedings of the Power and Energy Engineering Conference (APPEEC), 2010 Asia - Pacific. 2010: 1-6, https://doi.org/10.1109/APPEEC.2010.5448850.
4. Chang C, Wang R, Li, Zhu Z, Xu H. Research of supercapacitor voltage equalization strategy on rubber - tyred gantry crane energy saving system. Energy and Power Engineering 2009; 2(1): 1 - 4, https://doi.org/10.1109/APPEEC.2009.4918668.
5. Corral - Vega PJ, Fernandez - Ramirez LM, Garcia - Trivino P. Hybrid powertrain, energy management system and techno - economic assessment of the rubber gantry crane powered by diesel - electric generator and supercapacitor energy storage system. Journal of Power Sources 2019; 412: 311 - 320, https://doi.org/10.1016/j.jpowsour.2018.11.027.
6. Directive (EU) 2018/410 of the European Parliament and of the European Council of 14 March 2018 to enhance cost - effective emission reductions and low - carbon investments. Official Journal of the European Union.

7. Duc TH, Suzuki K, Tsume M, Tasaki R, Miyoshi T, Terashmia K. A switched optimal control approach to reduce transferring time, energy consumption, and residual vibration of payload's skew rotation in crane systems. Control Engineering Practice 2019; 84: 247 - 260, https://doi.org/10.1016/j.conengprac.2018.11.018.
8. Flynn MM, McMullen P, Solis O. High-Speed Flywheel and Motor Drive Operation for Energy Recovery in a Mobile Gantry Crane. In Proceedings of the 22nd Annual IEEE Applied Power Electronics Conference 2007: 1151-1157, https://doi.org/10.1109/APEX.2007.357660.
9. Flynn MM, McMullen P, Solis O. Saving energy using flywheels. IEEE Industry Applications Magazine 2008; 14 (6): 69 - 76, https://doi.org/10.1109/MIAS.2008.929351.
10. Honczarenko J, Berliński A. Modelowanie energochłonności procesów transportowych w zautomatyzowanych systemach montażowych. Technologia i automatyzacja montażu 2011; 4: 49 - 52.
11. Ianuzzi D, Piegari L, Tricoli P. Use of supercapacitor for energy saving in overhead travelling crane drives. In Proceedings of the International Conference on Clean Electrical Power 2009: 562 - 568, https://doi.org/10.1109/ICCEP.2009.5211983.
12. International Standard IEC 60034 - 1. Rotating electrical machines - Part 1: Rating and performance.
13. Kosucki A, Malenta P, Stawiński Ł, Halusiak S. Energy consumption and overloads of crane hoisting mechanism with system of reducing operational loads. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19 (4): 508-515, https://doi.org/10.17531/ein.2017.4.3.
14. Kosucki A, Malenta P. The possibilities of reducing the operational load of hoisting mechanisms in case of dynamic hoisting. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18 (3): 390-395, https://doi.org/10.17531/ein.2016.3.10.
15. Kosucki A. Badanie transportu ładunków przy wykorzystaniu skojarzonych ruchów mechanizmów suwnic pomostowych. Rozprawy Naukowe Nr 1175 Zeszyty Naukowe 474 Politechnika Łódzka 2013
16. Li Z, Wang G, Shen K, Zhao T, Ni X. Research of the Crane System Based on Technology of Storing - Energy with Super-capacitor. In Proceeding of the International Conference on Electrical and Control Engineering 2010: 4046-4049, https://doi.org/10.1109/iCECE.2010.984.
17. Parise G, Honorati A, Parise L, Martirano L. Near zero energy load systems: The special case of port cranes. In Proceedings 2015 IEEE/IAS 51st Industrial & Commercial Power Systems Technical Conference (I&CPS) 2015: 1 - 6, https://doi.org/10.1109/ICPS.2015.7266430.
18. Repo A, Montonen J, Sizonenko V, Lindh P, Pyrhonen J. Energy efficiency of hoisting motors. In Proceedings of the International Conference on Electrical Machines (ICEM) 2014: 144 - 149, https://doi.org/10.1109/ICELMACH.2014.6960172.
19. Sikora M, Szczybra K, Wróbel Ł, Michalak M. Monitoring and maintenance of a gantry based on a wireless system for measurement and analysis of the vibration level. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (2): 341 - 350, https://doi.org/10.17531/ein.2019.2.19.
20. Yifei T, Zhaohui T, Wei Y, Zhen Y. Research on energy-saving optimization design of overhead crane. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2013; 15 (4): 449 - 457.
21. Zelic A, Zuber N, Stostakov R. Experimental determination of lateral forces causes by bridge crane skewing during travelling. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (1): 90 - 99, https://doi.org/10.17531/ein.2018.1.12.
22. Zhou W, Xiaohua X. Energy efficiency of overhead cranes. Preprints of the 19th World Congress, The International Federation of Automatic Control 2014: 19 - 24, https://doi.org/10.3182/20140824-6-ZA-1003.00961.

Andrzej KOSUCKI
Łukasz STAWIŃSKI
Piotr MALENTA
Jakub ZACZYŃSKI
Justyna SKOWROŃSKA
Lodz University of Technology
Institute of Machine Tools and Production Engineering
ul. Stefanowskiego 1/15, 90-924 Łódź

E-mails: andrzej.kosucki@p.lodz.pl, lukasz.stawinski@p.lodz.pl, piotr.malenta@p.lodz.pl, jakub.zaczynski@dokt.p.lodz.pl, justyna.skowronska@dokt.p.lodz.pl

Article citation info:
TABASZEWSKI M, SZYMAŃSKI GM. Engine valve clearance diagnostics based on vibration signals and machine learning methods. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 331–339, http://dx.doi.org/10.17531/ein.2020.2.16.

Maciej TABASZEWSKI
Grzegorz. M. SZYMAŃSKI

ENGINE VALVE CLEARANCE DIAGNOSTICS BASED ON VIBRATION SIGNALS AND MACHINE LEARNING METHODS

DIAGNOSTYKA LUZU ZAWORÓW SILNIKA SPALINOWEGO Z WYKORZYSTANIEM SYGNAŁU DRGANIOWEGO I METOD UCZENIA MASZYNOWEGO*

A dynamic advancement of the design of combustion engines generates a necessity of introduction of strategies of operation based on the information related to their technical condition. The paper analyzes problems related to vibration based diagnostics of valve clearance of a piston combustion engine, significant in terms of its efficiency and durability. Methods of classification have been proposed for the assessment of the valve clearance. Experiments have been performed and described that aimed at providing information necessary to develop and validate the proposed methods. In the performed investigations, the vibration signals were obtained from a triaxial accelerometer located in the engine cylinder head. A parameterization of the obtained vibration signal has been carried out for the engine operating under different engine loads, rotation speeds and valve clearance settings. The parameterization pertained to the specific features of the vibration signals, the derivative of the vibration signal as a function of time as well as the envelope of this derivative. In the first approach, the authors developed a classifier in the form of a set of binary trees that additionally allowed distinguishing the features significant in terms of the identification of adopted classes. For comparison, the authors also developed classifiers in the form of a neural network as well as a k-nearest neighbors algorithm using the Euclidean metric. Based on the performed investigations and analyses a method of valve clearance assessment has been proposed.

Keywords: *Combustion engine, diagnostics, vibration, machine learning***.**

Dynamiczny rozwój konstrukcji silników spalinowych generuje potrzebę wprowadzenia strategii eksploatacji jednostek napędowych, opartej na znajomości ich stanu technicznego. W artykule poddano analizie zagadnienia, związane z drganiową diagnostyką luzu zaworów tłokowego silnika spalinowego, istotnego ze względu na efektywność pracy silnika i jego trwałość. Zaproponowano wykorzystanie metod klasyfikacji do oceny poprawności luzu zaworowego. Przeprowadzono i opisano eksperymenty, które miały na celu dostarczenie informacji koniecznych do zbudowania i zweryfikowania zaproponowanych metod. W przeprowadzonych badaniach pozyskano sygnały drganiowe z trójosiowego czujnika przyspieszeń drgań zlokalizowanego na głowicy silnika. Dokonano parametryzacji uzyskanych przebiegów czasowych sygnału drganiowego dla silnika pracującego pod różnym obciążeniem, z różnymi prędkościami obrotowymi oraz z różnymi luzami zaworowymi. Parametryzacja dotyczyła zarówno cech sygnału przyspieszeń drgań, pochodnej przyspieszeń drgań względem czasu jak i obwiedni tej pochodnej. W pierwszym podejściu zbudowano klasyfikator w postaci zbioru drzew binarnych, który przy okazji pozwolił na wyodrębnienie istotnych, ze względu na przyjęte klasy, cech. Dla porównania zbudowano także klasyfikatory w postaci sieci neuronowej jak i algorytmu k – najbliższych sąsiadów z metryką euklidesową. Na podstawie przeprowadzonych badań i analiz zaproponowano metodę oceny luzu zaworowego.

Słowa kluczowe: *silnik spalinowy, diagnostyka, drgania, uczenie maszynowe.*

1. Introduction

Internal combustion engines are commonly applied in vehicles and stationary equipment. They convert the energy contained in the fuel into mechanical work of the rotating crankshaft and, like all mechanical equipment, are subject to wear and tear and aging. The engine durability is described with design properties and, to a great extent, depends on the conditions of operation and the nature of the loads. As the engine degradation processes advance (variable temperatures, tribological processes, cavitation, chemical and electrochemical corrosion, aging etc.) the reliability and efficiency parameters deteriorate. As a consequence, the object wears, fails or is withdrawn from operation for economic or environmental reasons. Ever since the beginning of combustion engines, it has been observed that one of the key problems having impact on the engine operation is its correct adjustment.

Degradation of the engine structure and engine incorrect adjustment may lead to the following phenomena in the internal combustion engine: deterioration of the engine efficiency, reduction of power, related to the reduction of mechanical efficiency, thermal efficiency and filling coefficient, increase in the emission of toxic compounds in the exhaust fumes, and possible damage to the engine components.

Fig. 1a presents the differences in the fuel consumption for different valve clearance adjustments, and Fig. 1b presents the changes in the impact velocities of the valve against the valve seat allowing for the cam lift h). From the data shown in Fig. 1a, it follows that the changes in the valve clearance may lead to increased fuel consumption by the investigated internal combustion engine by approx. 9%, while the analysis of Fig. 1b leads to a conclusion that with the increasing valve clearance (lines: blue and green Fig.1b), the velocity of the impact of the valve against the valve seat grows, causing additional unwanted dynamic loads on the engine cylinder head of the engine.

There are many methods of diagnosing the technical condition of combustion engines. They can be divided into methods utilizing

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

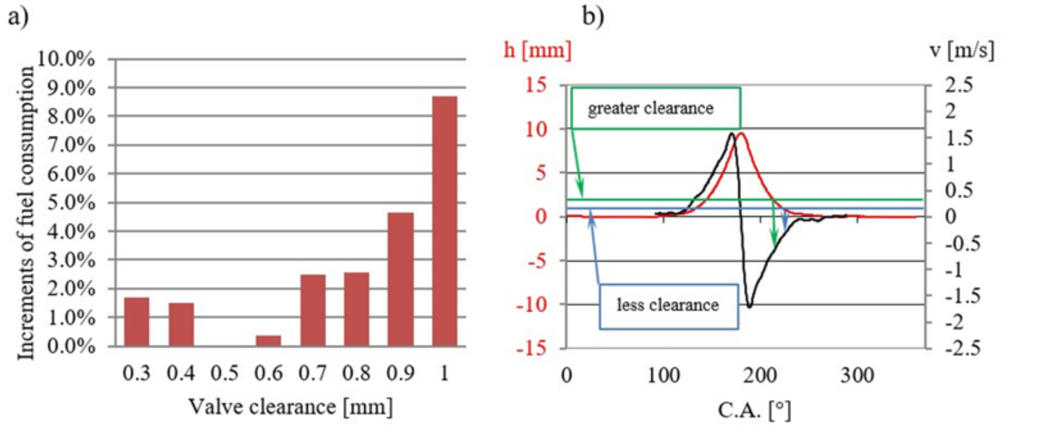


Fig. 1. a) Increments of fuel consumption as a function of valve clearance, b) cam curve (red line), velocity of the point on a cam (black line) and changes in the impact velocity of the valve against the valve seat in the engine caused by the changes in the valve clearance settings (vertical arrows: green and blue), C.A.– crankshaft angle, h – cam lift, v – velocity of the point on a cam

the operating processes (indicating, torque variations as a function of crankshaft angle, measurement of the exhaust fumes temperature and pressure above and underneath the piston, fueling parameters, exhaust opacity etc.) and residual processes (vibration, noise, thermal processes, electrical processes etc.). Based on the investigations into the operating processes, we may infer the overall condition of a combustion engine, while the residual processes carry information on the condition of individual subassemblies and kinematic pairs, which is why the residual processes are utilized as autonomous or auxiliary diagnostic methods. All methods based on vibration and noise analysis aiming at determining the technical condition of an object are called vibroacoustic diagnostics.

The engine timing mechanism, as one of the fundamental components of piston combustion engines, as a matter of the mere principle of its operation, is a source of vibroacoustic signal because the operation of the timing system, which includes such actions as valve opening and closing, cooperation of the cam with the valve lifters, canceling of the play between them etc. result in multiple impacts of the mating components, which naturally generates vibration. The use of vibration signals to diagnose the internal combustion engine timing system is presented in the following works:

Article [12] discusses the option of applying the distribution of the wavelet packet when filtering the acoustic signal of an internal combustion engine in order to diagnose excessive valve clearance. The authors have prepared an algorithm enabling selection of selected details and approximation of the wavelet analysis to low-frequency components, which constitute noise, as well as to high-frequency components, containing information about possible increase of the engine valve clearance. Then, based on selected components of the high-frequency acoustic signal, a method was developed for automatic detection of increased valve clearance, assuming that the energy share of the emitted acoustic signal should be determined when opening and closing individual valves. The authors of [13] have described research on the diagnosis of the condition of exhaust valves in large marine diesel engines. The tests were carried out on a four-cylinder two-stroke marine diesel engine with a piston diameter of 500 mm at MAN B&W Diesel Research Centre in Copenhagen, Denmark. The experiments involved three different states of the valve, two of which concerned artificially induced valve burnout situations. The basic monitoring measurements were vibration and structural stress waves, also known as acoustic emission (AE). The results showed that AE signals have a significant advantage over the other sensors involved, indicating sensitivity to both mechanical action and mechanical-smooth combustion. The recorded data has been pre-processed and the functions extracted using PCA (principal component analysis). On the basis of a number of applied heuristics and statistics, the search for the optimal subspace of the main components was conducted. The selected feature space has been used to classify the condition of the exhaust valve by using both standardized neural feedback classifiers and linear discriminators. Article [28], however, calculates cone-kernel distributions (CKD) of vibration acceleration signals obtained from the cylinder head in eight different states of the camshaft mechanism and presents them in gray images. Non-negative matrix factorization (NMF) was used to decomposing multidimensional data, and neural network ensemble (NNE), which has a better generalization capacity for classification than a single neural network, was used to make an intelligent diagnosis based on time-frequency distributions. On the basis of the experimental results, it has been shown that damage to the diesel engine's timing mechanism can be accurately classified using the proposed method. Paper [31] analyzes the issues related to vibration diagnostics of automatic clearance compensators for valves of internal combustion engines. Scientific experiments have been described and carried out to provide the information necessary to build and verify diagnostic models to assess the technical condition of those components of an internal combustion engine which are essential for engine performance and durability. Based on the obtained diagnostic model, an algorithm for evaluating the technical condition of automatic valve clearance compensators was developed. Whereas in [32] Wigner-Ville distributions (WVD) of vibration acceleration signals were calculated, which were recorded on the cylinder head in eight different valve train states and displayed in gray images. After their standardization, probabilistic neural networks (PNNs) were directly used to classify time-frequency images. In this way, the diagnosis of damage to the timing mechanism has been transferred to the classification of time-frequency images. Experimental results show that faults in diesel valve assemblies can be accurately classified using the proposed methods.

The application of various techniques of analyzing vibroacoustic signals to assess the technical condition of internal combustion engines was presented by the authors of the following works:

The authors of [1] have shown that the acoustic signals coming from a combustion engine are rich in information concerning the engine operation parameters and its condition. Unfortunately, this information is complex and has a lot of background noise. Based on the analysis of the measured parameters, they showed that engine failures caused by a drop in compression ratio, changes in injection pressure, changes in the exhaust system, changes in the suction and exhaust valve clearance can be diagnosed. The engine's technical condition was assessed on the basis of the analysis of changes in the rms value and the kurtosis of acoustic signals from each engine cylinder. The authors of [2] stated that acoustic signals caused by mechanisms often have to be described by non-linear models in the time domain. In the frequency domain, however, the linear model is in many cases sufficient to describe sound propagation channels. In the paper they compared the calculation methods in terms of accuracy, computation time and the possibility to perform tests during the operation of the facility. Article [3] presents a fast and automatic method of engine diagnostics based on one acoustic emission (AE) parameter. The method is based on a comparison of vibrations and AE energy with reference values to determine whether the state of the engine is defective. The method was used in the test engine and proved to be satisfactory. Work [9] concerns the monitoring of large diesel engines by analyzing changes in crankshaft angular speed. The focus was on a 20-cylinder diesel engine with natural crankshaft frequencies in the operating speed range. Angular speed changes were modelled at the free end of the crankshaft. Modelling included both the dynamic behavior of the crankshaft and the moments of excitation. Since the engine is very large, the first crankshaft turning modes are in the low frequency range. A model with a flexible crankshaft is required. The moments of excitation depend on the pressure curve in the cylinder. The latter was modelled using a phenomenological model. The mechanical and combustion parameters of the model were optimized using real data. An automated diagnosis based on a system using artificial intelligence has been proposed. Neural networks were used to recognize angular velocity patterns under normal and faulty conditions. The reference patterns required at the training stage were calculated from a model, calibrated using a small number of actual measurements. Promising results were obtained. During the verification tests, damage consisting of a fuel leakage was successfully diagnosed. In [10], a coupled simulation of piston dynamics and engine tribology (tribodynamics) was performed using quasi-static and transient numerical codes to model piston impacts on the cylinder wall. Confirmation of suitability of the proposed methods was determined on the basis of experimental measurements carried out on a single-cylinder petrol engine in laboratory conditions by measuring the vibration acceleration of the engine block surface. The authors of [34] proposed a system for the diagnosis of damage to combustion engines using wavelet packet transform (WPT) and artificial neural network (ANN) techniques. Article [5] discusses the application of the ionic current sensor for detecting combustion resonance in a direct injection diesel engine. A modified glow plug is used to measure ionic current in addition to its main function of heating the combustion chamber. Comparison was made of combustion resonance determined from the signals of the ionic current sensor, cylinder pressure transducer and engine vibration sensor. It has been found that the ionic current signal can be used to determine synchronization, amplitude, frequency and duration of resonance. The sensor output can be used as a feedback signal to the ECU (electronic control unit) to minimize engine vibration and noise. Article [8] concerns state-of-the-art diagnostic strategies and techniques based on vibroacoustic signals which can be used to monitor and diagnose internal combustion engines (ICEs) both on the test bench and under operating conditions. This article presents for the first time a short summary of sound and vibration generation in ICE in the context of further discussion on vibroacoustic diagnostics. A review of monitoring and diagnostic techniques described in the literature using vibration and acoustic signals is also presented.

On the basis of the analysis of the achievements to date in the field of vibroacoustic diagnostics of internal combustion engine systems, it has been concluded that research was carried out on the application of parameters of vibration signals to assess the technical condition of internal combustion engine assemblies or processes occurring therein. The research concerned methodological issues (e. g. determination of engine operating conditions during vibration measurements, selection of measurement points) and modelling issues (building diagnostic models and their validation). In the investigations both simple methods of signal description (e.g. point measures) as well as highly advanced techniques of signal processing (e.g. artificial neural networks, time-spectrum analysis) were used. In the analyzed works on the vibration diagnostics of internal combustion engines, methods were recognized in which diagnostic models were based on responses of the object structure to impulse excitations. The analyses related to the responses provided a hint as to the optimal choice of the measurement point and possibly the frequency range to be covered by the analysis.

This paper presents a new approach to identification of valve clearance of an engine operating at different loads and speeds based on an appropriately processed vibration signal. A methodology of valve clearance class identification has been proposed based on the absolute vibration accelerations measured on the cylinder head and a supervised learning systems - classifiers. Such a solution allows an automatic assessment of the correctness of the valve clearance adjustment on an operating engine without the necessity of seeking a mathematical model describing the relation between the vibration signal and the valve clearance.

Classifiers, as supervised taught systems, are used in a wide range of areas to process very large data resources and automate the inference process. It is impossible to fully review the applications of these algorithms, or even their fields of application. For example, we can only mention such different areas as: prediction of students' exam results [6], monitoring of urban changes [11], classification of road roughness [15], segmentation of apple defects [19], intelligent system of rotor machine damage detection [20], classification of network traffic [24], or image processing and analysis [35]. There are many known attempts to use machine learning in machine and component diagnostics to determine the technical condition as well as the characteristics of the condition. For example, in [16] the use of a convolution network with some modifications for the classification of machine condition is considered. The authors presented promising results of the method on the example of signals from rolling bearings. In [18], the authors successfully used k-means clustering and classification using the SVM (Support Vector Machine) method to assess the state of wear of packaging machine blades. In [23], statistical distance measures were used to distinguish the states of damage to rolling bearings. The authors of [27] demonstrated the effectiveness of classification methods with respect to monitoring the condition of piston compressors installed in refrigeration equipment. Among others, neural networks were used here, but also extreme learning machines (ELMs). In the study on wind turbine diagnostics [29], time signal representations in the form of images were used, various texture characteristics were used and classified using these characteristics. A discussion of many machine learning methods in the context of general diagnostic applications can also be found in [17].

Machine learning methods were also used in the diagnostics of internal combustion piston engines. Paper [4] compares different classification methods used to identify the phenomenon of ignition loss. The authors of [7] diagnosed the state of the motor injectors using a vibration signal, discrete wavelet transform and neural network. In turn, in [21] the problem of engine cylinder shutdown was analyzed using the same methods. Many failures of the combustion engine were also identified using probabilistic methods [33]. However, the problem of non-optimal valve clearance was not considered here. In [14] classifications were used to identify many damages - including too small or too large valve clearance. The research was conducted using advanced methods of extreme machine learning. The analysis of the work shows that the accuracy of damage classification obtained by the researchers does not exceed 96%. It seems that it is desirable to find a simpler method for classification of clearance due to the possibility of its easy practical implementation in on-board diagnostics, therefore, in the opinion of the authors of this paper, further search and research in this direction is necessary. It is also important to maintain high reliability of the diagnosis despite the simplicity of the method.

The classification process can be technically carried out using a number of methods with specific properties and possibilities. Among a number of possibilities, you can mention classification trees, neural networks, distance classifiers, approximation classifiers, fuzzy classifiers, etc. There are three methods used in this paper which, according to the authors, are the easiest to implement in the on-board engine diagnostics: a classification tree, artificial neural networks MLP and a classifier of k-nearest neighbors.

The advantage of the tree structure is the human friendly way of representation of knowledge that can be obtained after the learning process. Another advantage is the lack of requirements for assumptions about the relationship between the dependent variable and the explanatory variables. In addition, the tree also allows distinguishing those diagnostic measures that are significant in the process of conditions classification. Those that will not be used by the algorithm are of no importance to the construction of the tree, hence, to the classification of the state or identification of malfunctions with the use of this method. Additionally, this method can be used in data sets containing numerous data shortages, which may be important in the case of databases from several sources.

Other classification methods used also in engine diagnostics are neural networks. Artificial neural networks, used here for classification, allow parallel processing of information. In the case of artificial neural networks, it is essential to optimize the network structure (number of hidden layers or number of neurons in the layers), which is an arduous process carried out mostly by trial and error method. The selection of input attributes is also important, which, in the case of classification trees, are built into the tree algorithm.

Another classifier used in this paper is the k-nearest neighbors classifier. Its advantage is undoubtedly the simplicity of implementation and easy-to-determine prediction of a given class. The downsides are the need to store big data in the memory and the need to optimize both the selection of parameter k, the measures of distance between the data and the set of attributes.

2. Methodology of research

The object of the investigations was a single cylinder research engine (SB 3.1) based on the SW 680 one. The SB 3.1 engine was designed for research purposes related to combustion and parameter assessment of SW 680 engines licensed from Leyland and manufactured in WSK Mielec. In the research engine, the following assemblies of the SW 680 engine were applied: connecting rod, piston with piston rings, cylinder sleeve, valves, timing system, injectors, cylinder head (adapted from the SW 680 engine). The design of the engine allows: measurement of the in-cylinder pressure, adjustment of the compression ratio in the range ε = 14-20, continuously variable onset of fuel pumping, continuous variable timing and changes in the balancing of the mass forces of the first order.

The investigations were carried out in two stages. In the first stage, impulse tests were performed aiming at identification of the resonance frequencies, which determined the measurement range and allowed determining the points of acquisition of vibration on the cylinder head. In the second stage, the vibration of the cylinder head was investigated for different settings of the valve clearance and the settings of the engine work point. Based on the results of the second stage, an algorithm of the valve clearance assessment in the investigated engine was developed. An overall diagram of the measurement system used for the recording of the vibration signals has been shown in Fig. 2.

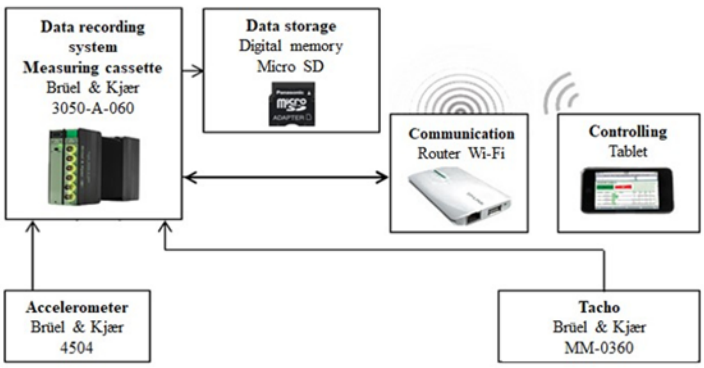


Fig. 2. Diagram of the system used for the measurement of the cylinder head vibration during engine operation

The research methodology was developed based on the assumptions of an active experiment [22]. During the experiment the following were adjusted: valve clearance, engine load and speed while recording the vibration accelerations of the engine cylinder head.

The measurements were carried out according to the principle of three starts, i.e. each series of measurements was performed three times and between each series of measurements the engine was shut off. The described method was used to avoid incidental values of the parameters of the vibration signal characteristics.

Vibration transducers (type 4504) by Brüel&Kjær were selected based on suggestions in [26,30] and the linear frequency range of selected transducers was 18 kHz. During the investigations, signals in the range 0.1 Hz–25 kHz were recorded. The accelerometers were fixed on the cylinder head with a glue. When selecting the measurement spots for the impulse tests, a principle was adopted that a transducer should be located in a accessible area possibly closest to the point where the vibration signal related to the valve operation is generated [25]. The directions of the vibration measurement were adopted as follows: direction X parallel to the diameter of the cylinder, direction Z parallel to the axis of the cylinder, direction Y perpendicular to the previous two (Fig. 3a). The sampling frequency was set at 65536 Hz. For the recording of the vibration signals, PULSE multianalyzer by Brüel&Kjær was used. It allows a parallel recording of fast-varying processes on 6 channels with the dynamics of up to 160 dB.

The special orientation of the transducers has been shown in Fig. 3a and the exact location of their fitting in Fig. 3b.

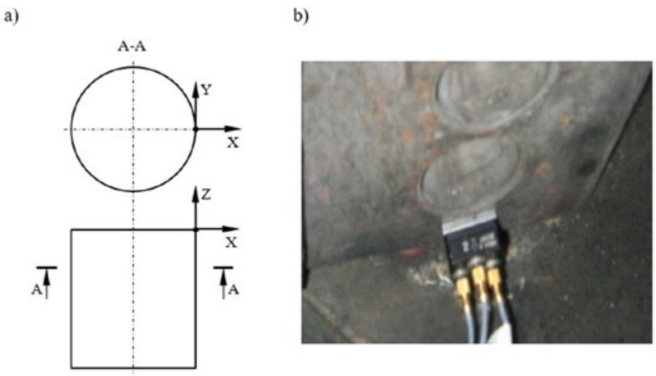


Fig. 3. a). Orientation of the directions of the vibration measurement on the cylinder head, b) view of the vibration transducer fitted on the cylinder head

The selection of the vibration measurement point was preceded by an analysis of the design of the cylinder head, the investigations described in [26] related to the determination of the influence of the diesel engine valve clearance on selected vibration parameters and the previously mentioned impulse tests consisting in hitting of the valves on the valve seats. The hits were carried out by removing of the reference shims from the valve stem and the valve lifter. This was repeated several times for each valve in order to eliminate incidental errors and to perform averaging. It was important to determine such a measurement point that would enable an assessment of the impact of each of the valves. Having performed the analyses of the impulse test results, a single point was selected. Given the dynamics of the signals recorded during the impulse tests, it was observed that direction X might carry the most information related to the valve clearance.

The tests on the operating engine were performed at the engine speed of 700 rpm, 1000 rpm, 1200 rpm, 1500 rpm, 1700 rpm and the engine load of 0 Nm, 22.5 Nm, 45 Nm, 67.5 Nm and 90 Nm. The coolant temperature was 75°C. For the above-mentioned conditions, an acceleration signal recording in three directions was performed. An example portion of the recording in direction X of the vibration signal has been shown in Fig. 4.

In figure 4, we can observe a series of engine work cycles. The dominating phenomenon are the events related to the ignition, while the vibration related to the valve closing is more difficult to spot, particularly due to the low values of the valve clearance settings.

Prior to the analysis, the vibration signals were subjected to angular selection. This means that only those fragments were analyzed that were synchronized with the valve closing in terms of time or angle.

In practice, this can be easily made having appropriate marks on the flywheel or even based on the acceleration signal itself taking into consideration the threshold of the peak value generated during ignition and an appropriate time window. In this way, over 32 thousand time signal portions were obtained for different engine loads and speeds.

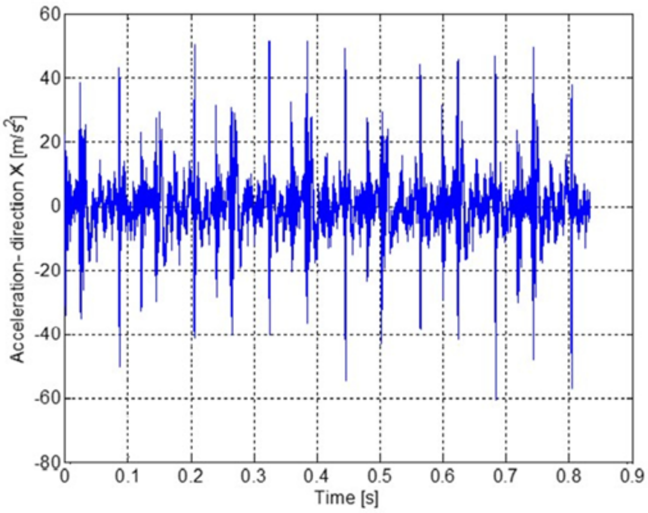


Fig. 4. Example waveform of the vibration acceleration recorded in direction X

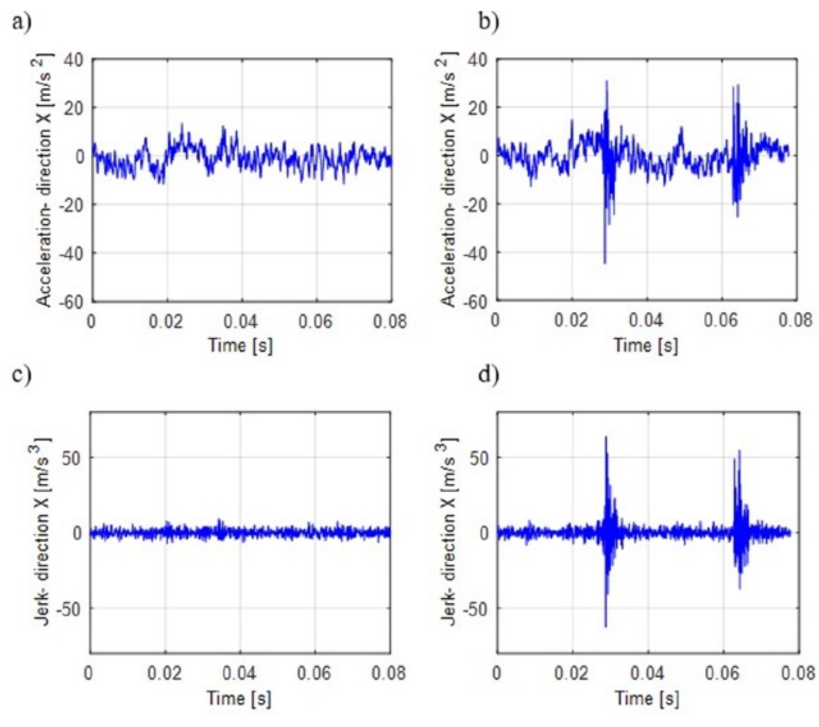


Fig. 5. Portions of tracings of accelerations (figures a, b) and the derivative of the vibration accelerations (figures c, d) related to the closing of the valves; figure a, c – clearance = 0.3 mm; figure b, d – clearance = 1.0 mm

Fig. 5 presents example portions of the recordings for the extreme values of the engine valve clearance of 0.3 mm and 1.0 mm.

In the case of a low value of the valve clearance setting, the moment of the valve closing is practically invisible (Fig. 5a) and in the case of extreme value of the valve setting, the said moment is easily discernible (Fig. 5b). We can clearly see the response of the system resulting from the impact of the valve on its seat. The most difficult to analyze are the indirect examples. It is noteworthy that individual portions differ from one another even for the same valve clearance setting, as shown in Fig. 4 and the obtained sets have a large spread. This is shown in Tab. 1 where the RMS values of the accelerations of selected signals and their standard deviations have been presented.

The design value of the valve clearance is 0.5 mm for the investigated engine. As results from the analysis of Tab. 1, due to an extensive spread of the results, correct determination of the valve clearance is impossible based exclusively on the effective value. It may be assumed that the determination of the valve clearance based on the amplitude measures themselves may have a significant level uncertainty. Given the above, it is substantiated to assess the signals using many measures and then select them in terms of their greatest information contribution to the identification of the class of a given valve clearance setting.

In order to emphasize the phenomena and reduce the impact of low frequencies, a rate of acceleration was also determined (jerk). Figures 5c and 5d present the tracings of this value corresponding to the presented waveforms of vibration accelerations.

In order to retrieve the most significant information from the isolated tracings, an envelope was determined of the signals using the Hilbert transform along with the magnitude of the analytical signal. Additionally, smoothing of the envelope was performed with the Brown's model of exponential smoothing:

$$\begin{cases} \hat{S}_1 = S_1 \\ \hat{S}_t = \alpha S_t + (1-\alpha)\hat{S}_{t-1} \quad \text{dla} \quad t > 1 \end{cases} \tag{1}$$

where: S_t – value of the original observations, $\hat{S}_t$ – values of the observations after smoothing, α –smoothing coefficient (arbitrarily adopted as α=0.1) and t – observation number.

Figures 6a and 6b present the obtained envelope of the waveforms from figures 5c and 5d and the envelope smoothed with the exponential smoothing model (figures 6c and 6d).

In order to train the classification system in distinguishing classes of valve clearance, a parameterization was carried out of the obtained portions of the tracings of the vibration accelerations, its derivatives and the smooth envelope. The obtained envelope was divided into two fragments corresponding to the time windows when the closing of the first and the seconds valves took place.

For the original acceleration signal, its derivative against time and the signal filtered in the frequency range above 2000Hz, the following signal measures were applied: abscissa of the center of gravity of the squared signal, regular and central moments of the first and second order, standardized moments of the same orders, RMS value of the signal, its peak, peak to peak value, average value, surface area under the curve, crest and form factors, peak to average value, as well as signal kurtosis. For the envelope, the counting of the samples was performed above the set levels (9 adopted levels)

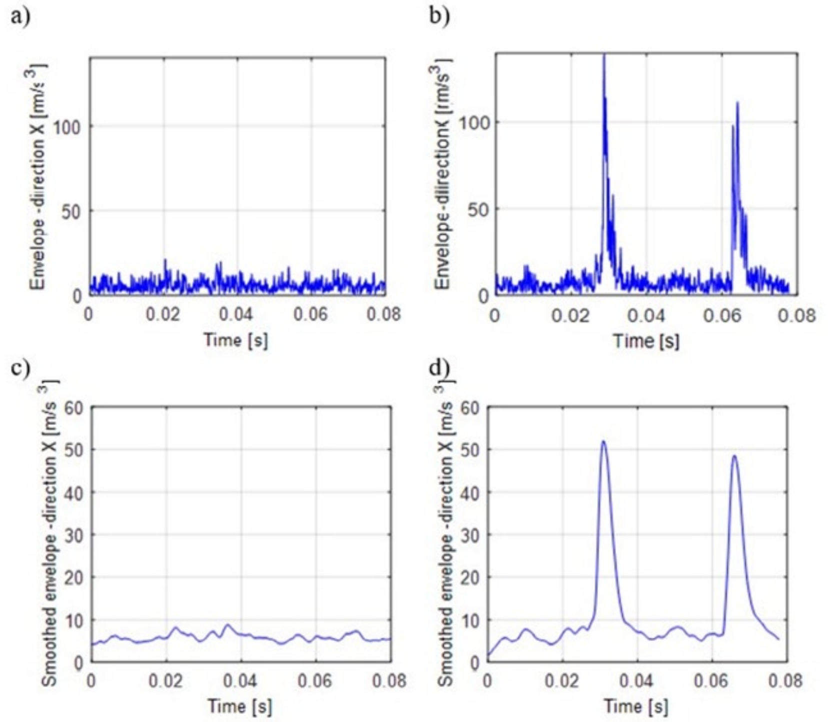


Fig. 6. Envelope of the signal of the acceleration derivative for two extreme cases of the valve clearance settings (figures a and b) and a smooth envelope (figures c and d); figures a, c – valve clearance setting = 0.3 mm; figures b, d – valve clearance setting = 1.0 mm

and the envelope was divided into two fragments related to the closing of the first and the second valve. The analyses were repeated for three recorded directions. Eventually, in excess of 300 parameters were generated. Additionally, the information related to the engine load and speed were considered as attributes. The available signal samples allowed generating 32054 training vectors. Compared to the number of parameters describing the training vectors, this is a set of small number of examples, hence, as described in the further part of the paper, a selection of diagnostic features was performed by reducing the training vectors to 15 parameters.

Each training vector was labeled in relation to the valve clearance setting and the data were divided into three classes: tight clearance (lower than 0.5 mm), optimum clearance (approx. 0.5 mm) and excess clearance (over 0.5 mm). Tab. 2 presents the number of available examples related to individual classes. As results from the table, the numbers of examples for individual classes vary widely, which determines the method of assessment of the results of classification errors.

Figure 7 presents the example space of features crated by two standardized (min-max standardization) attributes marked P5 and P46.

3. Data analysis

In the paper, the authors propose a system of identification of valve clearance of an operating engine based on the machine learning methods. To this end, three methods were compared: a set of three CART (Classification and Regression Tree) binary trees, using the OvA (one versus all) strategy, *k*-nearest neighbors classifier and a one direction MLP (multilayer perceptron) neural network with three outputs related to each of the classes. The classifier based on a set of three trees independently trained to recognize individual classes of valve clearance. The affiliation to a given class was determined by the positive response of one of the three trained classifiers. With such an approach, situations are possible when none of the classifiers give information at the output on the affiliation to a given class or more than one tree responds positively. A solution

Table 1. Mean effective values of the accelerations and their standard deviations measured in the frequency range of up to 6kHz for different measurement directions and different engine valve clearance

Clearance [mm]	Direction X		Direction Y		Direction Z	
	Average RMS m/s^2	Standard deviation m/s^2	Average RMS m/s^2	Standard deviation m/s^2	Average RMS m/s^2	Standard deviation m/s^2
0,3	9.40	4.32	12.47	4.29	10.43	3.50
0,4	9.10	4.49	12.74	4.01	10.48	3.52
0,5	10.11	4.64	12.99	4.29	12.15	4.45
0,6	8.25	2.59	13.12	3.97	11.01	3.80
0,7	10.43	4.42	13.71	4.29	12.48	4.47
0,8	9.35	2.85	13.59	4.07	12.96	4.48
0,9	11.78	5.51	13.41	4.02	14.05	4.73
1,0	12.33	5.08	14.29	4.40	16.53	5.50

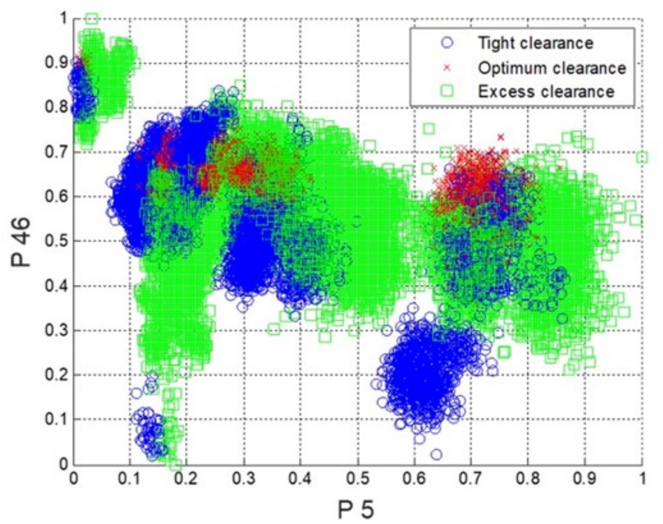


Fig. 7. Available training examples on a plane determined by two arbitrarily selected attributes

Table 2. Number of available examples representing classes

Class	Number of examples
Tight clearance	8160
Optimum clearance	4016
Excess clearance	19878

was adopted that a lack of unanimous classification forces another. In practice, when a situation like this takes place, the recognition system will classify subsequent time portions omitting uncertain cases. One may also repeat the classification a number of times in real time and make a decision based on the significance of the majority of the recognitions of a given class. Given the large number of data generated by an operating engine, in a relatively short time, ignoring the uncertain classifications does not generate any problems in practice. A similar problem occurs in the proposed solution based on the neural network. Here, we can apply the approach, in which the affiliation to a given class is decided by a neuron in the output layer that gives the highest output value. It is unlikely that exactly the same value appears at more than one output. It may happen, however that the response at all network outputs will be on a relatively low level, which may be interpreted as an unrecognized case. Through trial and error, it was assumed that the value at the network output must exceed the threshold of 0.7 to accept the affiliation of a given case to a given class. If this value is not exceeded at any of the network outputs, it is assumed that an unrecognized case occurred and the action is the same as in the case of the set of classification trees.

In order to reduce the size of the feature vectors, it was determined which of the features were the most useful in the classification of the valve clearance. In the first place, a classification was made through the previously discussed classifier in the form of classifying trees. It is known that the tree construction algorithms apply a certain measure of goodness of the division and evaluate individual features based on this measure. In the applied algorithm, the Gini index was used as the measure of the division quality. Out of all available measures, only those were taken into account that were used to build the trees. Then, based on the subsequent step of the analysis, it was observed that some of these measures were significantly linearly correlated. Upon removing some of them from the set of data and performing tests, no significant growth of classification error was observed. At the end, a set of 15 features was obtained, whose further reduction resulted in a greater or smaller increase in the testing error. The final set of features included such measures as: RMS value and the coefficient of shape of the acceleration signal, normal and central moments of the first and second order of the acceleration signal, mean and maximum value, kurtosis from the acceleration signal derivative, number of samples above certain levels (of low values) of the envelope of the derivative of the acceleration signal and the filtered signal as well as engine speed. Majority of the measures selected by the tree construction algorithm pertained to the measurement directions X and Y. The information on the engine load turned out insignificant.

For comparison, the same features were applied for the other classifiers. Obviously, other methods of feature selection could also be applied here. It is possible that they would reveal other attributes that are better for different classifiers. Comparing different possibilities, however, was not the aim of these investigations.

In the k-nearest neighbors classifier, the value of parameter k was modified in the range 1 to 11 along with the measure of distance while observing the testing error. Eventually, parameter k=6 and Euclidean metric were adopted .

Similarly, the optimization of the MLP network was carried out by comparing the errors on the test set. The best results were obtained for the network with two hidden layers with 7 and 6 sigmoidal neurons in individual layers. Increasing as well as decreasing of the number of neurons resulted in a growth of the learning error. For the learning process, the Levenberg – Marquardt gradient algorithm was applied.

Each time, the evaluation of the classifier errors was performed with the Hold-Out repeatable method for which 25% of examples as a test set and 100 test repetitions were arbitrarily adopted. It is noteworthy that this method has a tendency to overvalue the recognition error [17]. The obtained value of standard deviation during the repetition of the test allows determining how resistant a given classification algorithm is to the change of the training data. Such a solution also has its downsides as it is unlikely that during the tests, all sets of data will be depleted.

Due to a significant difference in the number of representatives of different classes, in order to compare the methods, a weighted classification error was calculated expressed with the formula:

$$\varepsilon = \frac{1}{K}\sum_{i=1}^{K}\frac{\sum_{j=1, j\neq i}^{K} a_{ij}}{K_i} \tag{2}$$

where: K – number of classes, K_i – number of elements in an *i-th* class, a_{ij} elements of the matrix of class distribution (confusion matrix) from outside of the diagonal.

Such a definition of error allows for the varied numbers of examples of different classes.

Table 3 presents the averaged results obtained during the tests of the compared methods while Tab. 4 presents matrices of the classes distribution. In the case of the neural networks as well as the k-NN classifier, results from the best classifiers have been presented (of the best selected network structure for the *k*-NN classifier of the optimum *k* parameter and distance measure). Due to the varied number of representatives of a given class, the number of errors was referred to the number of representatives of a given class.

From the comparison of the classifiers, it results that, out of the analyzed solutions of valve clearance identification for the investigated engine, the most advantageous is the set of binary classifiers. The *k*-NN classifier in the applied algorithm does not identify uncertain cases. This explains the greater classification error compared to other classifiers. ‘Difficult’ cases are classified and unrecognized cases are deemed uncertain. There is a possibility of introducing the nearest neighbors algorithm (k,l), for which, if there is an insufficient number of neighbors voting for a given class (fewer than *l*), the classification is deemed uncertain. Yet, this introduces another parameter *l* to optimize the classifier. As has been mentioned earlier in the paper, in the case of sets of trees and networks, uncertain classifications are rejected and the decision as to the class affiliation may be made based on subsequent signal samples. Upon rejecting of the uncertain classifications, the set of trees allows classifying the valve clearance with the accuracy of 99% despite the influence of different engine loads and speeds. In the case of other classifiers, the results are worse (approx. 98%) but acceptable, particularly, since, for the problem at hand, one may repeat the diagnosis in a short time.

The classifier based on classification trees was characterized by the highest resistance to change of the training set measured with the error standard deviation. All classifiers make the most errors in correct identification of the optimum valve clearance and it is most frequently

Table 3. Results of the classifiers tests

Method	Average weighted error of classification [%]	Error standard deviation – measure of resistance of classifier to the training set [%]	Share of uncertain classifications [%]
Set of trees	0.93	0.13	3.14
k-NN (k=6)	1.98	0.30	0
MLP [7,6]	1.73	0.70	3.27

Table 4. Matrix of the distribution of classes for the set of classification trees

		Real classes		
		Tight clearance [%]	Optimum clearance [%]	Excess clearance [%]
Recognized classes	Tight clearance	99.4	0.4	0.2
	Optimum clearance	0.1	98.1	0.1
	Excess clearance	0.5	1.5	99.7

Table 5. Matrix of the distribution of classes for the k-NN classifier (k=6)

		Real classes		
		Tight clearance [%]	Optimum clearance [%]	Excess clearance [%]
Recognized classes	Tight clearance	99.2	1.1	0.4
	Optimum clearance	0.4	96.0	0.8
	Excess clearance	0.4	2.9	98.8

Table 6. Matrix of the distribution of classes for the classifier in the form of a neuron network [7,6] (7 and 6 neurons in the hidden layers)

		Real classes		
		Tight clearance [%]	Optimum clearance [%]	Excess clearance [%]
Recognized classes	Tight clearance	98.8	0.6	0.7
	Optimum clearance	0.2	97.1	0.4
	Excess clearance	1.0	2.3	98.9

confused with the excess valve clearance (1.5%, 2.9 % and 2.3 % of the cases of optimum valve clearance were classified as excess valve clearance). The best results in this respect has the set of classification trees. In the case of this classifier, out of the compared methods, the most successfully identified valve clearance was the excess one.

4. Conclusions

Following the performed analyses, a method can be proposed of classification of the engine valve clearance based on vibration signals measured on the cylinder head. For the engine under investigations this is successful with approx. 99%. Given the possibility of multiple repetition of the classification process almost in real time and selecting the most frequently occurring class, one may decide on the class of the valve clearance with a high level of certainty. In order to determine such a state for the investigated engine, as few as 15 easily obtainable number parameters is sufficient. Obviously, for other type of engines, it will be necessary to perform the entire process of classifier construction.

In the paper, in order to solve the presented problem, the authors proposed an application of a set of three binary trees specialized in identification of each class of valve clearance. Such an approach allows an obtainment of relatively small classification errors (for the analyzed case) and automates the selection of **components** of the training vector. Additionally, the tree allows an easy generation of human friendly rules that are also easy to implement in a system that would operate autonomously.

Acknowledgements

The presented research has been subsidized by the Ministry of Science and Higher Education P05/52/DSPB/0277 and 02/21/SBAD/3558.

References

1. Albarbar A, Ball A, Starr A. On acoustic measurement-based condition monitoring of internal combustion engines. Insight: Non-Destructive Testing and Condition Monitoring 2008; 50(1): 30-34, https://doi.org/10.1784/insi.2008.50.1.30.
2. Albertson F, Bodén H, Gilbert J. Comparison of different methods to couple nonlinear source descriptions in the time domain to linear system descriptions in the frequency domain - Application to a simple valveless one-cylinder cold engine. Journal of Sound and Vibration 2006; 291(3-5): 963-985, https://doi.org/10.1016/j.jsv.2005.07.046.
3. Arroyo J, Muñoz M, Moreno F, Bernal N, et al. Diagnostic method based on the analysis of the vibration and acoustic emission energy for emergency diesel generators in nuclear plants. Applied Acoustics 2013; 74(4): 502-508, https://doi.org/10.1016/j.apacoust.2012.09.010.
4. Babu A K, Raj A A, Kumersan G. Misfire detection a multi- cylinder diesel engine: a machune learning approach. Journal of Engineering Science and Technology 2016; 11(2): 278-295.
5. Badawy T, Shrestha A, Henein N. Detection of Combustion Resonance Using an Ion Current Sensor in Diesel Engines. Journal of Engineering for Gas Turbines and Power 2012; 134(5): 052-802, https://doi.org/10.1115/1.4004840.
6. Ahmed A B, Elaraby I S. A prediction for Student's Performance Using Decision Tree ID3 Method. India - World Journal of Computer Application and Technology 2014; 2(2): 43-47.
7. Czech P, Bąkowski H. Diagnosing of car engine fuel injectors damage using dwt analysis and PNN neural networks. Transport Problems 2013; 8(3): 85-91.
8. Delvecchio S, Bonfiglio P, Pompoli F. Vibro-acoustic condition monitoring of Internal Combustion Engines: A critical review of existing

techniques. Mechanical Systems and Signal Processing 2018; 99: 661-683, https://doi.org/10.1016/j.ymssp.2017.06.033.
9. Desbazeille M, Randall R B, Guillet F, El Badaoui M, et al. Model-based diagnosis of large diesel engines based on angular speed variations of the crankshaft. Mechanical Systems and Signal Processing 2010; 24(5): 1529-1541, https://doi.org/10.1016/j.ymssp.2009.12.004.
10. Dolatabadi N., Theodossiades S., Rothberg S.J., On the identification of piston slap events in internal combustion engines using tribodynamic analysis. Mechanical Systems and Signal Processing 2015; 58: 308-324, https://doi.org/10.1016/j.ymssp.2014.11.012.
11. Erbek F S, Özkan C, Taberner M. Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. International Journal of Remote Sensing 2004; 25(9): 1733-1748, https://doi.org/10.1080/0143116031000150077.
12. Figlus T, Liščák Š, Wilk A, Łazarz B. Condition monitoring of engine timing system by using wavelet packet decomposition of a acoustic signal. Journal of Mechanical Science and Technology 2014; 28(5): 1663-1671, https://doi.org/10.1007/s12206-014-0311-3.
13. Fog TL, Hansen LK, Larsen J, Hansen HS, et al. On condition monitoring of exhaust valves in marine diesel engines. Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop 1999; 554-563.
14. Gao F, Lv J. Fault Diagnosis for Engine Based on Single-Stage Extreme Learning Machine. Mathematical Problems in Engineering 2016; 2016: 1-10, https://doi.org/10.1155/2016/7939607.
15. Gorges C, Öztürk K, Liebich R. Impact detection using a machine learning approach and experimental road roughness classification. Mechanical Systems and Signal Processing 2019; 117: 738-756, https://doi.org/10.1016/j.ymssp.2018.07.043.
16. Jia F, Lei Y, Lu N, Xing S. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. Mechanical Systems and Signal Processing 2018; 110: 349-367, https://doi.org/10.1016/j.ymssp.2018.03.025.
17. Korbicz J, Kościelny J M, Kowalczuk Z, Cholewa W (Eds.). Models, Artificial Intelligence, Applications. Springer-Verlag Berlin Heidelberg, 2004.
18. Lahrache A, Cocconcelli M, Rubini R. Anomaly detection in a cutting tool by k-means clustering and support vector machines. Diagnostyka 2017; 18(3): 21-29.
19. Leemans V, Magein H, Destain M F. Defect segmentation on 'Jonagold' apples using colour vision and a Bayesian classification method. Computers and Electronics in Agriculture 1999; 23: 43-53, https://doi.org/10.1016/S0168-1699(99)00006-X.
20. Lei Y, He Z, Zi Y. Application of an intelligent classification method to mechanical fault diagnosis. Expert Systems with Applications 2009; 36(6): 9941-9948, https://doi.org/10.1016/j.eswa.2009.01.065.
21. Madej H, Czech P. Discrete wavelet transform and probabilistic neural network in IC engine fault diagnosis. Eksplotacja i Niezawodnosc - Maintenance and reliability 2010; 4(48): 47-54.
22. Matzke W. Four-stroke engine timing, Warszawa: WKiŁ, 1967.
23. Mechefske C K, Mathew J. Fault detection and diagnosis in low speed rolling element bearings Part II: The use of nearest neighbour classification. Mechanical Systems and Signal Processing 1992; 6(4): 309-316, https://doi.org/10.1016/0888-3270(92)90033-F.
24. Nguyen T T, Armitage G. A survey of techniques for internet traffic classification using machine learning. IEEE Communications Surveys Tutorials 2008, 10(4): 56-76, https://doi.org/10.1109/SURV.2008.080406.
25. Niziński S, Michalski R. Diagnostics of technical objects. Radom: Department of Vehicle and Machine Operation, Faculty of Technical Sciences, University of Warmia and Mazury in Olsztyn, Institute for Sustainable Technologies in Radom, 2002.
26. Osiecki J, Ziemba S. Basics of mechanical vibration measurements, Warszawa: PWN, 1968.
27. Potočnik P, Govekar E. Semi-supervised vibration-based classification and condition monitoring of compressors. Mechanical Systems and Signal Processing 2017; 93: 51-65, https://doi.org/10.1016/j.ymssp.2017.01.048.
28. Qinghua W, Youyun Z, Lei C, Yongsheng Z. Fault diagnosis for diesel valve trains based on non-negative matrix factorization and neural network ensemble. Mechanical Systems and Signal Processing 2009; 23(5): 1683-1695, https://doi.org/10.1016/j.ymssp.2008.12.004.
29. Ruiz M, Mujica L E, Alférez S, Acho L, et al. Wind turbine fault detection and classification by means of image texture analysis. Mechanical Systems and Signal Processing 2018; 107: 149-167, https://doi.org/10.1016/j.ymssp.2017.12.035.
30. Serridge M, Licht T R. Piezoelectric accelerometers and vibration preamplifiers, Nearum: Brüel & Kjær, 1987.
31. Szymański G.M., Tomaszewski F., Diagnostics of automatic compensators of valve clearance in combustion engine with the use of vibration signal. Mechanical Systems and Signal Processing 2016; 68-69: 479-490, https://doi.org/10.1016/j.ymssp.2015.07.015.
32. Wang C, Zhang Y, Zhong Z. Fault diagnosis for diesel valve trains based on time-frequency images. Mechanical Systems and Signal Processing 2008; 22(8), 1981-1993, https://doi.org/10.1016/j.ymssp.2008.01.016.
33. Wong P K, Zhong J, Yang Z, Vong C M. Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis. Neurocomputing 2016; 174: 331-343, https://doi.org/10.1016/j.neucom.2015.02.097.
34. Wu J, Liu C. An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. Expert Systems with Applications 2009, 36: 4278-4286, https://doi.org/10.1016/j.eswa.2008.03.008.
35. Zhang H, Li J, Huang Y, Zhang L. A Nonlocal Weighted Joint Sparse Representation Classification Method for Hyperspectral Imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2014; 7(6): 2056-2065, https://doi.org/10.1109/JSTARS.2013.2264720.

Maciej Tabaszewski
Poznan University of Technology
Institute of Applied Mechanics
ul. Jana Pawła II 24, 60-965 Poznań, Poland
E-mail: maciej.tabaszewski@put.poznan.pl

Grzegorz. M. Szymański
Poznan University of Technology
Institute of Transport
ul. Piotrowo 3, 60-965 Poznań, Poland
E-mail: grzegorz.m.szymanski@put.poznan.pl

Article citation info:
SUN B, LI Y, WANG Z, LI Z, XIA Q, REN Y, FENG Q, YANG D, QIAN C. Physics-of-failure and computer-aided simulation fusion approach with a software system for electronics reliability analysis. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 340–351, http://dx.doi.org/10.17531/ein.2020.2.17.

Bo SUN
Yu LI
Zili WANG
Zhifeng LI
Quan XIA
Yi REN
Qiang FENG
Dezhen YANG
Cheng QIAN

PHYSICS-OF-FAILURE AND COMPUTER-AIDED SIMULATION FUSION APPROACH WITH A SOFTWARE SYSTEM FOR ELECTRONICS RELIABILITY ANALYSIS

METODA I OPROGRAMOWANIE DO ANALIZY NIEZAWODNOŚCI URZĄDZEŃ ELEKTRONICZNYCH OPARTE NA POŁĄCZENIU METODOLOGII FIZYKI USZKODZEŃ I SYMULACJI KOMPUTEROWEJ

Electronics, such as those used in the communication, aerospace and energy domains, often have high reliability requirements. To reduce the development and testing cost of electronics, reliability analysis needs to be incorporated into the design stage. Compared with traditional approaches, the physics of failure (PoF) methodology can better address cost reduction in the design stage. However, there are many difficulties in practical engineering applications, such as processing large amounts of engineering information simultaneously. Therefore, a flexible approach and a software system for assisting designers in developing a reliability analysis based on the PoF method in electronic product design processing are proposed. This approach integrates the PoF method and computer-aided simulation methods, such as CAD, FEM and CFD.The software system integrates functional modules such as product modeling, load-stress analysis and reliability analysis, which can help designers analyze the reliability of electronic products in actual engineering design. This system includes software and hardware that validate the simulation models. Finally, a case study is proposed in which the software system is used to analyze the filter module reliability of an industrial communication system. The results of the analysis indicate that the system can effectively promote reliability and can ensure the accuracy of analysis with high computing efficiency.

Keywords: *physics of failure, reliability analysis, electronics, prognostics and health management, computer-aided simulation, software system.*

Urządzenia elektroniczne, na przykład te używane w łączności, lotnictwie i energetyce, często muszą spełniać wysokie wymagania dotyczące niezawodności. Aby zmniejszyć koszty rozwoju i testowania tego typu urządzeń, należy opracować metodę analizy niezawodności, którą można wykorzystywać już na etapie projektowania. Metodologia fizyki uszkodzeń (PoF) pozwala, lepiej niż tradycyjne podejścia, rozwiązywać problemy związane z niezawodnością już na etapie powstawania projektu. Jednak jej zastosowanie w praktyce inżynierskiej nastręcza wielu trudności, związanych, między innymi, z koniecznością jednoczesnego przetwarzania dużych ilości informacji inżynieryjnych. W związku z tym, w przedstawionej pracy zaproponowano elastyczne podejście oraz system oprogramowania, które mogą być wykorzystywane przez projektantów do opracowania analizy niezawodności produktu elektronicznego w opaciu o PoF na etapie projektowania. Podejście to stanowi połączenie metody PoF i metod symulacji komputerowej, takich jak CAD, FEM i CFD. System oprogramowania zawiera moduły funkcjonalne, takie jak modelowanie produktu, analiza obciążeń, analiza niezawodności i inne, które mogą wspomagać projektantów w analizie niezawodności projektowanych przez nich produktów elektronicznych. Na system ten, oprócz oprogramowania składa się także sprzęt komputerowy, który służy do walidacji modeli symulacyjnych. W artykule przedstawiono studium przypadku, w którym zaproponowany system oprogramowania wykorzystano do analizy niezawodności modułu filtra wykorzystywanego w systemie łączności przemysłowej. Wyniki analizy pokazują, że opracowane oprogramowanie skutecznie poprawia niezawodność urządzeń jak też zapewnia dokładność analizy przy jednoczesnej wysokiej wydajności obliczeniowej.

Słowa kluczowe: *fizyka uszkodzeń, analiza niezawodności, urządzenia elektroniczne, prognostyka i zarządzanie zdrowiem, symulacja komputerowa, system oprogramowania.*

1. Introduction

The reliability discipline is regarded as a 'black box art' or 'number game' in science. Usually, electronics reliability analysis conducts a series of tasks in practice that focus on potential failure mode identification and reliability parameter calculation. Traditional electronics reliability analysis mainly depends on statistical experiments to obtain product failure information. Traditional electronics reliability analysis is mainly based on qualitative or semi-quantitative analysis, which ignores failure mechanism analysis in the product design stage. For this problem, the physics of failure (PoF) methodology is more suitable. The PoF approach focuses on failure mechanisms and root causes of failure in products and emphasizes the quantitative analysis and description of physical and chemical processes for product failure[10, 30].

The PoF was formally conceptualized in the first of a series of symposia in 1962 organized by the Rome Air Development Center (RADC) of the US Air Force [2]. After that, the PoF approach became an important research topic in the reliability field. Since 1967, the IEEE Reliability Physics Symposium (IRPS) has continued to present research related to the PoF [2]. In addition, several related research institutions have had an important influence on the development of the PoF. Computer Aided Life Cycle Engineering (CALCE) at the University of Maryland carried out many studies on the PoF approach and its application and presented the PoF-based reliability prediction method [15, 21-23]. The RADC and the Research Foundation of the Illinois Institute of Technology also promoted the development of the PoF [2, 4]. In China, the School of Reliability and Systems Engineering, Beihang University, cooperates with CALCE and has also performed many studies on PoF-based reliability prediction and fault prognostics for electronics [26-28]. In addition to traditional failure mechanism analysis, the PoF approach is extensively applied to accelerated tests [3] and lifetime assessment [12, 13, 16, 31]. For example, Bretts et al [24] introduced physical models into accelerated tests to estimate the time to failure (TTF) of resistors. As the PoF approach has developed, some researchers have attempted to introduce this approach into prognostics and health management (PHM). Pecht et al [19, 25] presented a PoF-based PHM approach for effective reliability prediction. K. Ma et al [16] researched the application of the PoF for prediction and design in power electronics systems. H. Oh et al [20] reviewed the application of the PoF in performing prognostics of insulated gate bipolar transistor modules.

Although the PoF approach has obtained abundant research results, there are many difficulties in using it in practical engineering applications. First, the PoF is a multidisciplinary application involving engineering, physics and chemistry. Abundant professional experiments or computer simulations are necessary to establish physical models, which brings more difficulties for engineers. Second, many kinds of simulation softwares are employed in the computer simulation process. The heterogeneous output files of different software make information extraction and communication more difficult. A large number of information files is inconvenient for unified engineering management. Therefore, professional PoF software is required to solve these difficulties. Due to the complexity of information processing, many research fields develop information systems to manage and process related information [5,6,17]. Professional software can build physical models, which can provide convenience for engineers. Moreover, according to experiment and engineering experience, the inserted physical models can be revised, which can reduce difficulties in obtaining information. At present, there is some professional PoF software in the commercial market, such as CALCE SARA [7], DFR Solutions Sherlock ADA [8, 18] and Reliass ASENT [9]. CALCE SARA and DFR Solutions Sherlock ADA can support a complete PoF analysis for microelectronics applications. The detailed comparability of those software programs can be found in [11]. However, there are still some problems with the above software. First, ASENT only supports thermal analysis. Second, except for DFR Solutions Sherlock, all of this software adopts a simplified numerical method in load-stress analysis, which reduces the accuracy of the analysis results. Additionally, ASENT and SARA do not support 3D modeling. In some engineering applications, to improve the credibility of reliability analysis results, finite element analysis based on 3D models is executed prior to providing the input data for reliability analysis. Information conversion among the different commercial software programs is complicated in application. Finally, in the reliability analysis process, none of the above software integrates the simulation model validation function.

In accordance with recent product development solutions, this paper presents a physics-of-failure and computer-aided simulation fusion approach. We employed an advanced PoF approach and product digital prototype method to develop professional PoF software. This software system integrates the CAD, CFD and FEA techniques into a computation environment and avoids information conversion among different commercial software. With the integrated technique, this software reduces the application difficulty of the PoF approach in the engineering field. The integrated computation environment is also equipped with corresponding hardware to achieve simulation model validation. The accuracy of the analysis results can be effectively improved through model validation. Parallel computing is applied to the integrated system to improve computational efficiency. This paper shows the details of the environment, such as the method application, information integration process and operating process.

The remaining part of this paper is organized as follows: section II illustrates the basic theory for PoF-based electronics reliability analysis and model validation; section III discusses the system architecture, the information transfer interface design for the software and the model validation process and related hardware setup; section IV introduces a typical case used in airborne electronics and the improvement of reliability; section V summarizes the characteristics, advantages and innovations of the software system.

2. Basic theory for system design

2.1. PoF-based reliability analysis method

In this software, an advanced PoF method is employed to conduct reliability analysis. Failure mechanism models are an important basis for applying the PoF method, and their foundation is based on the understanding of product failure laws. The failure mechanism model usually describes the functional relationship between a product's life, reliability or performance parameters and geometric parameters, material characteristics and various typical environmental load parameters (such as temperature, humidity, and vibration). A typical functional relationship can be expressed as [29]:

$$TTF_i = f\left(\mathbf{g},\mathbf{m},\mathbf{e},\mathbf{o},\cdots\right), \quad (1)$$

where TTF_i is the time to failure under the i th failure mechanism, $\mathbf{g}$ is the geometric parameter vector, $\mathbf{m}$ is the material parameter vector, $\mathbf{e}$ is the environmental parameter vector, and $\mathbf{o}$ is the operation load parameter vector.

Generally, product failure is caused by multiple failure mechanisms. To cover all failure mechanisms, a multipoint distribution fusion approach is used to obtain the probability density function (PDF) of the product lifetime. Due to model parameter uncertainty, the calculated TTF is different each time based on the failure mechanism model. Therefore, the failure information matrix can be obtained by Monte Carlo simulation as follows:

$$I=\begin{bmatrix}A_1\\A_2\\\vdots\\A_n\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1m}\\a_{21}&a_{22}&\cdots&a_{2m}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\cdots&a_{nm}\end{bmatrix}=\begin{bmatrix}f\left(x_{A_1}\right)\\f\left(x_{A_2}\right)\\\vdots\\f\left(x_{A_n}\right)\end{bmatrix}\tag{2}$$

where A_n represents the failure information vector of a failure mechanism. Based on the failure information vector, the corresponding PDF $f\left(x_{A_n}\right)$ of the failure mechanism can be calculated by the distribution fitting method.

In addition, some failure modes of components are not independent. Under the coupled influence of temperature and vibration, the correlation between failure modes often shows a positive correlation; that is, the generation of one failure mode accelerates the failure process of another failure mode. Since each failure mode of a component often causes the overall failure of the component, the component can be regarded as a series relationship among various failure mode conditions. The component failure mode correlation can be used as the series failure correlation of each failure mode. To resolve this issue, the copula approach is used to construct the correlation model. The copula function organically links the joint probability density function of a multivariate random variable with the marginal probability density function of each variable [32, 33]. The correlation among the random variables is considered, and the solving process of the joint probability density function of the multivariate random variables can be simplified.

According to Sklar's theorem [32], the following joint cumulative distribution function can be obtained:

$$F\left(x_{A_1},x_{A_2},\cdots,x_{A_n}\right)=C\left(F_{A_1}\left(x_{A_1}\right),F_{A_2}\left(x_{A_2}\right),\cdots,F_{A_n}\left(x_{A_n}\right)\right).\tag{3}$$

From the above formula, the PDF of the joint probability density function can be obtained as:

$$\begin{aligned}f_C\left(x_{A_1},x_{A_2},\cdots,x_{A_n}\right)&=\frac{\partial^n F}{\partial x_{A_1}\partial x_{A_2}\cdots\partial x_{A_n}}\\&=\frac{\partial^n C\left(F_{A_1}\left(x_{A_1}\right),F_{A_2}\left(x_{A_2}\right),\cdots,F_{A_n}\left(x_{A_n}\right)\right)}{\partial F_{A_1}\partial F_{A_2}\cdots\partial F_{A_n}}\frac{\partial F_{A_1}}{\partial x_{A_1}}\frac{\partial F_{A_2}}{\partial x_{A_2}}\cdots\frac{\partial F_{A_n}}{\partial x_{A_n}}\\&=c\left(F_{A_1}\left(x_{A_1}\right),F_{A_2}\left(x_{A_2}\right),\cdots,F_{A_n}\left(x_{A_n}\right)\right)f\left(x_{A_1}\right)f\left(x_{A_2}\right)\cdots f\left(x_{A_n}\right)\end{aligned}\tag{4}$$

Then, the failure correlation matrix is as follows:

$$J=\begin{bmatrix}C_1\\C_2\\\vdots\\C_m\end{bmatrix}=\begin{bmatrix}FM_1&FM_2&&\\FM_1&FM_2&FM_3&\\\vdots&\vdots&\ddots&\\FM_1&FM_2&\cdots&FM_n\end{bmatrix}=\begin{bmatrix}f_{C_1}\left(x_{A_1},x_{A_2}\right)\\f_{C_2}\left(x_{A_1},x_{A_2},x_{A_3}\right)\\\vdots\\f_{C_m}\left(x_{A_1},x_{A_2},\cdots,x_{A_n}\right)\end{bmatrix}\tag{5}$$

where C_m represents the failure mode vector. The Monte Carlo method is used to sample random numbers from the PDF $\left[f\left(x_{A_1}\right),f\left(x_{A_2}\right),\cdots,f\left(x_{A_n}\right),f_{C_1}\left(x_{A_1},x_{A_2}\right),f_{C_2}\left(x_{A_1},x_{A_2},x_{A_3}\right),\cdots,f_{C_m}\left(x_{A_1},x_{A_2},\cdots,x_{A_n}\right)\right]$, which yields the TTF results of the corresponding failure mechanism $\left(t_{A_1}^{(1)},t_{A_2}^{(1)},\cdots,t_{A_n}^{(1)},t_{C_1}^{(1)},t_{C_2}^{(1)},\cdots,t_{C_m}^{(1)}\right)$. Then, the TTF samples of a product can be represented as:

$$TTF^{(1)}=\min\left(t_{A_1}^{(1)},t_{A_2}^{(1)},\cdots,t_{A_n}^{(1)},t_{C_1}^{(1)},t_{C_2}^{(1)},\cdots,t_{C_m}^{(1)}\right).\tag{6}$$

By repeating the above process n times, the TTF sample sets can be obtained, which are denoted as $\left(TTF^{(1)},TTF^{(2)},\cdots,TTF^{(n)}\right)$. Using these sample sets to perform the distribution goodness test, the lifetime PDF $g(x)$ of the product is obtained.

The above process can be summarized as follows: First, determine the prior distribution of the input parameters in the PoF model. Then, a Monte Carlo simulation is used to obtain the sampling values of the input parameters. The sampling values are substituted into the PoF model to obtain the TTF sample values corresponding to the n th failure mechanisms. The copula method is used to obtain the PDF of correlation failure. A competition model is used to determine the minimum TTF, and the selected PoF model is sampled and fitted again. The TTF distribution function of the product is obtained. The algorithm is summarized in Table 1.

Table 1. PoF-based reliability analysis algorithm

Generate failure information matrix
For i =1:n
Sample $(\mathbf{g},\mathbf{m},\mathbf{e},\mathbf{o})\sim U(a,b\|Z)$ based on the prior distribution of model parameters
Sample $\left(a_{i1},a_{i2},\cdots,a_{im}\right)\sim f(\mathbf{g},\mathbf{m},\mathbf{e},\mathbf{o})$ based on the corresponding failure mechanism model
Fit the distribution $f\left(x_{A_i}\right)\sim\left(a_{i1},a_{i2},\cdots,a_{im}\right)$
End
Generate failure correlation matrix
For i = 1:m
Construct the joint probability density function $f_{C_m}\left(x_{A_1},x_{A_2},\cdots,x_{A_n}\right)$ based on Eq.(4)
End
Multi-point distribution fusion
For j =1:n
Sample $\begin{pmatrix}t_{A_1}^{(j)},t_{A_2}^{(j)},\cdots,t_{A_n}^{(j)},\\t_{C_1}^{(j)},t_{C_2}^{(j)},\cdots,t_{C_m}^{(j)}\end{pmatrix}\sim\begin{bmatrix}f\left(x_{A_1}\right),f\left(x_{A_2}\right),\cdots,f\left(x_{A_n}\right),f_{C_1}\left(x_{A_1},x_{A_2}\right),\\f_{C_2}\left(x_{A_1},x_{A_2},x_{A_3}\right),\cdots,f_{C_m}\left(x_{A_1},x_{A_2},\cdots,x_{A_n}\right)\end{bmatrix}$ based on the PDF of the failure mode
Calculate the minimum from $TTF^{(j)}\sim\left(t_{A_1}^{(j)},t_{A_2}^{(j)},\cdots,t_{A_n}^{(j)},t_{C_1}^{(j)},t_{C_2}^{(j)},\cdots,t_{C_m}^{(j)}\right)$ based on Eq.(6)
Generate the TTF sample vector $\left(TTF^{(1)},TTF^{(2)},\cdots,TTF^{(n)}\right)$
End
Calculate PDF of product lifetime
Fit the distribution $g(x)\sim\left(TTF^{(1)},TTF^{(2)},\cdots,TTF^{(n)}\right)$

2.2. Model validation approach

2.2.1. Thermal model validation approach

In the thermal simulation model validation, the temperatures of the corresponding temperature test points are generally compared under the same steady state. If the error is within the allowable range, the simulation model is taken to reflect the real situation. Since thermal test values are generally discrete values and do not satisfy the completely random sampling rule, this paper selects the Theil inequality coefficient (TIC) [14] approach to verify the thermal simulation model. Suppose that x_i is the simulation model output sequence and z_i is the actual system output sequence. The data length is N. Then, the function formula is as follows:

$$\rho(x,z)=\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-z_i)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)^2}+\sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i)^2}}=\frac{\sqrt{\sum_{i=1}^{N}(x_i-z_i)^2}}{\sqrt{\sum_{i=1}^{N}(x_i)^2}+\sqrt{\sum_{i=1}^{N}(z_i)^2}} \tag{7}$$

If $\rho=0$, the simulation model output sequence is exactly the same as the actual system output sequence. If $\rho=1$, the simulation model output sequence is completely uncorrelated with the actual system output sequence. Therefore, the closer the TIC value is to 0, the higher the accuracy of the simulation model. If the TIC is greater than the prediction threshold δ, the simulation result will not satisfy the accuracy requirements. Considering the model simplification and measurement noise, the threshold δ is set to 0.4. The product model needs to be revised until the TIC is smaller than the threshold δ. The revised model parameters include the actual power consumption, equivalent thermal resistance of the device and equivalent thermal conductivity of the material.

2.2.2. Vibration model validation approach

In vibration analysis, the modal analysis results are generally used to regulate the simulation model. Therefore, the modal assurance criterion (MAC) [1] is chosen as the model validation approach. The formula for the MAC is as follows:

$$MAC_{ij}=\frac{\left[\varphi_i^T\varphi_j\right]^2}{\left(\varphi_i^T\varphi_i\right)\left(\varphi_j^T\varphi_j\right)} \tag{8}$$

where φ_i and φ_j are modal vectors calculated by a simulation analysis and a physical test, respectively. The MAC can be used to check the consistency or mutual independence between the simulated modal results and the experimental modal results. In theory, if MAC=1, the experimental modal vector is exactly the same as the simulated modal vector. If MAC=0, it means that the two modes are orthogonal; that is, the experimental modal vector has no linear relationship with the simulated modal vector. However, due to the model simplification, the external noise interference of measurement data and improper data processing, the calculation result of the MAC will be affected. Generally, when MAC > 0.7, the two modes can be considered to have a good linear relationship, and the simulation model reflects the real vibration. When MAC < 0.2, the simulation model does not reflect the real vibration.

3. System architecture

3.1. Program structure

The Physics of Failure-based electronics reliability analysis (PofEra) is an integrated multidisciplinary simulation analysis system that employs the advanced PoF and product digital prototype methods to assess electronics reliability. Graphical modeling, finite element analysis (FEA) and computational fluid dynamics (CFD) analysis are integrated into the system. This system can save and import product parameter information and environmental load information and enable the visualization of the model and analysis results. Moreover, the different function modules can execute either separately or in a sequential order. The PofEra comprises seven main program modules: product modeling (PM), the mission profile (MP), load-stress analysis (LSA), reliability analysis (RA), simulation validation (SV), reliability optimization (RO) and a basic database (DB). Fig. 1 shows the detailed function structure of the different function modules.

a) PM module: Supports constructing the CAD model of circuit boards, which covers the device, circuit board and PTH modeling. In addition, this module enables the input model to be provided by other CAD software (such as CATIA/UG).
b) MP module: Provides the environment and time information for load-stress analysis and failure prediction.
c) LSA module: Enables the thermal and vibration analyses to be carried out under the corresponding mission profiles. The thermal and vibration analyses apply the CFD program and FEA program, respectively.

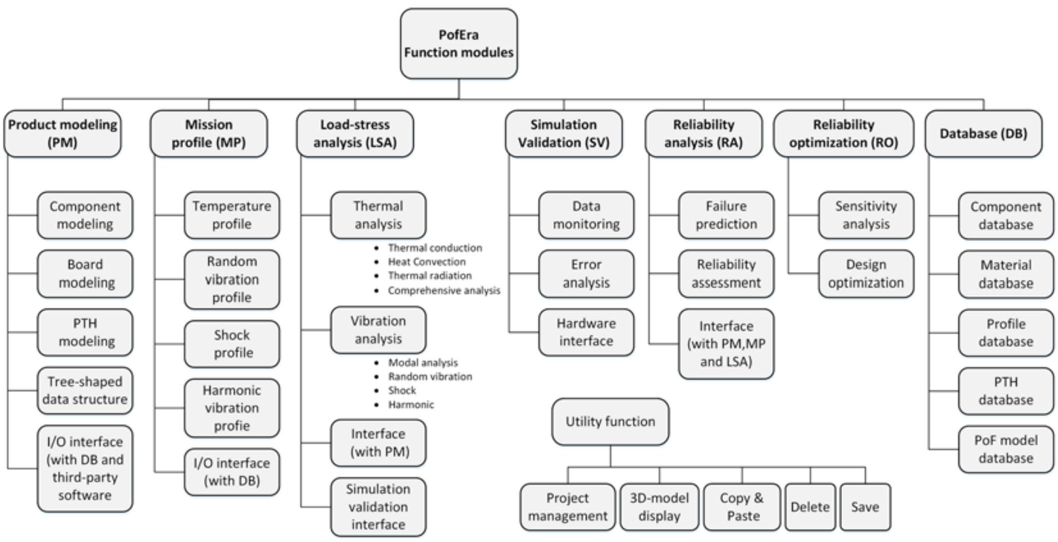


Fig. 1. Function modules of PofEra

d) SV module: Validates the simulation credibility through a consistency test between the simulation and practical test results.
e) RA module: Predicts the reliability or remaining useful life for the components and circuit board based on the mission profile.
f) RO module: Proposes improved measures according to the sensitivity analysis results to avoid failure. Health management can be realized based on the remaining useful life results.
g) DB module: Stores and provides part of the basic data, including material data, device data and PTH data.

3.2. Software hierarchy

The software hierarchy of the PofEra system is composed of four hierarchies—the software foundation hierarchy, information management hierarchy, function model hierarchy and user interface hierarchy—as shown in Fig. 2. First, the software foundation hierarchy mainly establishes a software system support environment, such as the operating system Microsoft .NET framework, CFD program, FEA program and database program. These platforms are the essential foundations to build the other hierarchies. For example, the CFD program is the solver of thermal analysis. Similarly, the FEA program is the solver of vibration analysis. Then, the information management hierarchy employs a database program to manage the relevant product information. In the electronics reliability analysis, there is a large amount of involved information, such as device information, package information, joint information and environment load information. Most of the information can be accumulated and reused. Therefore, the reasonable organization and management of product information can improve software efficiency. Additionally, the function module hierarchy contains individual realization and interface relationships of different function modules. Finally, the user interface hierarchy provides a user interface that is used to manage and apply various function modules. Each function module can be individually applied using the GUI, which can obtain the required output of all information in table-based and graphical forms.

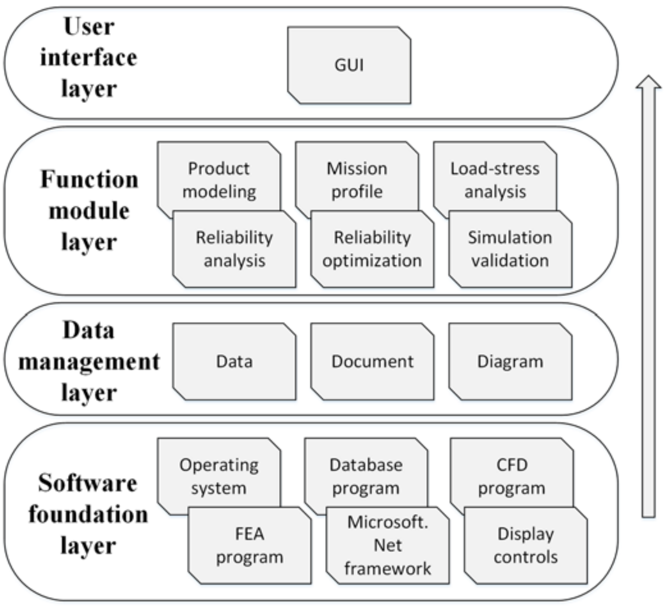


Fig. 2. Software hierarchy of PofEra

3.3. Information transfer interface design

The PofEra system achieves information integration for the comprehensive reliability simulation analysis of electronic products. The PofEra system avoids information interaction among the different software. Therefore, the system has complicated information transmission relationships. Fig. 3 illustrates the information transmission relationships among different modules. First, the PM module provides the product design information of the CAD model to the LSA module. According to the CAD model, the thermal and vibration analysis generate the CFD and FEA models, respectively. Then, the corresponding simulation analyses are carried out based on the local load information (temperature and vibration values) from the MP module. The SV module receives the experimental information from the hardware to verify the simulation model. Then, failure prediction in the RA module applies partial design information, simulation results and profile time information to predict the TTF of the device. Based on the device prediction results, a multipoint information fusion algorithm is employed to fit the life distribution of the circuit board and obtain the corresponding mean time to failure (MTTF) in the reliability analysis. Finally, according to the assessment results, the improved measures are fed back to the designers to optimize the electronic product. Because the PofEra system is not related to the product design, the transmission of feedback information is indicated by the dashed line.

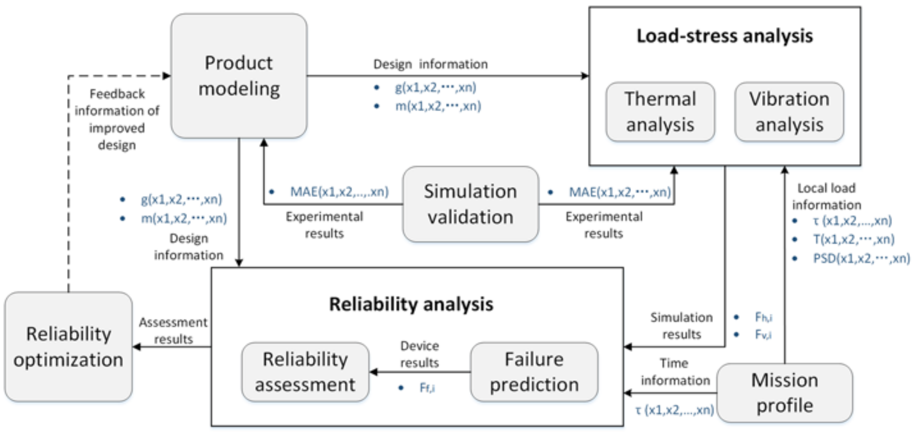


Fig. 3. Information transmission relationship of PofEra

The different function modules will generate different kinds of data types. Therefore, several data types should be considered in the application process, such as stl, XML and sheet. For the convenience of data transmission, most data should be transferred to the union data format; the XML format is selected as the union format. The transmission and transferability of data is achieved, as shown in Fig. 4.

The API program is developed in the C# language, which contains the PM program, MP program, LSA program, RA program, and RO program. The API program provides the commands and runs the analysis programs, which link the different modules. The function of the different programs is illustrated in Table 2. Then, the analysis results from the data of each module location are presented graphically using the GUI program. The display controls are used to display the 3D-model and graphical data. In the execution process, part of the data come from the database. The database can perform data selection and updating as well as deletion and insertion for the data tables.

3.4. Parallel computing

In addition, according to the requirement of the mission profile, multiple sets of load-stress analyses will be executed under different stress conditions, which is often very time-consuming. To improve the

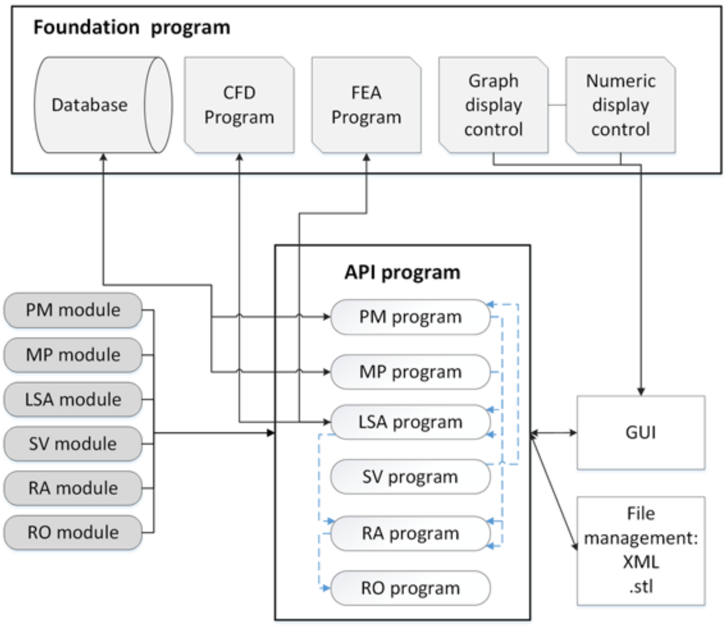


Fig. 4. Achievement of information transmission and transferability

computing efficiency, distributed resource allocation and parallel computing are applied to the PofEra system. The foundation programs are installed at a public server, including the CFD program, FEA program and SQL server. The clients access the corresponding services through a server. The thermal and vibration analysis based on the CFD program and FEA program can be simultaneously executed on the public server. Then, multiple sets of load-stress analysis results can be called by the reliability analysis module. The parallel computing schematic diagram is shown in Fig. 5.

Parallel computing can be divided into two phases: first, establishing a computing task execution sequence; second, task assignment and calculation. The task assignment process is shown in Fig. 6.

a) To complete failure prediction, the corresponding calculations are divided into three parts: thermal analysis $N_{h,i}$, vibration analysis $N_{v,i}$ and failure prediction $N_{f,i}$, where $i = 1,2,\cdots,n$. According to the random task allocation method, all the calculation tasks are reordered to form a new task calculation sequence. First, a random number sequence $\{R_{3n}\}$ is generated corresponding to the task sequence $Q = \{N_{3n}\}$. Then, the sequence $\{R_{3n}\}$ is rewritten in order from small to large to form the sequence $\{S_{3n}\}$. A new computation task sequence $Q' = \{L_{3n}\}$ is obtained according to the sequence $\{S_{3n}\}$. The new task sequence can make task assignment more uniform on each processor.

b) All computing tasks are assigned to the p computing nodes according to a new computing task sequence $Q' = \{L_{3n}\}$. After performing thermal analysis and vibration analysis, the corresponding data files ($F_{h,i}$, $F_{v,i}$, $i = 1,2,\cdots,n$) are generated. If the assigned failure prediction task has no corresponding data file for thermal analysis and vibration analysis, the calculation node suspends the response of the task assignment and performs the thermal analysis and vibration analysis that have

Table 2. Function of the API program

API program	Function
PM program	1. calling the basic data from the database 2. generating the '.stl' files containing the modeling information 3. generating the XML files containing the product design information
MP program	1. calling the basic data from the database 2. generating the XML files containing the profile information
LSA program	1. calling the '.stl' files and profile XML files for the CFD or FEA program 2. reading the analysis results from the CFD or FEA program 3. generating the XML files containing the analysis results
SV program	1. reading the experimental results from the hardware equipment 2. completing the error analysis between the simulation and test results 3. generating the XML files containing the error analysis results
RA program	1. calling the XML files (design information, profile and analysis results) 2. calculating the TTF of devices 3. generating the XML files containing the prediction results 4. assessing product reliability based on the prediction XML files and generating the assessment results
RO program	1. calling the XML files (prediction and assessment results) 2. carrying out the sensitivity analysis

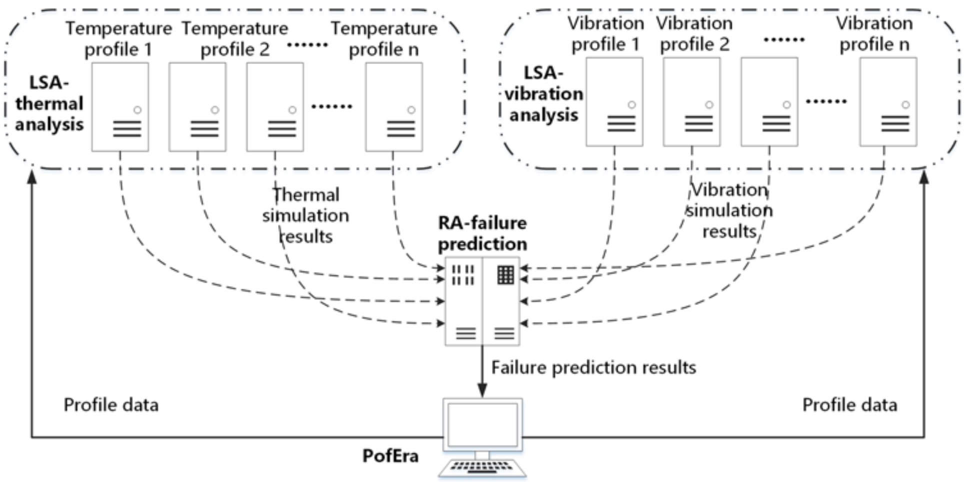


Fig. 5. Parallel computing schematic diagram

not yet been performed. The performed load analysis task will be removed in the task sequence to avoid repeated analysis. Then, the assigned failure prediction task is executed, and the other task assignment in the task sequence is accepted after the failure prediction task is executed.

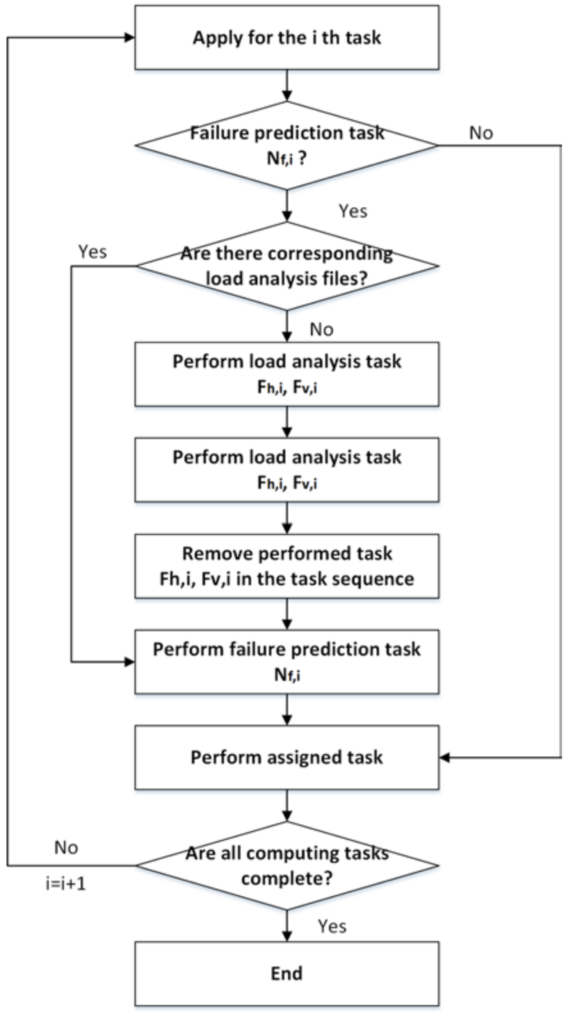


Fig. 6. The task assignment process

3.5. Model validation

The validation of digital simulation models is an important phase to promote the accuracy of assessment results. Through a consistency test between the simulation and physical test results, simulation credibility can be validated. Therefore, the simulation validation module provides the interface between the software and special hardware. Fig.7 shows the schematic diagram of the thermal and vibration tests. The hardware contains two types of channels to receive the thermal and vibration test data. The thermal test applies temperature sensors to collect temperature data. According to the thermal simulation results, certain high-temperature devices are selected as test objects to complete the thermal validation. The vibration test applies the hammer modal test to perform vibration validation. A displacement sensor is equipped on the tested circuit board. Then, the circuit board is hit multiple times with a hammer equipped with an acceleration sensor to collect feedback signals. The feedback signals are applied to calculate the various order modes of the circuit board. Then, the test data are inputted into the SV module. Based on the TIC and MACmodel validation functions, the module will automatically report the validation results of important devices. The product model will be revised based on validation results until the simulation results approach the test results. After determining the accuracy of the model, the simulation model can be applied to simulation analysis under other conditions, thus eliminating the conditional limitations of physical experiments.

4. Case study

As shown in Fig. 8, the filter module for an industrial communication system is selected as the validation case. The filter module is used to filter the interference signal and ensure the accuracy of the communication signal and the normal operation of the communication system.Table 3 displays some important device data. These devices are used to construct the filter circuit and perform the filter function. In the actual task environment, the filter module is not greatly affected by vibration, so this case only focuses on thermal effects. Fig. 9 is a partial task profile, which provides temperature and task time data for simulation analysis.

According to the above information, the CAD model of the filter module is established in the PM module, as shown in Fig. 10. In the modeling process, the model is properly simplified such that some non-important information is ignored. After completing the simulation model, the accuracy of the model needs to be verified. Here, a

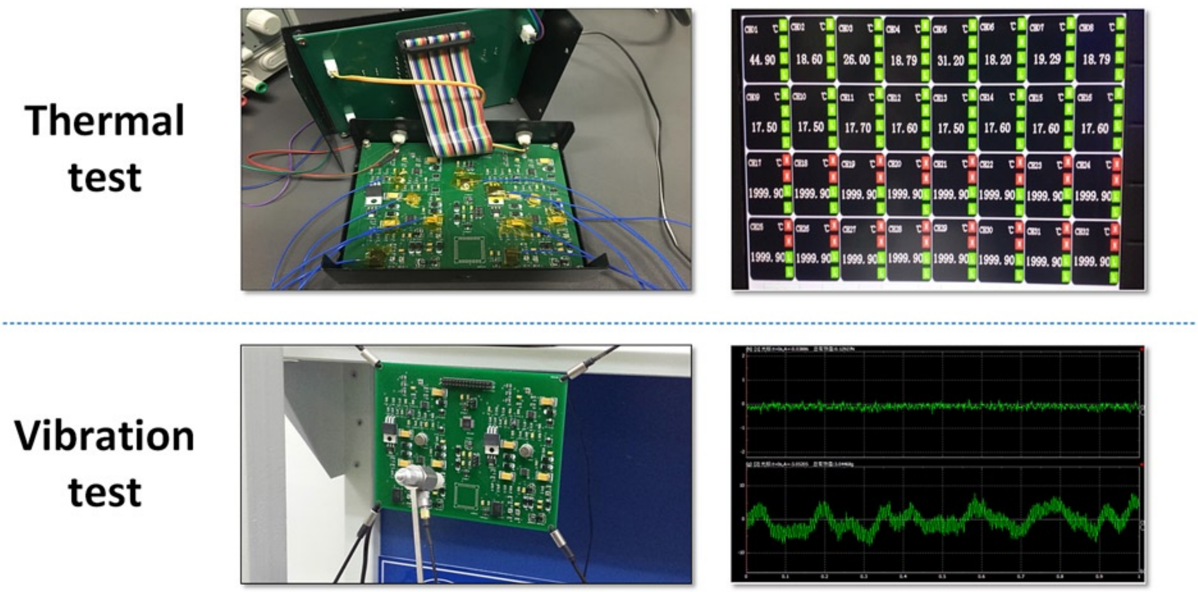


Fig. 7. Schematic diagram of the thermal and vibration tests

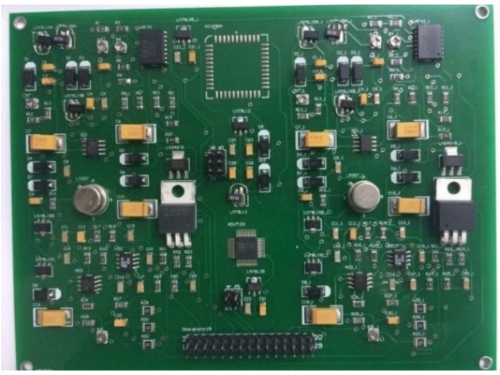


Fig. 8. Filter module

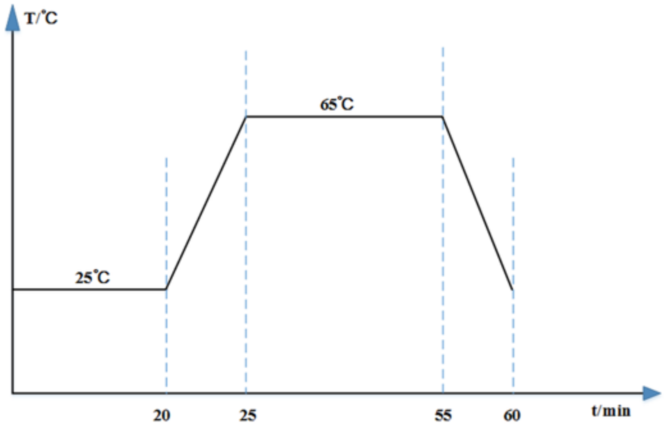


Fig. 9. Temperature profile

probability of potential failure. The amount of heat may lead to welding spot fatigue or open circuiting of the device. In the blue region, the two devices have a metal package, so they can achieve better heat dissipation.

After completing the thermal simulation analysis, failure prediction and reliability analysis can be sequentially executed. The expected lifetime of the filter module is assumed to be 5.5 years. The filter module runs 12 times a day according to the task profile, as shown in Fig. 9. As shown in Fig. 13(a), compared with the other devices, the red devices have a lower lifetime. For example, the TTF of C7 is 1.27×10^5 h. The underlying failure mechanism is first-order thermal fatigue. The devices marked in red can generate higher heat themselves, which easily causes thermal fatigue failure of the lead and welding spot. The devices in the white box are surface-mount devices that are close to high temperature areas and may have a life expectancy lower than the design value due to thermal fatigue failure. Then, based on the device failure prediction result, the reliability analysis result of the filter module can be obtained, as shown in Fig. 14(a). The MTTF of the filter module is 4.14×10^4 h, which is below the expected lifetime. Therefore, with a first round of reliability analysis completed, the analysis results can provide useful information for product reliability design improvements in the design stage. For the thermal design of circuit boards, there are generally three improved measures that can be adopted:

a) Choose other devices with better heat dissipation performance, such as metal-packaged devices.
b) Modify the layout of the original devices.
c) Introduce forced-cooling measures, such as air or liquid cooling.

thermal test is adopted to verify the model by comparing the thermal simulation results under the same environmental conditions. The thermal test is carried out under normal temperature conditions, that is, 25°C. High-power devices are selected as test objects by the test equipment. The SV module receives the hardware monitoring signal and displays the temperature results. The TIC is 0.04, and the model accuracy is satisfactory. The test results are shown in Table 4.

Then, based on the CAD model, thermal simulation analysis can be carried out. The corresponding simulation result is shown in Fig. 11. For electronic products, heat has a great impact on reliability. The red marks in Fig. 11 indicate regions with high heat generation. Generally, the devices in this region have a great

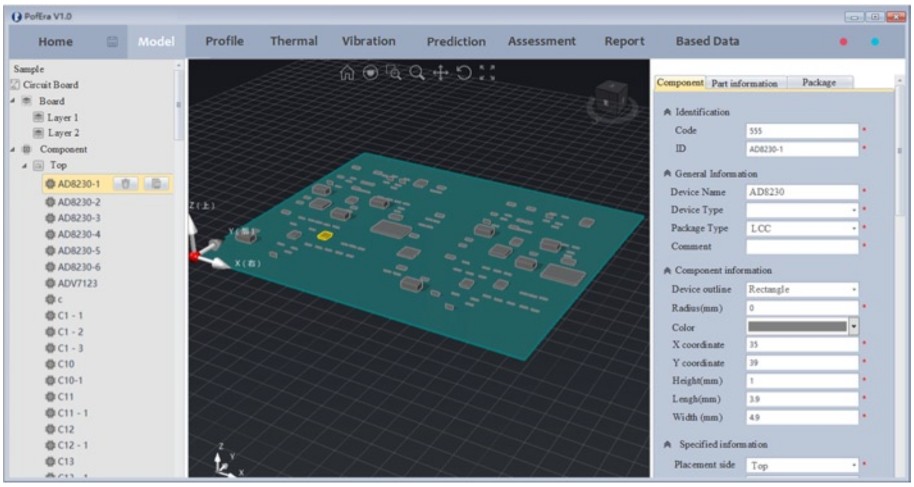


Fig. 10. The CAD model of the filter module

Table 3. Part of the device information

No.	Value	Device type	Length*Width *Height(mm)	Weight(g)	Power Consumption(W)
C1	100 pF	Capacitor	5.2*2.7*1.8	0.8	0.001
C2	47 μF	Capacitor	5.2*2.7*1.8	0.8	0.001
C3	0.22 μF	Capacitor	5.2*2.7*1.8	0.8	0.001
D1	/	Diode	2.5*1.7*1	0.1	0.001
D2	/	Diode	2.5*1.7*1	0.1	0.001
R1	5 kΩ	Resistor	2.2*1*0.7	0.06	0.001
R2	5.5 kΩ	Resistor	2.2*1*0.7	0.06	0.001
R3	25 kΩ	Resistor	2.2*1*0.7	0.06	0.001
AD620AN	/	Operational Amplifier	4.9*3.9*1.5	0.4	0.3

Table 4. Part of the device information

Number	**Measurement (℃)**	**Simulation (℃)**
C6	43	45
C7	43.5	47
D1	35	30
D2	30	32
R1	26	29
R2	27.5	29.6
AD620AN	28	29.6

In this paper, we assume that the air-cooling measure is selected to achieve better heat dissipation and improve the product reliability. Then, a new thermal simulation and reliability analysis are carried out. The corresponding analysis results are displayed in Figs. 12-14. It can be seen from Fig. 12 that the heat concentration region disappears. Compared with Fig. 11, the maximum temperature is reduced by approximately 4℃. The TTF of C7 is increased to 2.12×10^5 h. The TTF is 66.9% higher than the predicted lifetime without improved measures. The prediction lifetimes of other devices have also increased. The MTTF of the filter module is raised to 5.23×10^4 h, which is an improvement of approximately 26.3%. In addition, it can be seen from Fig. 14 that the initial inflection point of the reliability curve has been delayed to 3.0×10^4 h.

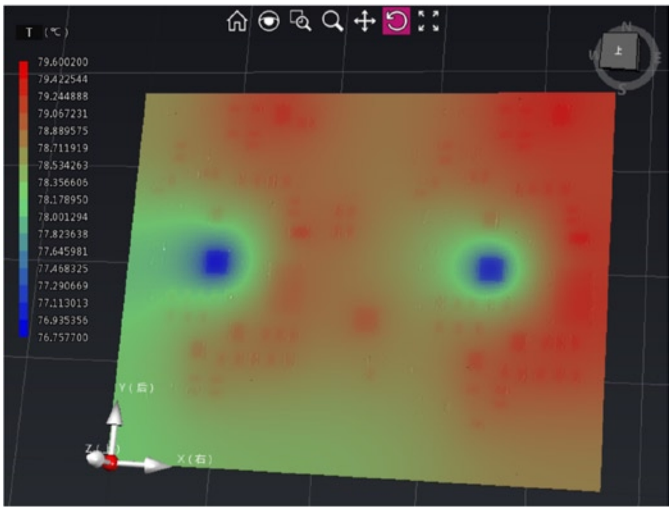


Fig. 11 The thermal simulation result of the filter module

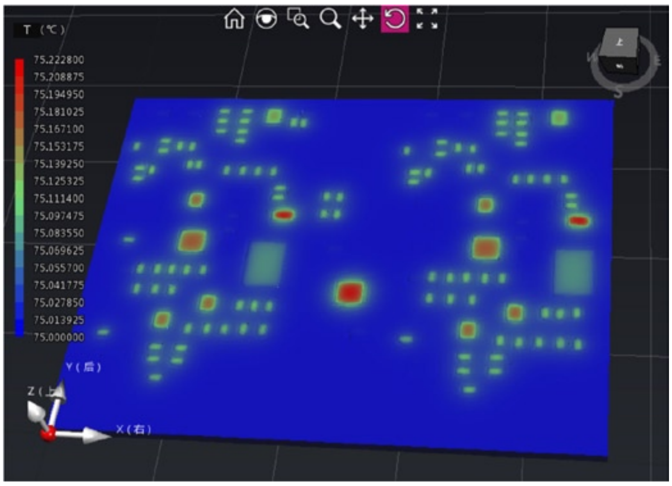


Fig. 12. The thermal simulation result with the improved measure

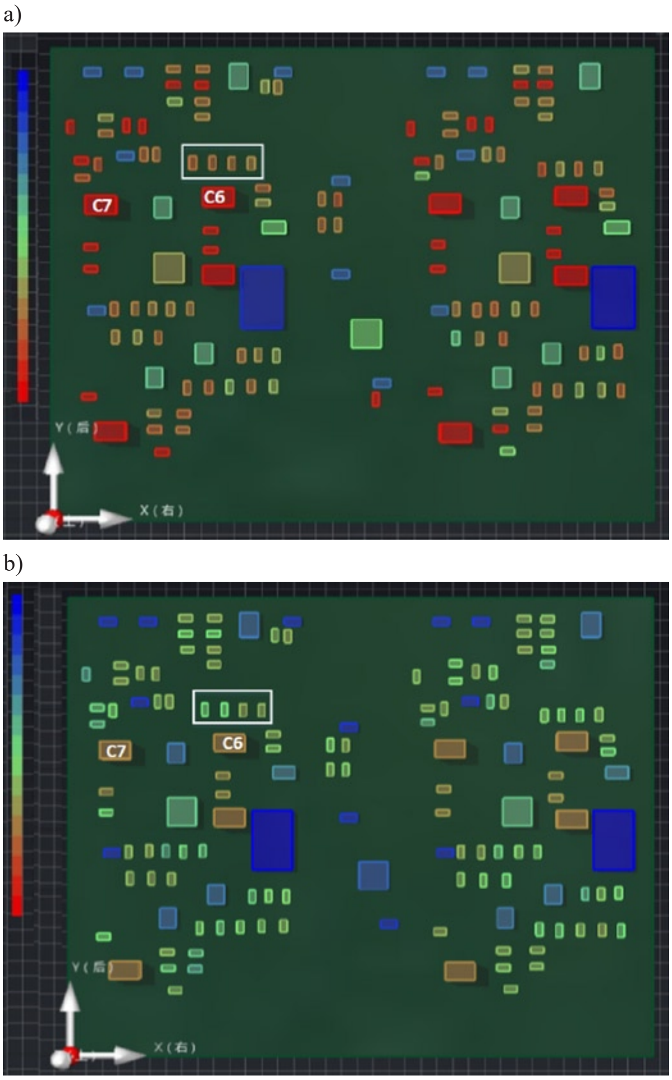


Fig. 13. Failure prediction result (a) without the improved measure (b) with the improved measure

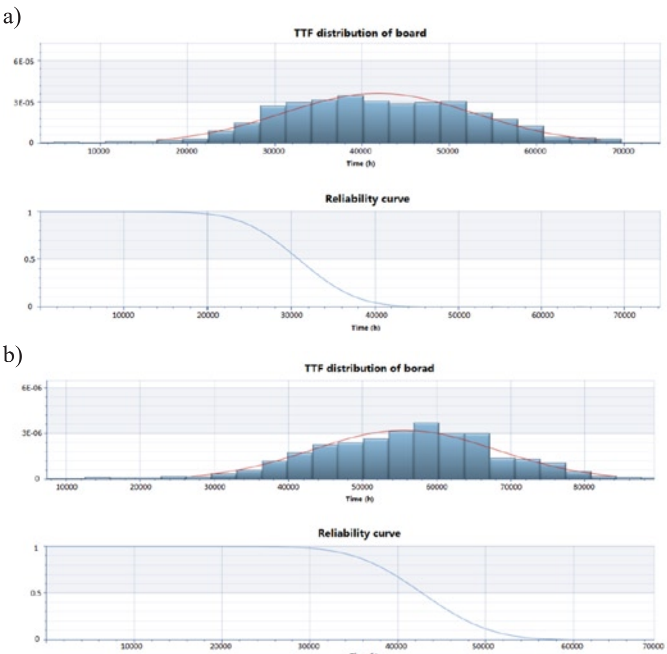


Fig. 14. Reliability analysis result (a) without the improved measure (b) with the improved measure

To validate the prediction accuracy, the errors between the prediction and measurement data need to be calculated. Therefore, the root-mean-square error (RMSE) is selected as the evaluation criterion to evaluatethe prediction performance as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(l_{pre}(i) - l_{true}(i)\right)^2}, \tag{9}$$

where $l_{pre}(i)$ is the predicted lifetime of devices and $l_{true}(i)$ is the true lifetime of devices.

As a widely used industrial product, we have collected the historical lifetime data of some filter module devices. These historical lifetime data can be used as the true lifetime data to evaluate the prediction accuracy. Then, we select eight typical devices and calculate the RMSE. As shown in Fig. 15, the average RMSE is 0.318. The small error value indicates that the prediction accuracy is high. The fluctuation of the curve is gentle, which indicates that the prediction results are more stable.

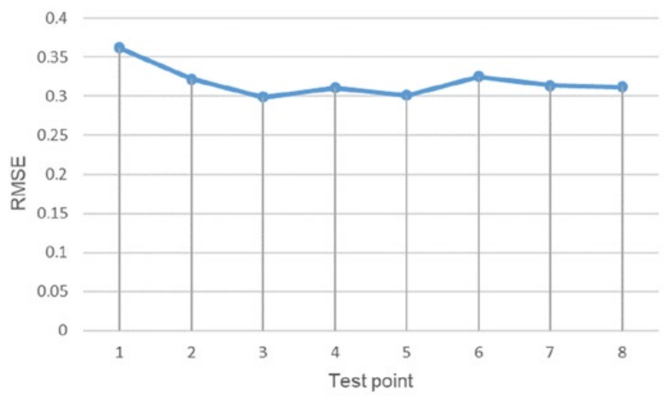


Fig. 15. Prediction error for the filter module

To compute the efficiency of parallel computing in this system, we choose serial computing under the same preconditions as the comparison object. The preconditions are as follows: First, the thermal condition and MonteCarlosimulation number (N=1000) in the failure prediction are the same for parallel computing and serial computing. Second, a computer with a quad-core CPU and 64 GB memory is used to execute parallel computing and serial computing. Therefore, under the same precondition, the calculation efficiency is compared according to the calculation time.The following efficiency function is designed to evaluate the improvement of computational efficiency:

$$CE_i = \frac{t_i}{\sum_{i \in s} t_i}, \tag{10}$$

where $s = (s_p, s_s)$ represents parallel computing and serial computing, respectively. t_i is the computational time of parallel computing or serial computing. The smaller the value of CE, the higher the computational efficiency. As shown in Fig. 16, the horizontal axis represents the number of simulation tasks. When the number of simulation tasks is 1, the computing efficiency of parallel computing and serial computing are equal. As the number of simulation tasks increases, the computing efficiency of parallel computing gradually increases, and the computing efficiency of serial computing gradually decreases. The average value of CE_{s_p} is 0.298, and the average value of CE_{s_s} is 0.702. Therefore, parallel computing is 40% more efficient than serial computing.

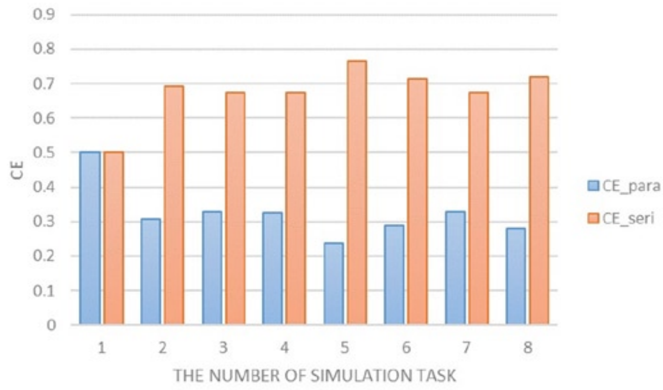


Fig. 16. The computational efficiency of parallel computing and serial computing

5. Conclusion

In this paper, the physics-of-failure and computer-aided simulation fusion approach as well as a corresponding software system are presented. The software system applies the PoF approach and computer-aided simulation method. Based on the integrated technology, the CAD, CFD and FEA techniques are integrated into the application system. This provides convenient and accurate access to the PoF method in engineering applications for electronic product reliability.

In the software system, the API programs have a significant impact on providing integration between the information and function modules in a uniform data format. Based on the API programs, the different function modules can not only complete their functions independently but also perform information interchange in the same software environment. Information conversion among different commercial software programs is avoided. Moreover, thermal and vibration testing are also integrated into the system framework, which allows the system to combine hardware and software simultaneously. The simulation validation and corresponding hardware tests can improve the accuracy of the reliability analysis. Therefore, with the integrated technique, the software reduces the application difficulty of the PoF approach in the engineering field. Because of the integration of advanced simulation techniques, the accuracy of the reliability analysis is also improved. Parallel computing is applied to the integration system to improve computational efficiency. Finally, based on the method and system of this paper, the remaining useful life can be evaluated by inputting the life load profile obtained from actual operation, and thus, health management can be realized.

Acknowledgement

The authors would like to thank the National Natural Science Foundation of China (No. 61673037) for support.

References

1. Allemang RJ. The modal assurance criterion - twenty years of use and abuse. Sound and Vibration, 2003, 37: 14-23.
2. Chatterjee K, Modarres M, Bernstein J B, Nicholls D. Celebrating fifty years of physics of failure. In 2013 Proceedings Annual Reliability and Maintainability Symposium (RAMS), IEEE 2013: 1-6, https://doi.org/10.1109/RAMS.2013.6517624.
3. Evans J, Cushing M J, Lall P, Bauernschub R. A physics-of-failure (POF) approach to addressing device reliability in accelerated testing of MCMs. In Proceedings of 1995 IEEE Multi-Chip Module Conference (MCMC-95), IEEE, 1995: 14-25.
4. Foucher B, Boullie J, Meslet B, Das D. A review of reliability prediction methods for electronic devices. Microelectronics Reliability 2002; 42(8): 1155-1162, https://doi.org/10.1016/S0026-2714(02)00087-2.
5. Gholami R, Watson R T Molla A, Hasan H, Bjørn-Andersen N. Information systems solutions for environmental sustainability: how can we do more?. Journal of the Association for Information Systems 2016; 17(8): 521-536, https://doi.org/10.17705/1jais.00435.
6. Hariharan A, Adam M T P, Lux E, Pfeiffer J, Dorner V, Müller MB, Weinhardt C. Brownie: A platform for conducting NeuroIS experiments. Journal of the Association for Information Systems 2017; 18(4): 264-296, https://doi.org/10.17705/1jais.00457.
7. http://www.calce.umd.edu/index.html
8. https://www.dfrsolutions.com
9. http://www.reliability-safety-software.com
10. Irias NG, Souza FA, de Paula H, Silva L A R. Challenges in using the physics-of-failure approach in practical applications. In 2017 IEEE Industry Applications Society Annual Meeting 2017: 1-8, https://doi.org/10.1109/IAS.2017.8101838.
11. Irias N G D N, de Paula H. Practical application of the physics-of-failure approach: Software and new resources for reliability prediction of power electronics systems. In 2017 Brazilian Power Electronics Conference (COBEP), IEEE 2017: 1-6, https://doi.org/10.1109/COBEP.2017.8257244.
12. Kimseng K, Hoit M, Tiwari N, Pecht M. Physics-of-failure assessment of a cruise control module. Microelectronics Reliability 1999; 39(10): 1423-1444, https://doi.org/10.1016/S0026-2714(99)00018-9.
13. Kostandyan E E, Sørensen J D. Physics of failure as a basis for solder elements reliability assessment in wind turbines. Reliability Engineering & System Safety 2012; 108, 100-107, https://doi.org/10.1016/j.ress.2012.06.020.
14. Kheir NA, Holmes WM. On Validating Simulation Models of Missile Systems. Simulation 1978; 30(4): 117-128, https://doi.org/10.1177/003754977803000404.
15. Lall P, Pecht M. An integrated physics-of-failure approach to reliability assessment. Advances in Electronic Packaging 1993: 509-524.
16. Ma K, Wang H, Blaabjerg F. New approaches to reliability assessment: Using physics-of-failure for prediction and design in power electronics systems. IEEE Power Electronics Magazine 2016; 3(4): 28-41, https://doi.org/10.1109/MPEL.2016.2615277.
17. Maass W, Parsons J, Purao S, Storey V C, Woo C. Data-driven meets theory-driven research in the era of big data: opportunities and challenges for information systems research. Journal of the Association for Information Systems 2018; 19(12): 1253-1273, https://doi.org/10.17705/1jais.00526.
18. McLeish J, Blattau N. CAE apps for physics of failure reliability & durability simulations. In 2014 Reliability and Maintainability Symposium, IEEE 2014: 1-6, https://doi.org/10.1109/RAMS.2014.6798444
19. Oh H, Azarian MH, Pecht M, White CH, Sohaney RC, Rhem E. Physics-of-failure approach for fan PHM in electronics applications. In 2010 Prognostics and System Health Management Conference, IEEE 2010: 1-6, https://doi.org/10.1109/PHM.2010.5413501.
20. Oh H, Han B, McCluskey P, Han C, Youn BD. Physics-of-failure, condition monitoring, and prognostics of insulated gate bipolar transistor modules: A review. IEEE Transactions on Power Electronics 2014; 30(5): 2413-2426, https://doi.org/10.1109/TPEL.2014.2346485.
21. Pecht M. Electronic reliability engineering in the 21st century. In Advances in Electronic Materials and Packaging 2001 (Cat. No. 01EX506), IEEE 2001: 1-7.
22. Pecht MG, Nash FR. Predicting the reliability of electronic equipment. Proceedings of the IEEE 1994; 82(7): 992-1004, https://doi.org/10.1109/5.293157.
23. Pecht M, Dasgupta A, Barker D, Leonard CT. The reliability physics approach to failure prediction modelling. Quality and Reliability Engineering International 1990; 6(4): 267-273, https://doi.org/10.1002/qre.4680060409.
24. Pecht M, Dasgupta A. Physics-of-failure: an approach to reliable product development. Journal of the IES 1995; 38(5): 30-34, https://doi.org/10.1109/IRWS.1995.493566.
25. Pecht M, Gu J. Physics-of-failure-based prognostics for electronic products. Transactions of the Institute of Measurement and Control 2009; 31(3-4): 309-322, https://doi.org/10.1177/0142331208092031.
26. Sun, B., Zhao, Y., Huang, W., Xie, J. S., & Kang, R. Case study of prognostic and health management methodology for electronics. Systems Engineering and Electronics, 2007; 29(6): 1012-1016.
27. Sun B, Zeng S, Kang R, Pecht MG. Benefits and challenges of system prognostics. IEEE Transactions on Reliability 2012; 61(2): 323-335, https://doi.org/10.1109/TR.2012.2194173.
28. Sun B, Jiang X, Yung KC, Fan J, Pecht MG. A review of prognostic techniques for high-power white LEDs. IEEE Transactions on Power Electronics 2016; 32(8): 6338-6362, https://doi.org/10.1109/TPEL.2016.2618422.
29. Singh SSK, AbdullAh S, Mohamed NAN. Reliability analysis and prediction for time to failure distribution of an automobile crankshaft. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2015; 17(3): 408-415, https://doi.org/10.17531/ein.2015.3.11.
30. White M. Microelectronics reliability: physics-of-failure based modeling and lifetime evaluation. Pasadena, CA: Jet Propulsion Laboratory, National Aeronautics and Space Administration, 2008.
31. Wang Y, Fang X, Zhang C, Chen X, Lu J. Lifetime prediction of self-lubricating spherical plain bearings based on physics-of-failure model and accelerated degradation test. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18(4): 528-538, https://doi.org/10.17531/ein.2016.4.7.
32. Wu XZ. Assessing the correlated performance functions of an engineering system via probabilistic analysis. Structural Safety 2015; 52: 10-19, https://doi.org/10.1016/j.strusafe.2014.07.004.
33. Zhang Y, Ma Y, Liu L, Ouyang 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, https://doi.org/10.17531/ein.2018.4.9.

Bo SUN
Yu LI
Zili WANG
Zhifeng LI
Quan XIA
Yi REN
Qiang FENG
Dezhen YANG
Cheng QIAN
School of Reliability and Systems Engineering
Beihang University
100191, Beijing, China

E-mails: sunbo@buaa.edu.cn, liyu822@buaa.edu.cn, wzl@buaa.edu.cn, handanlinzhang@163.com, quanxia@buaa.edu.cn, renyi@buaa.edu.cn, fengqiang@buaa.edu.cn, dezhenyang@buaa.edu.cn, cqian@buaa.edu.cn

Article citation info:
ROMANIUK M, HRYNIEWICZ O. Estimation of maintenance costs of a pipeline for a U-shaped hazard rate function in the imprecise setting. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 352–362, http://dx.doi.org/10.17531/ein.2020.2.18.

Maciej ROMANIUK
Olgierd HRYNIEWICZ

ESTIMATION OF MAINTENANCE COSTS OF A PIPELINE FOR A U-SHAPED HAZARD RATE FUNCTION IN THE IMPRECISE SETTING

ESTYMACJA KOSZTÓW EKSPLOATACYJNYCH RUROCIĄGU DLA U-KSZTAŁTNEJ FUNKCJI INTENSYWNOŚCI USZKODZEŃ PRZY NIEPRECYZYJNYM PODEJŚCIU

In this paper, we discuss imprecise settings for an evaluation of the maintenance costs of a water distribution system (WDS). Moments of failures of pipes are modelled using a newly proposed three-piece convex hazard rate function (HRF) for which number of previous failures is taken into account, too. Both fuzzy sets and shadowed sets are used to model the impreciseness of important parameters of this HRF and the costs of maintenance services. Contrary to more classical and widely-used approaches to cost analysis (i.e. a constant yield or nominal value of money), a strictly stochastic process (i.e. the one-factor Vasicek model) of an interest rate is assumed in the analysis of maintenance costs. This approach models future behaviour of the interest rate (i.e. the future value of money) in a more realistic way. Respective algorithms together with exemplary results of numerical simulations for two setups, which are related to fuzzy and shadowed sets, are also provided.

Keywords: *water distribution system, maintenance costs, convex hazard rate function, Monte Carlo simulations, fuzzy sets, shadowed sets.*

W niniejszym artykule omawiamy nieprecyzyjne podejścia do problemu obliczenia kosztów eksploatacji systemu dystrybucji wody (WDS). Czasy uszkodzeń rur modelowane są z wykorzystaniem nowo zaproponowanej trzyczęściowej wypukłej funkcji intensywności uszkodzeń (hazard rate function, HRF) dla której brana jest pod uwagę również liczba wcześniejszych uszkodzeń. Do modelowania nieprecyzyjności istotnych parametrów tej HRF oraz kosztów działań serwisowych są wykorzystywane zarówno zbiory rozmyte jak i zbiory cieniowane. W przeciwieństwie do bardziej klasycznych i szeroko wykorzystywanych podejść do analizy kosztów eksploatacji (tzn. stałej stopy procentowej lub wartości nominalnej pieniądza), założono ściśle stochastyczny proces (tzn. jednoczynnikowy model Vasicka) dla stopy procentowej. Podejście to modeluje przyszłe zachowanie stopy procentowej (czyli przyszłej wartości pieniądza) w bardziej realistyczny sposób. Zaprezentowano również odpowiednie algorytmy wraz z przykładowymi wynikami symulacji numerycznych dla dwóch zestawów parametrów, związanych ze zbiorami rozmytymi i cieniowanymi.

Słowa kluczowe: *system dystrybucji wody, koszty eksploatacji, wypukła funkcja intensywności uszkodzeń, symulacje Monte Carlo, zbiory rozmyte, zbiory cieniowane.*

1. Introduction

To deliver water of desirable quality and in necessary quantity, various maintenance services (like repairs and replacements of connections) for a water distribution system (WDS) are necessary. The literature devoted to different aspects of these problems, like modelling reliability of a WDS or a calculation of costs of the maintenance services, is abundant. We refer the reader to some detailed and interesting reviews, e.g., [16, 29, 30]. The articles related to the problem of maintenance of a WDS are really diversified, too. Some of them discuss hydraulic and physical characteristics of parts of a WDS (see, e.g., [5, 18]), other focus on a "macro-management" of a WDS (see, e.g., [22]) or its "micro-management" scale (see, e.g., [1]), or propose the application of artificial neuronal nets in a monitoring system (see, e.g., [23]) or even artificial intelligence (see, e.g., [24]).

In this paper, we discuss a simulation approach to the estimation of the maintenance costs related to repairs and replacements of pipes. Usually, to calculate these costs in an appropriate manner, a relatively long-time horizon has to be considered. Such a time interval covers 20, 50 or even 60 years (see, e.g., [12]). But in most of the papers, a constant rate for money flow or even nominal costs are considered, which is rather counter-intuitive, because one unit of money, which is paid now, and the same unit in 50-60 years, are not equal. Moreover, the assumption about a constant discount factor is too strong and unrealistic for the mentioned long time horizons. Therefore, a variable interest rate should be considered in a more real-life approach. Then, in this paper, for modelling such variable interest rates we apply the one-factor Vasicek model.

In the literature, many models of intensities of malfunctions of parts of a WDS have been proposed. Some of them are related to physical aspects of a pipe and are described by formulas (like the Hazen-Williams equation, see, e.g., [12]). Other models are based on Markov or semi-Markov processes (see, e.g., [13, 17, 25]) or are described using a hazard rate function (HRF, see, e.g., [30] for a comprehensive review). In this paper, we propose a new kind of HRF that describes three important stages of the "life" of a pipe and takes into account an increasing deterioration of a material of a pipe. The proposed HRF completely fulfils requirements formulated by some authors (see, e.g., [30]). It can be easily adapted to real-life data. Moreover, the respective generation algorithm is very efficient in the Monte Carlo simulations which are widely used in modelling complex systems (see, e.g., [3]).

This paper can be seen as further development of ideas proposed earlier in [25, 26], as some new concepts are also considered. Thus, our contribution to problems considered in this paper is five-fold.

First, a new, more general hazard rate function, which describes the intensities of malfunctions of pipes in a WDS, is introduced. Contrary to other HRFs, which were previously discussed in the literature, it has some appealing features. It is U-shaped and it models three important states of a connection: a starting burn-in period (immediately after a repair or an installation of a pipe, when the intensity of malfunctions is a decreasing function), a middle stable state (when initial problems with a connection have passed, so the level of failures is relatively low), and a later wear-out period (when the intensity of malfunctions is an increasing function, with higher values than during its stable state). This HRF also depends on number of previous repairs of the given pipeline, so an increasing deterioration, which is caused by recurring stresses related to repairs, can be taken into account. Moreover, a numerically efficient algorithm for the generation of random times of failures for this HRF is also provided. It leads us to a direct application of the Monte Carlo (MC) simulations to simulate the behaviour of the whole WDS. Contrary to [26], the newly proposed HRF has three, instead of only two states (i.e. it is U-shaped instead of its previous V-shaped version).

Second, apart from the mentioned HRF, we use a special random distribution with a decreasing intensity function and finite support to model times of services related to malfunctions of a WDS (i.e., repairs and replacements of pipes). A numerically efficient algorithm for the generation of random times of these services is provided, too. Therefore, the MC approach can be used to simulate the respective values of the mentioned times.

Third, almost all of the parameters of the model (apart from these related to the variable interest rate, an unconditional replacement age, and time horizon for the Monte Carlo simulations) are fuzzified to express our limited information about their real values. Moreover, we provide a general framework for using different kinds of fuzzy numbers (like triangular fuzzy numbers, trapezoidal fuzzy numbers or left-right fuzzy numbers) as the parameters of the model to calculate the present value of the maintenance costs or other important characteristics of a WDS. This framework can be then applied to other models of failure intensities for a WDS. Therefore, we generalize our previous considerations from [25] to a more general fuzzy approach including new types of fuzzy numbers.

Fourth, apart from the imprecise setting related to fuzzy numbers, we apply shadowed sets to describe the parameters of the model. An introduction of shadowed sets can be very fruitful because it enables us to both consider impreciseness (which is, e.g., related to an expert's opinion) and to dramatically limit amount of necessary numerical simulations to estimate the considered characteristic of the model (like the present value of the maintenance costs) in comparison with the previously mentioned fuzzy setting.

Fifth, apart from the theoretical framework for both fuzzy numbers and shadowed sets setups, we provide respective numerical algorithms together with examples of numerical simulations based on the Monte Carlo approach. Some important measures of a WDS, like the present value of the maintenance costs, are approximated using fuzzy numbers and shadowed sets. Therefore, apart from a "strict" value (related to the core of a fuzzy / shadowed set), an additional "imprecise" interval (related to the support of a fuzzy / shadowed set) is calculated.

It should be pointed out, that the stochastic model of the interest rate (i.e., the one factor-Vasicek model) is directly embedded into our Monte Carlo simulations, as in [26]. To our best knowledge, application of the variable interest rate is still a new idea, which is not considered in other papers. But there are significant differences in outputs (like estimated values of costs of a WDS, see, e.g., [25, 26]), between models with a constant yield and with a variable discount factor. Therefore the stochastic model of the interest rate should be preferred in an estimation of the maintenance costs.

This paper is organized as follows. In Section 2, a new type of a U-shaped hazard rate function, which describes times of the failures of a connection, is presented. A numerical algorithm for a respective density, which is based on this HRF, is also provided. In Section 3 and 4, a model of maintenance costs for a WDS based on constant and random variable costs of repairs and replacements, together with a new random distribution for modelling times of repairs and replacements, are discussed. Section 5 is devoted to two imprecise settings related to fuzzy numbers and shadowed sets. Respective algorithms for these setups are also provided there. We illustrate these two approaches in Section 6 with examples of numerical analysis using the Monte Carlo simulations. The paper is concluded in Section 7.

2. Properties of U-shaped HRF

Let us suppose, that the considered WDS is modelled by a graph of connections G. In this graph, each connection (i.e., a pipeline which is a part of this WDS) is represented as an edge, and possible sources or outflows are denoted by nodes. In the following, we focus only on the edges of the graph G, i.e., the connections of the WDS. Let us assume, that these connections behave in a statistically independent way, i.e., time of a malfunction of one pipe does not influence the quality and possible malfunctions of other pipes. In the literature, multistate systems with embedded dependencies of components are also considered (see, e.g., [2]).

We assume that times of failures for each connection are described by a hazard rate function (abbreviated further as HRF) $\lambda(x|n_r)$, which is given by the formula:

$$\lambda(x|n_r)=\begin{cases}-a_0x+b_0+\alpha_r n_r & if \quad x\in\left[0,x_0^*\right)\\ a_1x+b_1+\alpha_r n_r & if \quad x\in\left[x_0^*,x_1^*\right)\\ a_2x+b_2+\alpha_r n_r & if \quad x\ge x_1^*\end{cases}, \tag{1}$$

where $a_0>0, a_1>0, a_2>a_1, b_0\ge a_0x_0^*, x_0^*>0$, $x_1^*>0, \alpha_r>0$ and

$$b_1=-x_0^*\left(a_0+a_1\right)+b_0 \quad , b_2=x_1^*\left(a_1-a_2\right)+b_1 .$$

As it is seen in Figure 1, this HRF is a U-shaped function with three linear segments, for which:

- $-a_0$ is a directional component of the descending linear part of this HRF (i.e., a left-hand side of this function, for which $x\in\left[0,x^*\right)$),
- a_1 is a directional component of the middle, ascending linear part (i.e., when $x\in\left[x_0^*,x_1^*\right)$),
- a_2 is a directional component of the right-hand side, ascending linear part (i.e., when $x\ge x_1^*$),
- b_0 is related to a vertical shift of the whole HRF,
- x_0^* and x_1^* are horizontal values of the points, where this HRF changes its behaviour,
- α_r is a parameter of deterioration, which is related to a single, previous malfunction of a connection,
- n_r is number of previous malfunctions of a connection, if there were repairs afterwards.

When a connection is replaced with a completely new component, then $n_r=0$ is set. It means that the previously mentioned parameter α_r reflects a level of fatigue related to prior malfunctions and repairs (see also [26]). Such an influence of the number of previous malfunctions on the current condition of a connection is frequently postulated in the literature (see, e.g., [30]). In real-life applications the above-mentioned parameters of the HRF (1) should be properly adjusted to the existing data, because number of failures varies, e.g., in [6], the average rate of failures is reported as values from 2.3 up to 34.8 (per 100 miles of pipes per a year) depending on a material of the considered pipe.

The values x_0^* and x_1^* are connected with three important states of a connection (see also, e.g., [4]): its initial burn-in period (after a previous repair or just after an installation of a new pipe), a stable state (when possible initial problems have passed, so the level of failures is relatively low) and a wear-out period (when the intensity of failures increases with passing time, because of existing problems of "old age" of the connection).

This new HRF, which is given by (1), is a more complex, U-shaped function if it is compared to its previous V-shaped counterpart, which was introduced in [26]. Moreover, a new additional state (the stable state) can be modelled using (1). Then, this HRF can be used to describe the intensity of malfunctions, taking into account three completely different states of quality of a pipeline and progress of fatigue of a connection (which is related to the number of previous repairs n_r). Therefore, this HRF can be better adjusted to real-life data, if it is compared to other types of functions discussed in the literature. This HRF also meets the requirements concerning functions describing the intensity of malfunctions (see [30] for additional details).

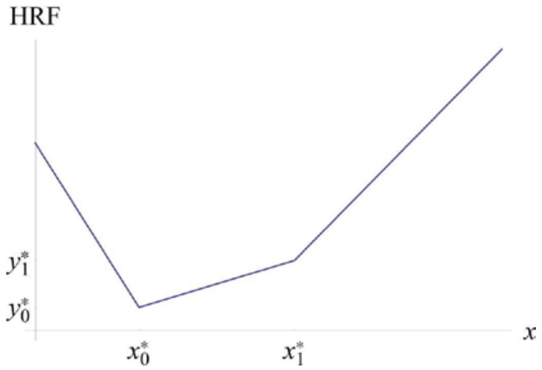


Fig. 1. Exemplary plot of the introduced HRF

To simplify further formulas, let us assume that:

$$b_0^{'} = b_0 + \alpha_r n_r \ , \ b_1^{'} = b_1 + \alpha_r n_r \ , \ b_2^{'} = b_2 + \alpha_r n_r \ .$$

Because:

$$\lambda(x) = \frac{f(x)}{R(x)} \ , \tag{2}$$

where $f(x)$ is a pdf (probability density function), $R(x)=1-F(x)$ and $F(x)$ is a cdf (cumulative density function), then in the case of (1), we get:

$$f(x) = \begin{cases} \left(-a_0 \, x + b_0^{'}\right)\exp\left(\frac{1}{2}a_0 \, x^2 - b_0^{'} \, x\right) & if \quad x \in \left[0, x_0^*\right) \\ \left(a_1 \, x + b_1^{'}\right)\exp\left(-\frac{1}{2}a_1 \, x^2 - b_1^{'} \, x + c_1\right) & if \quad x \in \left[x_0^*, x_1^*\right) \\ \left(a_2 \, x + b_2^{'}\right)\exp\left(-\frac{1}{2}a_2 \, x^2 - b_2^{'} \, x + c_2\right) & if \quad x \geq x_1^* \end{cases} \tag{3}$$

where:

$$c_1 = \frac{1}{2}\left(a_0 + a_1\right)\left(x_0^*\right)^2 + \left(-b_0^{'} + b_1^{'}\right)x_0^*,$$

$$c_2 = c_1 + \frac{1}{2}\left(-a_1 + a_2\right)\left(x_1^*\right)^2 + \left(-b_1^{'} + b_2^{'}\right)x_1^*$$

In Figure 2, an exemplary plot of the density (3) with the two previously mentioned points, where the HRF changes its behaviour ($x_0^* = 0.8, x_1^* = 3$, namely), is provided.

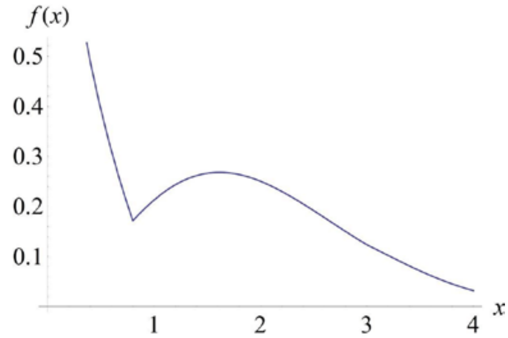


Fig. 2. Exemplary plot of the density for the introduced HRF

To simulate random times of malfunctions, it is necessary to provide a numerically efficient algorithm, which generates random variables based on (3). It can be done using the composition method and the inversion method (see, e.g., [28] for a necessary introduction). If the composition approach is applied, then a respective pdf $f(x)$ is given by:

$$f(x) = \sum_{i=1} f_i(x) p_i \ ,$$

where $f_i(x) \geq 0$ is a density and $p_i \geq 0$ is a discrete probability for $i = 1, 2, \ldots$. In the case of (3), we have:

$$p_1 = P\left(X \in \left[0, x_0^*\right)\right) = 1 - \exp\left(\frac{1}{2}a_0\left(x_0^*\right)^2 - b_0^{'} \, x_0^*\right),$$

$$p_2 = p_2^{'} \ \exp\left(x_0^*\left(-b_0^{'} + \frac{1}{2}a_0 x_0^*\right)\right),$$

$$p_3 = 1 - p_1 - p_2 \ , \tag{4}$$

where

$$p_2^{'} = 1 - \exp\left(\left(x_0^* - x_1^*\right)\left(b_1^{'} + \frac{1}{2}a_1\left(x_0^* + x_1^*\right)\right)\right) \ .$$

Using the inversion method, for $f_1(x)$ (if $x \in \left[0, x^*\right)$), $f_2(x)$ (when $x \in \left[x_0^*, x_1^*\right)$) and $f_3(x)$ (when $x \geq x_1^*$), the respective inversions of their cdfs (compare also with [26]) are given by

$$F_1^{-1}(y) = \frac{b_0^{'} - \sqrt{\left(b_0^{'}\right)^2 + 2a_0 \ln\left(1 - p_1 \, y\right)}}{a_0} \ ,$$

$$F_2^{-1}(y)=\frac{-b_1'+\sqrt{\left(b_1'+a_1x_0\right)^2-2a_1\ln\left(1-p_2'y\right)}}{a_1},$$

$$F_3^{-1}(y)=\frac{-b_2'+\sqrt{\left(b_2'+a_2x_1^*\right)^2-2a_2\ln(1-y)}}{a_2}. \quad (5)$$

Then, Algorithm 1 can be directly used to simulate times of the failures in a numerically efficient way.

Algorithm 1 (Generation procedure for the HRF)
Input: A set of the parameters of the HRF (1).
Output: A random time of a failure *X*.

Calculate p_1,p_2,p_3 *, which are given by (4);*
Generate independent random values U,Y *from the uniform standard distribution* $U[0;1]$*;*

if $U\le p_1$
$X=F_1^{-1}(Y)$ *(see (5));*
else
if $U\le p_2$
$X=F_2^{-1}(Y)$ *(see (5));*
else
$X=F_3^{-1}(Y)$ *(see (5));*
return *X*

The expected value of the density (3) can be numerically computed, e.g., for $a_0=0.6,a_1=0.2,a_2=0.8,b_0=0.65,x_0^*=0.5,x_1^*=10,\alpha_r=0.1,n_r=0$ we get 4.53037 (about 4.5 years if time unit is assumed to be a year, see also values in Table 2) and when one malfunction has happened (i.e., for $n_r=1$) this value changes to 3.765 (about 3 and 3/4 years).

3. Model of maintenance times

We assume that each connection in time *t* can be in one of the following states: *working, under repair, under replacement*. We also assume that immediately after a failure, the respective connection is repaired or replaced by a new one, so there is no waiting time for a necessary service.

As it was assumed, working times WT_i (i.e., times between malfunctions) are *iid* random variables described by the density (3). In the following, repairing times RT_i (times, when a connection is repaired) and replacement times PT_i (times, when a connection is replaced with a new one) are also modelled by a new kind of a probability distribution.

A replacement of a pipe is related to a deterministic and unconditional replacement age P^*. If the current sum of working and repairing times for the considered connection is larger than P^*, i.e.,

$$\sum_{i=1}^{j}WT_i+RT_i>P^*, \quad (6)$$

then this connection is replaced with a new one (instead of one more repair). Afterwards, $n_r=0$ is set, so this replacement "clears" a deterioration process of the given pipeline.

To model the mentioned repairing and replacement times we apply a special distribution related to a decreasing linear intensity function, i.e.,

$$\lambda(x)=-cx+cd \quad if \quad x\in[0,d],$$

where $c>0,d>0$, and *c* is its directional component while *d* is a right-hand side limit of its support. Then, using (2), the respective density is equal to:

$$f(x)=\frac{1}{1-\exp\left(-\frac{cd^2}{2}\right)}c(d-x)\exp\left(c\left(\frac{x^2}{2}-dx\right)\right) \quad if \quad x\in[0,d], (7)$$

An exemplary plot of this density can be seen in Fig. 3. To simulate random values from (7), the inversion method can be directly applied, so we get

$$F^{-1}(y)=d-\frac{1}{\sqrt{c}}\sqrt{cd^2+2\log\left(1+\left(\exp\left(-cd^2/2\right)-1\right)y\right)}. \quad (8)$$

Then, Algorithm 2 can be used to simulate values of the repairing and the replacement times. In the following, to distinguish parameters of these two types of services, we use c_R,d_R (in the case of the repairing times) or c_P,d_P (for the replacement times, respectively). The expected value of the density (7) can be numerically calculated, e.g., for $c=10,d=0.014$ we get 0.004666 (i.e., we have $\frac{5}{360}=0.013889,\frac{1}{360}=0.0027778$, so d is about 5 days for 360 days calendar, and the expected value is then about 1.68 day, see also values in Table 2).

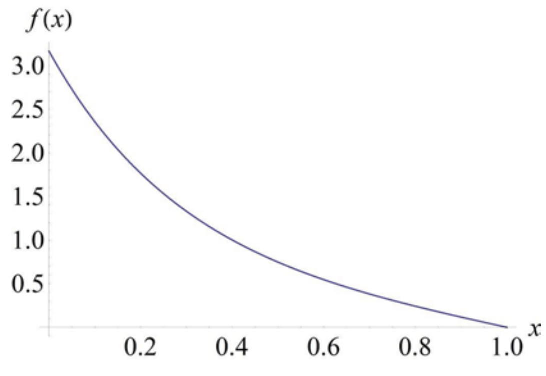


Fig. 3. Exemplary plot of the density for the repairing and the replacement times

Algorithm 2 (Generation procedure for the repairing / replacement times)

Input: Parameters c_R,d_R or c_P,d_P.
Output: A random repairing / replacement time *X*.

Generate independent random value U *from the uniform standard distribution* $U[0;1]$*;*

$X=F^{-1}(Y)$ *(see (7));*
return *X*

To simulate the whole WDS, we assume that its connections behave in a statistically independent way. Then, using the MC approach, the random times of malfunctions $t_1, t_2, \ldots$ (which are directly related to the working times), together with the repairing and the replacement times (i.e., RT_i and PT_i), can be generated.

4. Model of maintenance costs

In the following, the costs of repairs and replacements are estimated using the Monte Carlo (MC) simulations. The Monte Carlo approach is applicable if, apart from the considered HRF (1), some numerically feasible probability distributions for the repairing times RT_i and the replacement times PT_i are also used, like the introduced density (7). Because of the MC simulations, there is no need to solve complex theoretical formulas to take into account the condition (5).

In this paper, we focus only on the maintenance costs related to the replacements and the repairs. Other types of costs (like costs of water losses, loss of revenues, etc. – see, e.g., [4, 12, 19]) are commonly considered in the literature. Among others, restoration and diagnostic costs should be also mentioned. They are very important, especially for the long-time horizon of analysis (see, e.g., [26] for an additional discussion).

Moreover, we assume that a value of the costs is related to a type of service (i.e., if it is a replacement or a repair), length of this service and a type of the connection. Therefore we have:

$$cost^{(j)}(t_i) = cost^{(j)}_{R,const} + cost^{(j)}_{R,Var}\left(RT_{t_i}\right),$$

where $cost^{(j)}(t_i)$ denotes a *total sum of costs* for the given j-th connection and time t_i, when a necessary service begins, $cost^{(j)}_{R,const}$ is a constant value (or a fixed cost, i.e., the value which is independent of length of a repair) and $cost^{(j)}_{R,Var}(\,.\,)$ is a variable cost of a repair (i.e., the value which is related to length of a repair). In the same manner, if for a replacement, we have:

$$cost^{(j)}(t_i) = cost^{(j)}_{P,const} + cost^{(j)}_{P,Var}\left(PT_{t_i}\right).$$

In the existing literature, the concept of a variable interest rate, which is used to estimate the present value (or the future value, see, e.g., [28]) of the maintenance costs, is still rarely used. But, as it was pointed out in [25, 26], the obtained results in the case of a variable rate significantly differ if they are compared to models with a constant yield. It is especially true if a long time horizon T (like 20 or even 50 years, which are quite common values for real-life WDSs) is taken into account. Then, to calculate the present value of the total sum of the costs of repairs and replacements, which is given by:

$$PV(cost) = \sum_{i,j} PV\left(cost^{(j)}(t_i)\right),$$

the one-factor Vasicek model (see, e.g., [8, 28]) is used to find a discounting factor $PV(.)$ for each respective cost $c^{(j)}(t_i)$. For this variable interest rate, a value of the interest rate r_t at time t is modelled by:

$$dr_t = a\left(b - r_t\right) + \sigma\, dW_t\,, \tag{9}$$

where W_t is the standard Brownian motion, b characterizes a long term mean level (i.e., the trajectory of r_t is directed to b during its long run), a reflects the speed of reversion towards b, and σ is instantaneous volatility (variability) of the trajectory related to the random component W_t. In the MC setting, an iterative scheme for a generation of increments Δr_t of the process (9) should be used (see, e.g., [8, 26] for a more detailed discussion and necessary formulas).

5. Imprecise setting of the model

As it was mentioned in Sec. 1, there are a few important cases, when our model can be described in an imprecise way. For example, data can be sparse or even unavailable, so to take into account opinions of the experts, the necessary parameters of the model are given as imprecise values (like, e.g., "the value of this parameter is about 5" or "this parameter is relatively low", see, e.g., [11] for an additional discussion). This impreciseness can be modelled using various types of fuzzy sets (see, e.g., [9, 11, 21, 26, 27, 28] for additional discussion) like, e.g., triangular fuzzy numbers (abbreviated further as TRFN), left-right fuzzy numbers (LRFN), trapezoidal fuzzy numbers (TPFN), interval-valued fuzzy numbers (IVFN), or using another approach, like shadowed sets (SHS). Of course, fuzzy or shadowed sets, which are applied in the considered setting, should be strictly related to available data and its interpretation.

In this paper, we further develop our previous works related to the fuzzy approach to describe a WDS (see [25, 26]). In the following, the whole model of the maintenance costs, together with the parameters of the introduced HRF, will be entirely fuzzified. Moreover, we also present a completely new approach, which is based on shadowed sets. Respective numerical algorithms for both these cases will be also provided.

5.1. Fuzzy approach

We start with some basic definitions and notation, which will be used in this paper. Additional details concerning the fuzzy approach can be found in, e.g., [7, 14, 28].

For a fuzzy subset $\tilde{A}$ of the set of real numbers R, we denote by $\mu_{\tilde{A}}$ its membership function $\mu_{\tilde{A}} : R \to [0,1]$ and by $\tilde{A}[\alpha] = \{x : \mu_{\tilde{A}}(x) \geq \alpha\}$ the α-level set (or the α-cut) of $\tilde{A}$ for $\alpha \in (0,1]$. Then $\tilde{A}[0]$ is the closure of the set $\{x : \mu_{\tilde{A}}(x) > 0\}$.

A fuzzy number $\tilde{a}$ is a fuzzy subset of R for which $\mu_{\tilde{A}}$ is a normal, upper-semicontinuous, fuzzy convex function with a compact support. Then for each $\alpha \in [0,1]$, the α-level set $\tilde{a}[\alpha]$ is a closed interval of the form $\tilde{a}[\alpha] = \left[a_L[\alpha], a_U[\alpha]\right]$, where $a_L[\alpha], a_U[\alpha] \in R$ and $a_L[\alpha] \leq a_U[\alpha]$.

A left-right fuzzy number (LRFN) is a fuzzy number with the membership function of the form:

$$\mu_{\tilde{a}}(x) = \begin{cases} L\left(\dfrac{x-a}{b-a}\right), & x \in [a,b] \\ 1, & x \in [b,c] \\ R\left(\dfrac{d-x}{d-c}\right), & x \in [c,d] \\ 0, & otherwise \end{cases},$$

where $L, R : [0,1] \to [0,1]$ are non-decreasing functions, such that $L(0) = R(0) = 0$ and $L(1) = R(1) = 1$. If both L and R are strictly linear functions, then this kind of LRFN is known as a trapezoidal fuzzy number (further abbreviated as TPFN), and we have:

$$\mu_{\tilde{a}}(x)=\begin{cases}\frac{x-a}{b-a}, & x\in[a,b]\\ 1, & x\in[b,c]\\ \frac{d-x}{d-c}, & x\in[c,d]\\ 0, & otherwise\end{cases}.$$

This kind of fuzzy number will be further denoted by $[a,b,c,d]$. Trapezoidal fuzzy numbers are one of the most commonly used types of fuzzy numbers (see, e.g., [9] for additional discussion). A triangular fuzzy number (TRFN) is a trapezoidal fuzzy number with its core reduced to a single value. Then its membership function has the form:

$$\mu_{\tilde{a}}(x)=\begin{cases}\frac{x-a}{b-a}, & x\in[a,b]\\ \frac{x-c}{b-a}, & x\in[b,c]\\ 0, & otherwise\end{cases}.$$

TRFN will be further denoted by $[a,b,c]$. Some examples of the mentioned types of fuzzy numbers can be found in Fig. 4.

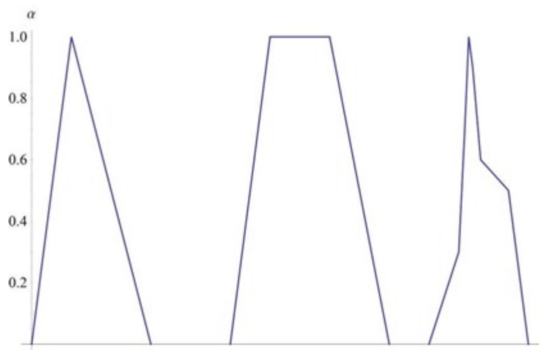


Fig.4. Examples of fuzzy numbers (the left-hand side to the right-hand side: a triangular fuzzy number, a trapezoidal fuzzy number, a left-right fuzzy number)

Let us assume, that our aim is to calculate the value of a function $f(x)$ for some parameter x, e.g., the present value of the maintenance costs $PV(c)$ depending on the constant cost of repair $c^{(j)}_{R,const}$. To approximate a fuzzy value $\tilde{f}(\tilde{x})$ for a fuzzy parameter $\tilde{x}$, two-step procedure (see Algorithm 3) is conducted. During the first step, monotonicity of $f(x)$ is checked and a specified value $\alpha\in[0,1]$ is set. If $f(x)$ is a non-decreasing function, then for the given α, the left endpoint $f_L(\tilde{x})[\alpha]$ of the respective α-level set $\tilde{f}(\tilde{x})[\alpha]$ is approximated using the crisp value $x_L[\alpha]$, i.e., $f_L(\tilde{x})[\alpha]=f(x_L[\alpha])$. And the same applies to the right endpoint $f_U(\tilde{x})[\alpha]$ and $x_U[\alpha]$, i.e., $f_U(\tilde{x})[\alpha]=f(x_U[\alpha])$. In contrary, if $f(x)$ is a non-increasing function, first $x_U[\alpha]$, then $x_L[\alpha]$ should be used to evaluate the respective α-cut , which is given by $[f_L(\tilde{x})[\alpha], f_U(\tilde{x})[\alpha]]$. Then the same step is repeated for other values of α. Usually, we start from $\alpha=0$ and end at $\alpha=1$ with some fixed increment $\Delta\alpha>0$. During the second step, the whole fuzzy number $\tilde{f}(\tilde{x})$ is constructed, based on the α-level sets which were previously calculated. The "missing" α-level sets are directly approximated using respective linear segments between the known α-cuts (see, e.g., [20, 26, 28] for further discussion). Of course, more complex functions (e.g., polynomials) can be also applied, but usually simple linear functions are sufficient. Let us illustrate the above procedure with a simple example:

Example. Let us suppose, that we are interested in an approximation of the maintenance costs $PV(c)$ for a fuzzy value of the constant cost of repair $\tilde{c}^{(j)}_{R,const}$, i.e., we would like to find $\widetilde{PV}(\tilde{c}^{(j)}_{R,const})$. As it is easily seen, the respective function is an increasing one in this case. Therefore, to find $\widetilde{PV}_L(\tilde{c}^{(j)}_{R,const})$ (or $\widetilde{PV}_U(\tilde{c}^{(j)}_{R,const})$, respectively) for the given α, the value $\tilde{c}^{(j)}_{R,const,L}$ (or $\tilde{c}^{(j)}_{R,const,U}$, respectively) should be applied. And we can start our approximation of the fuzzy output using $\alpha=0$ with an increment $\Delta\alpha=0.1$ up to $\alpha=1$.

Algorithm 3 (Approximation of the fuzzy output)

Input: A function $f(x)$, an increment $\Delta\alpha>0$.

Output: An approximation of $\tilde{f}(\tilde{x})$.

$\alpha=0$
while $\alpha\le 1$ **do**
  Check monotonicity of $f(x)$;
  Calculate $[f_L(\tilde{x})[\alpha], f_U(\tilde{x})[\alpha]]$ *using* $[f(x_L[\alpha]), f(x_U[\alpha])]$ *(if* $f(x)$ *is a non-decreasing function) or* $[f(x_U[\alpha]), f(x_L[\alpha])]$ *(otherwise);*
  $\alpha=\alpha+\Delta\alpha$;

Approximate missing values $[f_L(\tilde{x})[\alpha], f_U(\tilde{x})[\alpha]]$ *using linear segments;*

return $\tilde{f}(\tilde{x})$

Because in the following we focus on the present value of the maintenance costs and its fuzzy counterpart, monotonicity of $PV(c)$ depending on the considered parameters is summarized in Table 1, where a plus sign denotes non-decreasing and a minus sign – a non-increasing function of the given parameter. However, a similar table can be prepared for other kinds of the desired fuzzy output.

Table 1. Monotonicity of $PV(c)$ depending on different parameters of the model

Parameter	a_0	a_1	a_2	b_0	x_0^*	x_1^*	α_r	c_R, c_P	d_R, d_P	Variable and constant costs
Monotonicity	-	+	+	+	-	-	+	-	-	+

5.2. Approach based on shadowed sets

We start with some basic definitions and notation, which will be used in this paper further on. Additional details concerning shadowed sets can be found in, e.g., [10, 21].

A shadowed set (SHS, for short) $\dot{S}$ in a universe of discourse X is a set-valued mapping $\dot{S}: X \to \{0,[0,1],1\}$ having the following interpretation (see, e.g., [21]):

- all elements of X for which $\dot{S}(x)=1$ are called a *core* of the shadowed set $\dot{S}$ and they embrace all elements that are fully compatible with the concepts conveyed by $\dot{S}$,
- all elements of X for which $\dot{S}(x)=0$ are completely excluded from the concept described by $\dot{S}$,
- all elements of X for which $\dot{S}(x)=[0,1]$, called *a shadow*, are uncertain.

Then, for a shadowed set $\dot{S}$ we have its core (defined by $\{x \in X, \dot{S}(x)=1\}$), the shadow $\{x \in X, \dot{S}(x)=[0,1]\}$ and the support $cl\{x \in X, \dot{S}(x) \neq 0\}$. The usage of the unit interval for the shadow shows that any element from the shadow could be excluded or exhibit partial membership or could be fully allocated to $\dot{S}$ (see, e.g., [10]).

In the following, we consider shadowed sets defined for real values only, i.e. when $X=R$ (see Fig. 5 for the respective example). Then, a shadowed set will be denoted by $[s_1,s_2,s_3,s_4]_{SH}$, where its core is given by the interval $[s_2,s_3]$, its shadow by $(s_1,s_2)\cup(s_3,s_4)$ and its support by $[s_1,s_4]$. Shadowed sets are conceptually close to rough sets (see, e.g., [21]), i.e. their cores can be treated as the regions whose elements belong to the concept under discussion, their shadows – the regions, where the membership grade is doubtful, and the outer parts – the regions whose elements definitely do not belong to the concept.

Fig. 5. Example of a shadowed set

There exist important links between concepts of a fuzzy set and a shadowed set. First, a fuzzy set can be approximated using a shadowed set. Based on the initial fuzzy set, a corresponding shadowed set, that at the same time captures "the essence" of this fuzzy set, reduces computational efforts related to a membership function, and simplifies the interpretation, can be constructed. The idea behind this approach was discussed in [21]. As he noted "we are usually far more confident about assigning values close 1 (thus counting the elements in) or 0 (therefore making the corresponding element excluded from the concept). On the other hand, the membership values (such as those around 0.5) always spark some hesitation and are always more difficult to place on a simple numeric scale". Therefore, a shadowed set is constructed (induced) from the initial fuzzy set with an elevation of some membership values ("close to 1" or "high enough") and with a reduction of others (which are "close to 0" or "low enough"). Then, the necessary computational effort related to using the obtained shadowed set (instead of its fuzzy counterpart) is reduced, because only two "cuts" (instead of the whole [0,1] interval) are used in further calculations. Because this procedure can reduce vagueness, some additional restrictions are taken into account to maintain its overall value.

In [10], a respective approximation procedure, which is related to the optimization of two weighting functions, is introduced. Then, a TPFN, which is given by $[a,b,c,d]$, can be approximated by a SHS $[s_1,s_2,s_3,s_4]_{SH}$ using formulas

$$s_1 = \frac{2}{3}a + \frac{b}{3}, s_2 = \frac{a}{3} + \frac{2}{3}b, s_3 = \frac{2}{3}c + \frac{d}{3}, s_4 = \frac{c}{3} + \frac{2}{3}d \,. \qquad (10)$$

Example. Let us suppose, that our TPFN is described by values $[1,3,4,7]$. Then, from (10) we get $[1.6667, 2.3333, 5, 6]_{SH}$.

Second, calculation of the value of a function, which is related to a shadowed set, can be seen as a simplified approach for a fuzzy set, or – stated in another way – based on interval calculations. Namely, to find a value of a function $\dot{f}(\dot{S})$ for some shadowed set $\dot{S}$ we have to take into account only two intervals $[s_1,s_4]$ and $[s_2,s_3]$ similarly as for a fuzzy set (but with two α-level sets only, i.e. when α=0 and α=1). Then, Algorithm 3 can be directly modified if parameters of the considered model are given with shadowed sets instead of fuzzy sets. This leads us to Algorithm 4.

Algorithm 4 (Approximation of the shadowed set output)

Input: A function $f(x)$.

Output: An approximation of $\dot{f}(\dot{x})$.

Check monotonicity of $f(x)$*;*

Calculate the core of $\dot{f}$ *using* $[f(s_3), f(s_4)]$ *(if* $f(x)$ *is a non-decreasing function) or* $[f(s_4), f(s_3)]$ *(otherwise);*

Calculate the shadow of $\dot{f}$ *using* $(s_1,s_2)\cup(s_3,s_4)$ *(if* $f(x)$ *is a non-decreasing function) or* $(s_4,s_3)\cup(s_2,s_1)$ *(otherwise);*

return $\dot{f}(\dot{x})$

6. Examples of numerical simulations

After providing the necessary algorithms, we present some examples based on the Monte Carlo simulations. In the following, numerical approximations of the present value of the maintenance costs for an exemplary WDS for both the fuzzy and the shadowed sets settings are discussed. The applied parameters of the considered model can be divided into four groups:

1. parameters of the given type of the connection, which are related to the introduced HRF (1), namely $a_0, a_1, a_2, b_0, x_0^*, x_1^*, \alpha_r, n_r$,
2. parameters, which depend on the respective connection, and are related to the maintenance costs $cost_{R,const}, cost_{P,const}, cost_{R,var}(.), cost_{P,var}(.)$ or the lengths of times of necessary services (i.e. repairs and replacements), like parameters of the random distributions for RT_i and PT_i,
3. parameters of the interest rate model, which are related to the one-factor Vasicek model (6), i.e. r_0, a, b, σ,
4. other parameters, like P^* and time range for the whole simulation T.

6.1. Numerical analysis for the fuzzy setting

We assume that all the parameters related to the considered HRF (1), the costs of the repairs and the replacements, and the length of the services are fuzzified, i.e., they are given as triangular or trapezoidal fuzzy numbers. Only the parameters of the one-factor Vasicek model and the considered times (i.e. P^* and T) are described by crisp (real) values. Time is measured in years.

In the following, we partially use fuzzified versions of parameters considered in [26] (see Table 2). As noted in Sect. 2 and Sect. 3, the respective expected value for the period between malfunctions is "about 4.5 years" and for the time of repair is "about 1.68 days".

For example, if $a_0 = [0.5, 0.6, 0.65, 0.7]$, then the directional component of the descending linear part of (1) is described by a TPFN for which its core is given by the interval $[0.6, 0.65]$ and its support by $[0.5, 0.7]$. Therefore, we are completely sure, that the considered value is in the interval $[0.6, 0.65]$ and also "certain to some extent", that this value is in $[0.5, 0.7]$, so this parameter is "about 0.6.0–65 plus 0.05 /minus 0.1" (for additional remarks concerning the fuzzy scales and their interpretation, see, e.g., [15]). And for $cost_{R,const} = [0.5, 1, 2]$ we can say, that the constant costs of a single repair are "about 1, minus 0.5 (50% of the core value) / plus 1 (100% of the core value)", so this parameter has longer right-hand support. In the same manner, if $d_R = [0.012, 0.014, 0.016]$, then the maximum time of a repair is "about 5 days". The unconditional replacement age P^* is equal to 5 years, and the time range of the simulations is equal to 50 years, which is a value commonly used in real-life applications for other WDS.

In this example, we use only triangular or trapezoidal fuzzy values of the parameters, but more complex types of fuzzy numbers can be also applied (like, e.g., LRFNs). However, these two types are the most commonly spotted in real-life applications because of their simple description, together with easy and direct interpretation.

To find fuzzy approximations of the desired output, one million Monte Carlo simulations for 10 pipes (which are identical, i.e. they have the same parameters) were conducted with the increment $\Delta\alpha = 0.1$. Using the rather low-end hardware (i5-7400 3 GHz, 8 GB RAM, Win 7 Pro) as for the modern standards and C++ (Visual Studio 2019), the whole simulation procedure took about 1 hour.

Some examples of the obtained output were summarized in Fig. 6–9. In Fig. 6 we can find fuzzy approximations of the minimum, the mean and the maximum of the costs of the single repair. As we can see, these values are described by a TRFN (the minimum) or TPFNs (the mean and the maximum), but the mean (given by $[0.766419, 1.32643, 1.37307, 2.39986]$) has the rather narrow core, mainly due to the existing discounting factor. Similar measures for the costs of the single replacement are given in Fig. 7. Moreover, fuzzy approximations of the means for the costs of the repairs and the replacements differ substantially, the first one is lower than the second (which is equal to $[4.79961, 6.83968, 6.953, 8.95973]$), and they have other kinds of skewness. Their relatively wide supports should be also noticed.

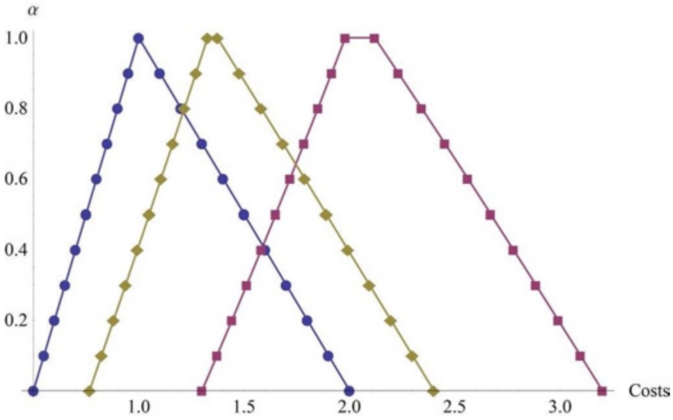


Fig. 6. Fuzzy approximations of the minimum (circles), the mean (diamonds), and the maximum (rectangles) of the repairs costs

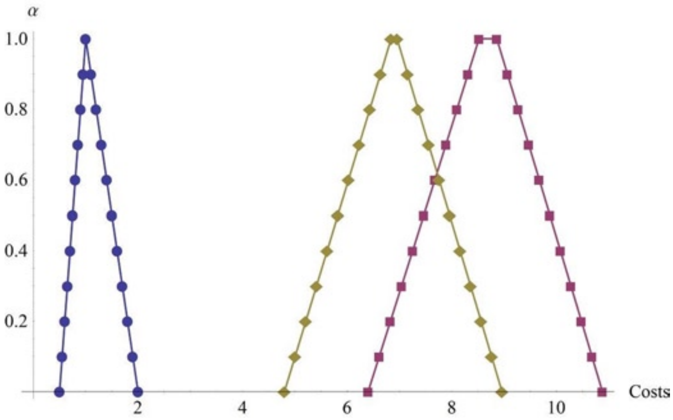


Fig. 7. Fuzzy approximations of the minimum (circles), the mean (diamonds), and the maximum (rectangles) of the replacements costs

Table 2. Fuzzy and crisp parameters applied in exemplary numerical simulations

Parameter	a_0	a_1	a_2	b_0
Value	$[0.5, 0.6, 0.65, 0.7]$	$[0.1, 0.2, 0.3]$	$[0.7, 0.8, 0.9, 1]$	$[0.6, 0.65, 0.7]$

Parameter	α_r	x_0^*	x_1^*
Value	$[0.05, 0.1, 0.15]$	$[0.4, 0.5, 0.6, 0.7]$	$[9, 10, 11, 12]$

Parameter	c_R	d_R	c_P	d_P
Value	$[8, 10, 11, 12]$	$[0.012, 0.014, 0.016]$	$[4, 5, 6]$	$[0.024, 0.026, 0.028, 0.03]$

Parameter	$cost_{R,const}$	$cost_{R,var}$	$cost_{P,const}$	$cost_{P,var}$
Value	$[0.5, 1, 2]$	$[50, 70, 80, 100]$	$[4, 6, 8]$	$[80, 90, 110, 120]$

Parameter	a	b	r_0	σ	P^*	T
Value	0.1	0.05	0.04	0.001	5	50

Another output is shown in Fig. 8, where the fuzzy approximation of the average number of the repairs $\bar{x}_R$ of the whole WDS is plotted. The obtained LRFN is very close to a TPFN, but clearly differs

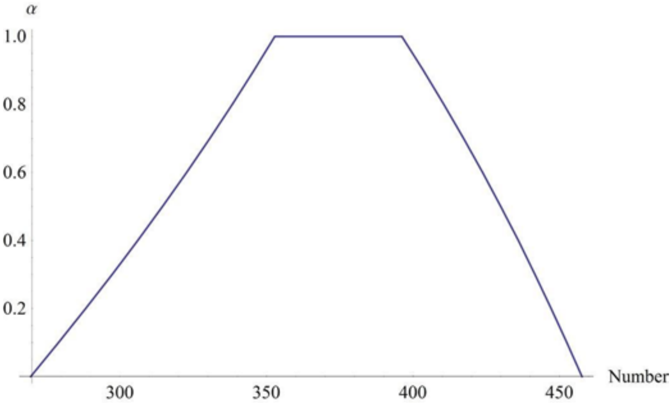
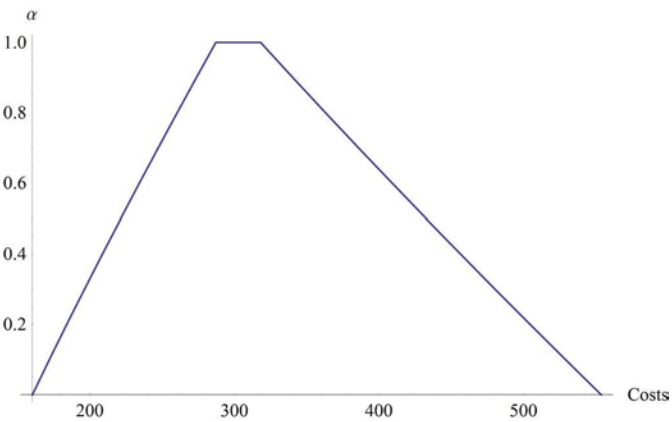


Fig. 8. Fuzzy approximation of the average number of repairs $\bar{x}_R$

from this kind of fuzzy number because of the visible curvatures of its membership function for the left- and the right-hand sides. Moreover, its support, which is equal to [154.314 , 349.395], is rather wide, then the obtained impreciseness is rather high in this case.

The most important simulation result, i.e. fuzzy approximation of the present value of the whole maintenance costs $PV(cost)$, is presented in Fig. 9. In this case, we have a fuzzy number that may be identified with a TPFN, because no curvatures are clearly visible. The obtained fuzzy number has the longer right-hand support (so we can expect some costs “on plus” rather than “on minus”), with the core given by the interval [359.268 , 384.822] and the rather wide support, which is equal to [209.44 , 621.271]. Once more, the obtained impreciseness for $\alpha = 0$ is rather high. Clearly, for the given other value of *α*, we can find the respective *α*-level set of $PV(cost)$, which approximates the result as the interval.

Fig. 9. Fuzzy approximation of the present value of the maintenance costs $PV(cost)$



It may be fruitful to compare the simulation results if one (or more) of the important parameters are changed. The whole optimization procedure can be even applied to minimize the whole maintenance costs (see, e.g., [26]). In the following, instead of $P^* = 5$, higher value of the unconditional replacement age, which is equal to 20 years (i.e. $P^* = 20$) is assumed. Some of the obtained results can be found in Fig. 10–11.

As we can see in Fig. 11, for the longer unconditional replacement age, the present value of the maintenance costs is highly reduced (to a TPFN given by $[72.4155,120.987,134.428,222.912]$, so even more than 65% for some α-cuts). Moreover, the average number of repairs

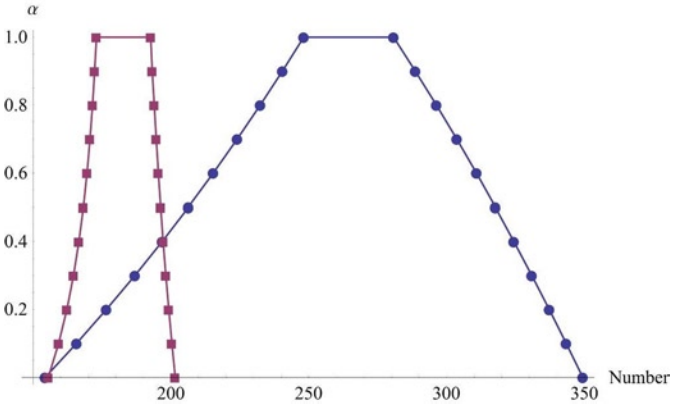


Fig. 10. Fuzzy approximation of the average number of repairs $\bar{x}_R$ for $P^* = 5$ (years, circles) and $P^* = 20$ (years, rectangles)

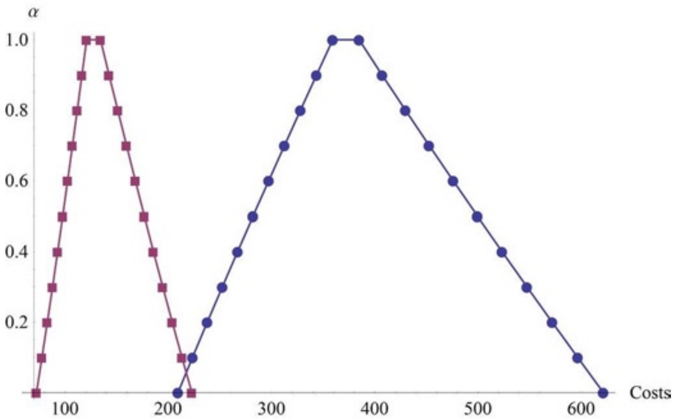


Fig. 11. Fuzzy approximation of the present value of the maintenance costs $PV(cost)$ for $P^* = 5$ (years, circles) and $P^* = 20$ (years, rectangles)

(see Fig. 10) is also significantly lower, with the highly reduced length of its support (hence, its impreciseness, too).

6.2. Numerical analysis in the case of the shadowed sets

As it was noticed in Sec. 1, it may be fruitful to use a concept of shadowed sets instead of fuzzy numbers. First, it may be easier for an expert to describe a considered parameter as a set of four values only, rather than formulate a more complex opinion concerning a whole course of a membership function. Second, calculations related to shadowed sets are usually easier, if they are compared with numerical efforts required for fuzzy numbers, because only two levels are necessary to obtain the desired output instead of, e.g., twenty α-level sets. Hence, only two simulation runs are required instead of twenty of them. But the obtained shadowed set still reflects some vagueness, which may be expressed as an imprecise opinion of an expert.

In the following, to compare the simulated results for both approaches, we approximate all fuzzy parameters (see Table 2) using shadowed sets and the formulas (10), e.g., we have:

$$a_0 = \left[0.5(3), 0.5(6), 0.6(6), 06.8(3)\right]_{SH},$$
$$a_1 = \left[0.1(3), 0.1(6), 0.2(3), 0.2(6)\right]_{SH},$$
$$a_2 = \left[0.7(3), 0.7(6), 0.9(3), 0.9(6)\right]_{SH},$$
$$b_0 = \left[0.61(6), 0.7(6), 0.9(3), 0.9(6)\right]_{SH}.$$

Values of other parameters can be easily computed using (10).

The simulated mean of the costs for single repair is plotted in Fig. 12, the mean of the costs for a single replacement in Fig. 13, the average number of repairs $\bar{x}_R$ in Fig. 14 and the present value of the maintenance costs $PV(c)$ in Fig. 15. If they are compared with their fuzzy counterparts, wider cores and narrower shadows are clearly visible. Therefore, the simulated shadowed sets are different, but still similar to the previously obtained fuzzy sets, e.g., now the mean of the single repair costs is equal to $[0.956047, 1.14272, 1.71829, 2.06055]_{SH}$. Generally, the areas of values, which are "completely sure", are wider and the areas, which are "doubtful", are narrower. But their "shapes" (i.e. when the left- or the right-hand shadow is wider in comparison with its right- or left-hand counterpart) are similar to the respective fuzzy outputs. It is especially clearly visible in the case of $PV(c)$ (compare Fig. 9 with Fig. 15).

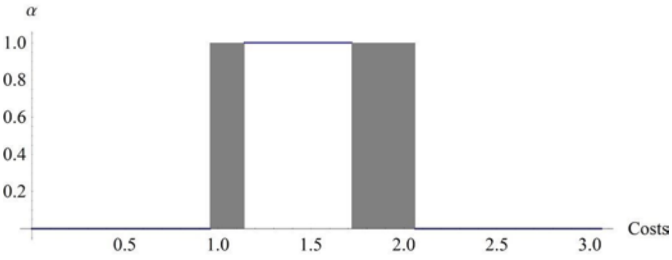


Fig. 12. Shadowed set approximation of the mean of the repairs costs

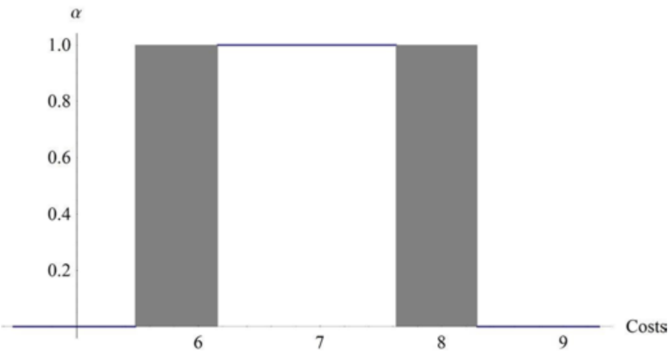


Fig. 13. Shadowed set approximation of the mean of the replacements costs

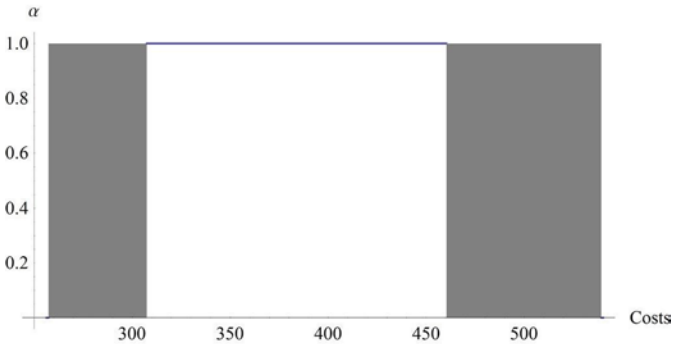


Fig. 14. Shadowed set approximation of the average number of repairs $\overline{x}_R$.

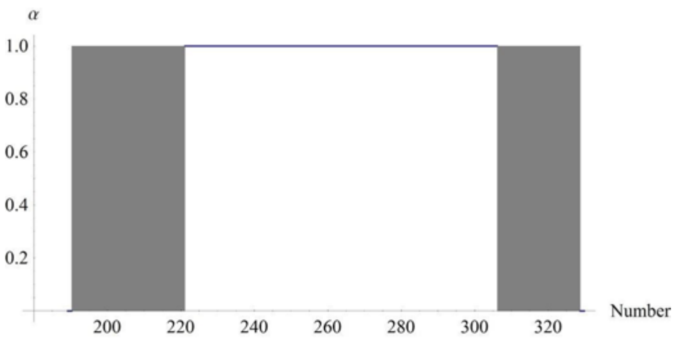


Fig. 15. Shadowed set approximation of the present value of the maintenance costs $PV(cost)$.

7. Conclusion

In this paper, a new kind of a hazard rate function for the time between malfunctions of a pipeline is proposed. This HRF is a U-shaped function, which also depends on the number of the previous repairs of the given connection. The numerically efficient simulation algorithm for this HRF is also provided. Additionally, times of the maintenance services (i.e. repairs and replacements) are modelled using the random distribution with decreasing intensity and finite support. The introduced models are then used to approximate the present value of the maintenance costs and other important characteristics for a WDS in two imprecise settings based on fuzzy numbers and shadowed sets. The general framework for these two imprecise setups together with the respective numerical algorithms are also discussed. These two approaches lead us to better incorporation of the experts' knowledge and a more proper, closer to real-life modelling of imprecise parameters of the considered model. To calculate the present value of the maintenance costs, the one-factor Vasicek model is used, as it doesn't incorporate error appearing in the case of a constant interest rate. The numerical examples of the maintenance costs and other important characteristics for a WDS are also analysed.

References

1. Amani N, Ali N M, Mohammed A H, Samat R A. Maintenance and management of wastewater system components using the condition index system, prediction process and costs estimation. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2013; 15(2): 161-168.
2. Blokus A, Kołowrocki K. Reliability and maintenance strategy for systems with aging-dependent components. Quality and Reliability Engineering International 2019, https://doi.org/10.1002/qre.2552.
3. Dąbrowska E, Soszyńska-Budny J. Monte Carlo Simulation Forecasting of Maritime Ferry Safety and Resilience, 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Bangkok 2018, 376-380, https://doi.org/10.1109/IEEM.2018.8607464.
4. Fadaee M J, Tabatabaei R. Estimation of failure probability in water pipes network using statistical model. Engineering Failure Analysis 2011; 18: 1184-1192, https://doi.org/10.1016/j.engfailanal.2011.02.013.
5. Farmani R, Kakoudakis K, Behzadian K, Butler D. Pipe Failure Prediction in Water Distribution Systems Considering Static and Dynamic Factors. Procedia Engineering 2017, 186: 117-126, https://doi.org/10.1016/j.proeng.2017.03.217.
6. Folkman S. Water Main Break Rates In the USA and Canada: A Comprehensive Study. Utah State University 2018.
7. Gil M A, Hryniewicz O. Statistics with Imprecise Data. In: Meyers R A (ed.). Encyclopedia of Complexity and Systems Science. New York: Springer-Verlag, 2009.
8. Glasserman P. Monte Carlo Methods in Financial Engineering. New York: Springer, 2004, https://doi.org/10.1007/978-0-387-21617-1.
9. Grzegorzewski P. Trapezoidal approximation of fuzzy numbers based on sample data. In: Hüllermeier E, Kruse R, Hoffmann F (eds.). Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications. Springer Berlin Heidelberg 2010:

402-411, https://doi.org/10.1007/978-3-642-14058-7_42.
10. Grzegorzewski P. Fuzzy number approximation via shadowed sets. Information Sciences 2013; 225: 35-46, https://doi.org/10.1016/j.ins.2012.10.028.
11. Hryniewicz O, Kaczmarek K, Nowak P. Bayes statistical decisions with random fuzzy data - An application for the Weibull distribution. Eksploatacja i Niezawodnosc - Maintenance and Reliability. 2015; 17(4): 610-616. https://doi.org/10.17531/ein.2015.4.18
12. Kleiner Y, Adams B J, Rogers J S. Long-term planning methodology for water distribution system rehabilitation. Water Resources Research 1998; 34(8): 2039-2051, https://doi.org/10.1029/98WR00377.
13. Kumar G, Jain V, Gandhi O P. Steady-state availability analysis of repairable mechanical systems with opportunistic maintenance by using semi-Markov process. Int J Syst Assur Eng Manag. 2014; 5(4): 664-678, https://doi.org/10.1007/s13198-014-0231-8.
14. Lee K H. First Course on Fuzzy Theory and Applications. Berlin Heidelberg: Springer, 2005.
15. Lubiano M A, Salas A, Carleos C, de la Rosa de Sáa S, Gil M Á. Hypothesis testing-based comparative analysis between rating scales for intrinsically imprecise data. International Journal of Approximate Reasoning 2017; 88: 128-147, https://doi.org/10.1016/j.ijar.2017.05.007.
16. Liu Z, Kleiner Y, Rajani B, Wang L, Condit W. Condition Assessment Technologies for Water Transmission and Distribution Systems. Washington DC: U.S. Environmental Protection Agency, 2012; EPA/600/R-12/017.
17. Malinowski J. A newly developed method for computing reliability measures in a water supply network. Operations Research and Decisions 2016; 26(4): 49-64.
18. Marchionni V, Cabrala M, Amadoa C, Covasa D. Water supply infrastructure cost modelling. Procedia Engineering 2015; 119: 168-173, https://doi.org/10.1016/j.proeng.2015.08.868.
19. Neelakantan T R, Suribabu C R, Lingireddy S. Optimisation procedure for pipe-sizing with break-repair and replacement economics. Water SAl 2008; 34(2): 217-224, https://doi.org/10.4314/wsa.v34i2.183642.
20. Nguyen H T. A Note on the Extension Principle for Fuzzy Sets. Journal Mathematical Analysis and Applications 1978; 64: 369-380, https://doi.org/10.1016/0022-247X(78)90045-8.
21. Pedrycz W. Shadowed Sets: Representing and Processing Fuzzy Sets. IEEE Transactions on Systems, Man and Cybernetics- Part B 1998; 28(1): 103-109, https://doi.org/10.1109/3477.658584.
22. Pietrucha-Urbanik K, Studziński A. Case study of failure simulation of pipelines conducted in chosen water supply system. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19 (3): 317-322, https://doi.org/10.17531/ein.2017.3.1.
23. Rojek I, Studziński J. Comparison of different types of neuronal nets for failures location within water-supply networks. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2014; 16 (1): 42-47.
24. Rojek I, Studziński J. Detection and localization of water leaks in water nets supported by an ICT system with artificial intelligence methods as away forward for smart cities. Sustainability 2019; 11(2), https://doi.org/10.3390/su11020518.
25. Romaniuk M. On simulation of maintenance costs for water distribution system with fuzzy parameters. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18(4): 514-527, https://doi.org/10.17531/ein.2016.4.6.
26. Romaniuk M. Optimization of maintenance costs of a pipeline for a V-shaped hazard rate of malfunction intensities. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20(1): 46-54, https://doi.org/10.17531/ein.2018.1.7.
27. Romaniuk M, Hryniewicz O. Discrete and smoothed resampling methods for interval-valued fuzzy numbers. IEEE Transactions on Fuzzy Systems 2020, https://doi.org/10.1109/TFUZZ.2019.2957253.
28. Romaniuk M, Nowak P. Monte Carlo methods: theory, algorithms and applications to selected financial problems. Warszawa: ICS PAS, 2015.
29. Røstum J. Statistical modelling of pipe failures in water networks. Ph.D. dissertation. 2000.
30. Scheidegger A, Leitão J P, Scholten L. Statistical failure models for water distribution pipes - A review from a unified perspective. Water Research 2015; 83: 237-247, https://doi.org/10.1016/j.watres.2015.06.027.

Maciej ROMANIUK
Systems Research Institute Polish Academy of Sciences
ul. Newelska 6, 01-447 Warszawa, Poland

Warsaw School of Information Technology,
ul. Newelska 6, 01-447 Warsaw, Poland

Olgierd HRYNIEWICZ
Systems Research Institute Polish Academy of Sciences
ul. Newelska 6, 01-447 Warszawa, Poland

E-mails: mroman@ibspan.waw.pl hryniewi@ibspan.waw.pl

Article citation info:
KORTA J, SKRUCH P. The concept of integrating vapor chamber into a housing of electronic devices for increased thermal reliability. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 363–369, http://dx.doi.org/10.17531/ein.2020.2.19.

Jakub KORTA
Paweł SKRUCH

THE CONCEPT OF INTEGRATING VAPOR CHAMBER INTO A HOUSING OF ELECTRONIC DEVICES FOR INCREASED THERMAL RELIABILITY

KONCEPCJA INTEGRACJI KOMORY PAROWEJ Z OBUDOWĄ UKŁADÓW ELEKTRONICZNYCH W CELU ZWIĘKSZENIA ICH NIEZAWODNOŚCI*

Systematic increase in computational power and continuous miniaturization of automotive electronic controllers pose a challenge to maintaining allowable temperature of semiconductor components, preventing premature wear-out or, in extreme cases, unacceptable shutdown of these devices. For these reasons, efficient and durable cooling systems are gaining importance in modern car technology design, showing critical influence on reliability of vehicle electronics. Vapor chambers (flat heat pipes) which could support heat management of automotive electronic controllers in the nearest future are passive devices, which transport heat through evaporation-condensation process of a working liquid. At present, vapor chambers are not commercially used in cooling systems of automotive controllers, being a subject of research and development endeavors aimed at understanding their influence on thermomechanical reliability of semiconductor devices used in cars. This paper presents a concept of an electronic controller aluminum housing integrated with a vapor chamber. The conceptual design was numerically validated in elevated temperature, typical for automotive ambient conditions. The paper discusses influence of the vapor chamber-based cooling system on the controller's thermal performance, as well as on its reliability, expressed as the expected lifetime of the device.

Keywords: *cooling, electronics, thermal analysis, heat pipe, vapor chamber, phase change, automotive, reliability.*

Systematycznie wzrastająca moc obliczeniowa oraz postępująca miniaturyzacja urządzeń elektronicznych stosowanych w pojazdach samochodowych powodują trudności w utrzymaniu temperatury pracy elementów półprzewodnikowych w dozwolonym zakresie, przyczyniając się do ich przedwczesnego zużycia, a w skrajnych przypadkach, uniemożliwiając nawet ich normalną pracę. Wydajne i trwałe układy chłodzące stają się więc nieodzownym komponentem współczesnych podzespołów samochodowych, o krytycznym znaczeniu dla ich niezawodności. Urządzeniami mogącymi w niedalekiej przyszłości wspomagać działanie układów chłodzenia systemów elektronicznych wykorzystywanych w motoryzacji są komory parowe (płaskie rurki cieplne), w których transport energii termicznej zachodzi poprzez przemianę fazową i samoistne przemieszczanie się czynnika roboczego. Współcześnie, tego rodzaju urządzenia nie są komercyjnie stosowane w układach chłodzenia sterowników samochodowych, pozostając przedmiotem prac badawczo-rozwojowych związanych z ich wpływem na szeroko pojętą niezawodność termomechaniczną urządzeń elektronicznych. W niniejszym artykule opisano koncepcję zintegrowania komory parowej z aluminiową obudową kontrolera elektronicznego pracującego w warunkach podwyższonej temperatury otoczenia, odpowiadającej warunkom użytkowania komponentów samochodowych. Ponadto, ocenie poddano wpływ zastosowania tego urządzenia na temperaturę pracy chłodzonego elementu półprzewodnikowego i jego niezawodność, wyrażoną jako przewidywany czas jego bezawaryjnego funkcjonowania.

Słowa kluczowe: *chłodzenie, elektronika, analiza termiczna, rurka cieplna, komora parowa, przemiana fazowa, motoryzacja, niezawodność.*

1. The need for cooling of electronic devices

In the recent years, high-power electronic devices have been gaining more and more interest in the automotive industry. This type of devices is used to increase vehicle comfort and safety performances. High-end automotive electronics is also utilized in multimedia equipment (infotainment) and various types of driver assistance systems, collectively called as ADAS (Advanced Driver Assistance Systems). Operation principle of ADAS devices is based on acquisition of various types of signals – generated by wireless communication systems, cameras, radars, ultrasonic transducers or lidars – and its high-performance processing. For safety reasons, the latter is frequently performed in near real time.

Acquisition and analysis of such big data streams requires utilization of significant computing powers, which results in an undesirable side effect of generation of large amounts of thermal energy. In consequence, this often leads to considerable rise in operating temperature of electronic components, resulting in premature wear-out or, in extreme cases, causing sudden shutdown of these components during operation, which is unacceptable for human safety reasons. Another important factor is the ambient temperature to which the ADAS systems are exposed. Automotive-grade components must typically operate in 50°C or more, which makes it difficult to maintain semiconductor chips within allowable operating temperature limits. Following the results presented in [20], high operating temperature was the reason for 49% of recalls in the automotive market, in the 2005-2015 period. As described in [6,16], alongside mechanical vibration, long-term high temperature exposure and temperature cycling are the main factors of electronics failure.

It is forecasted, that despite continuous development and optimization of computing algorithms and improvements of hardware tech-

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

nology, the value of heat fluxes generated by complete automotive electronic systems (in particular automated driving controllers) will rise to 1kW and above [8]. In order to make operation of these systems possible, new cooling solutions for vehicle electronics are being developed, which must meet the automotive requirements, that is, be thermally efficient, characterized by high degree of reliability and be low-cost, allowing for large scale production. Such cooling systems must allow for dissipation of high density heat fluxes, being a result of generating several dozen of watts on a surface of a single semiconductor device die (typically, several square centimeters). An example of such a device is Nvidia Xavier system-on-a-chip, dedicated to autonomous driving tasks, for which the TDP (thermal design power) is 30W, generated on a die surface area equal to 3.5cm^2.

The undesired influence of elevated temperature on electronic devices is broadly described in the scientific and engineering literature. The authors of [10] described an overview of temperature-driven failure mechanisms of semiconductor devices and the influence of these phenomena on device reliability. Temperature cycling effect on reliability of high-power automotive transistors is described in [18]. In [11], the authors discussed an attempt to assess fatigue resistance of a car electronic controller components, loaded by variable operating temperature. A novel methodology of conducting accelerated temperature aging tests is proposed in [19]. The authors of [22] presented a methodology of predicting electronic devices reliability, which is based on the so-called PoF (physics of failure) model, incorporating the influence of die temperature on thermomechanical behavior of semiconductor device. Analysis of temperature effect on expected lifetime duration of contemporary ADAS electronic devices and systems is described in [4]. The same paper provides an insight into a methodology allowing for diagnosis of this sort of problems in the early stage design phase. A detailed description of up-to-date analytical and experimental methods for assessment of electronic devices reliability is presented in [5]. Two-phase cooling solutions, based on liquid evaporation – condensation cycle, and its application to electronics heat management is described in [1] and [3]. Both of these papers describe active cooling systems which were used to dissipate thermal energy to the ambient. Despite high cooling capacity, this type of solutions are rarely used in the automotive sector, due to the need for costly and space-occupying auxiliary devices, that is, heat exchangers, pumps, valves, hoses and others. The authors of [13] described phase change phenomenon and its exploitation in a passive cooling device – VC (vapor chamber), building and validation of numerical model describing VC behavior and its comparison to the efficiency of copper heat spreader. Similar study is presented in [17], which provides thorough characterization of vapor chambers and copper heat spreaders, used to disperse heat from semiconductor devices. In both, [13] and [17], the two-phase devices have been identified as superior over standard copper plates.

Despite great popularity of two-phase cooling devices in consumer electronics, there is a limited number of reports available in the public domain, describing its application in the automotive industry. The author of [12] provides a general overview of possible implementations of VC and heat pipes in road vehicles. The following regions of possible applications are enlisted in this work: temperature control of LED headlamps, electric vehicle battery pack cooling systems and heat management of electronic control units. An attempt of using heat pipes to dissipate thermal energy from a motorcycle headlamp equipped with LEDs is described in [15]. By comparison to the classical solution, the proposed two-phase cooling provided small improvement to the estimated lifetime duration of the headlamp. Application of VC to dissipate heat from a high power transistor module is studied in [14]. According to the findings presented in this paper, the use of VC resulted in decreased operating temperature of the transistors by ca. 9°C compared to a standard, copper-based solution. Simultaneously, application of VC decreased the temperature gradient measured on the analyzed device, positively influencing mechanical reliability of the module. Automotive radio passive cooling system based on a heat pipe is described in [9]. The device was tested in room conditions, in natural convection cooling. With respect to the solution based on an aluminum radiator, the two-phase cooling system provided ca. 3°C reduction in temperature of the main heat source component.

This paper describes a concept of integrating a small VC into an aluminum housing of an electronic device operating in ambient temperature typical for automotive environment. The aim of the presented analysis was to understand the influence of using flat heat pipe on the junction temperature (t_j) of the main heat source device and to estimate the change in the expected failure-free operating lifetime of this component. Considering limited number of studies available in the public domain and dedicated to passive, two-phase cooling systems of vehicle controllers, this paper can be a valuable source of information for scientists and engineers working in the field of thermomechanical reliability of automotive-grade electronic devices.

Section 2 of the paper presents general classification of cooling methods used to control temperature of electronic devices and discusses physical mechanisms of heat transfer. Section 3 describes operating principles of two-phase cooling devices. The concept of VC integration with an aluminum housing of an electronic device is describes in Section 4. Section 5 discusses the effect of this modification to the expected operation lifetime duration of the device. Section 6 summarizes and concludes the paper.

2. Cooling methods of electronics

Cooling systems of electronics can be divided into two categories: passive and active systems. Cooling solutions belonging to the first group take advantage of natural mechanisms of heat exchange, that is, conduction, convection and radiation. Active cooling systems utilize auxiliary, externally powered devices, like fans or pumps, in order to enhance natural heat transfer processes.

Other concept of cooling technologies classification assumes a straightforward criterion of occurrence of phase change process: electronics cooling can be obtained by heat transfer with no phase change effect (e.g., natural or forced air convection) or by employing phase change cycles (e.g., the already mentioned evaporation – condensation process, which takes place in VCs). Selection of the appropriate cooling technique is driven by the values of heat flux and heat flux density, which must be removed from a device. In the automotive applications, passive cooling is preferred, because of maintenance-free character of this technology. Moreover, by contrast to the active solutions, passive cooling does not require external power supply, which improves the overall energy balance of the system.

Heat transfer mechanisms, that is, conduction, convection and radiation can be described by equations (1) – (3) [2]. In the case of heat conduction along a specified direction, the heat flow rate can be expressed by the equation (1):

$$\dot{Q} = -k\frac{\Delta TA}{\Delta x} \tag{1}$$

where: $\dot{Q}\left[\mathrm{W}\right]$ – heat flow ratio, $\Delta T\left[\mathrm{K}\right]$ – temperature difference measured along the heat flow, $\Delta x\left[\mathrm{m}\right]$ – flow distance, $A\left[\mathrm{m}^2\right]$ – cross-section area of the conductor, $k\left[\frac{\mathrm{W}}{\mathrm{m}\cdot\mathrm{K}}\right]$ – heat conductivity of the conductor material.

Heat convection is a macroscopic-scale phenomenon of heat transport, which takes place in fluids. Natural convection is driven by non-uniform fluid density, while forced convection is caused by external means, for example, operation of a fan or pump. Convective heat flow ratio can be described by equation (2):

$$\dot{Q} = Ah(T_s - T_\infty) \quad (2)$$

where: $T_s[\mathrm{K}]$ and $T_\infty[\mathrm{K}]$ – temperature measured on a heat source and at a distance, respectively, $h\left[\frac{\mathrm{W}}{\mathrm{m^2K}}\right]$ denotes convection heat transfer coefficient, which depends on shape, size, convection type and other factors.

Radiative heat transfer is driven by propagation of electromagnetic waves which are generated by movement of electrically charged particles. Similarly to convection and conduction, thermal radiation takes place in ambient temperatures above absolute zero. Heat radiated by a source in unit of time can be expressed by equation (3):

$$\dot{Q} = \varepsilon\sigma A\left(T_S^4 - T_A^4\right) \quad (3)$$

where: ε – dimensionless coefficient of thermal emissivity, $\sigma = 5.67x10^{-8}\left[\frac{\mathrm{W}}{\mathrm{K^4m^2}}\right]$ –Stefan – Boltzmann constant, $T_A[\mathrm{K}]$ – ambient temperature.

In reality, all three phenomena simultaneously participate in heat dissipation process, but it is a cooling system designer responsibility to decide on individual contribution of each heat transfer mechanism.

3. Phase change heat transfer

Efficient passive cooling systems are based on phase change heat transfer mechanism. In such solutions, the phase change is most frequently observed between liquid and gaseous phases, in the evaporation – condensation process. In cooling solutions of electronics, the phase change process is most commonly used in heat pipes and VCs. As schematically shown in Fig. 1, functionally separate regions can be distinguished in the body of these devices.

In the evaporator region, which is heated by an external heat source, working fluid changes its phase to vapor. The amount of thermal energy consumed in this process is equal to the liquid's latent heat of evaporation. The accumulated thermal energy is stored in the vapor and transported through the device cavity by natural convection process, towards the condenser region. Lower temperature of the condenser initializes heat transfer out of the vapor, causing reversed phase change process, that is, condensation. Simultaneously, pressure field inside the device causes condensed liquid to flow towards evaporator region. Liquid transport is enhanced by capillary forces between working medium and grooves or porous structures frequently embedded inside the device.

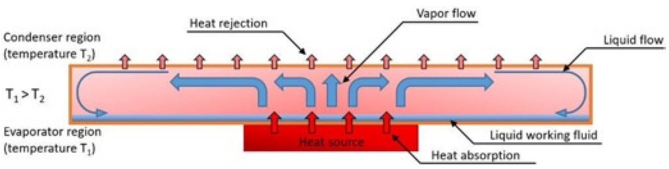


Fig. 1. Vapor chamber operating principle

In VCs integrated in consumer electronic devices, the most frequently used working medium is demineralized water. Other liquids, like methanol or acetone, are less popular due to higher level of toxicity and lower latent heat of evaporation h_e values they are characterized by. Table 1 combines h_e values for the mentioned working fluids.

The overall effect of using VC is continuous absorption of heat in the evaporator region, and its release in condenser area. Hence, thermal energy is continuously transported from source to opposite side of the device, where it can be easily dissipated to the ambient. Additionally, typically large VC condenser regions lead to a decrease in density of dissipated heat flux.

Table 1. Latent heat of evaporation for selected working fluids typically used in two-phase cooling [2].

Working fluid	Latent heat of evaporation $h_e\left[\frac{\mathrm{kJ}}{\mathrm{kg}}\right]$
Water	2257
Methanol	1100
Acetone	518

Subsequent paragraphs of this paper describe a concept of integrating a VC into aluminum housing of an electronic device. Thermal efficiency of this passive cooling system is compared with a standard solution, based on heat conduction in solid material of the enclosure. Assessment of both designs was done by means of numerical modelling and analysis by the CFD (Computational Fluid Dynamics) method.

4. Concept of vapor chamber integration into an electronics housing

Figure 2 depicts the two-part enclosure of the electronic device which was the subject of the described analyses. The housing was made from aluminum alloy EN AC-44300 ($k_{Al} = 130\frac{\mathrm{W}}{\mathrm{m \cdot K}}$) and covered by a number of fins, allowing for improved heat exchange with the surroundings. Additionally, Fig. 2 shows the header used to supply power and communicate with electronic devices inside the controller.

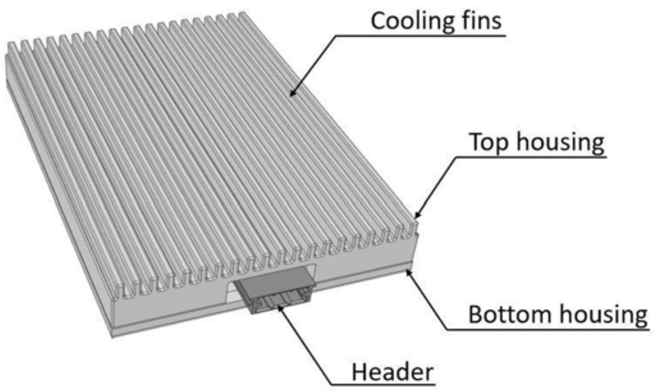


Fig. 2. The analyzed electronic device enclosed in an aluminum housing

The general view of the electronic device layout (i.e., PCB with electronic components) used in thermal analysis is shown in Fig. 3. Additionally, the analyzed model contained screws used for closing the assembly. One of the components (Component I) mounted on the PCB dissipated $P = 18\,\mathrm{W}$ of power, while the total power dissipation of the complete device reached 20.84W (Fig. 4). In the standard cooling solution which was analyzed in the first design iteration, Component I was connected to the aluminum housing through a pedestal. Additionally, a layer of thermal interface material was used between the heat source and aluminum pedestal, to decrease thermal interface resistance between these components. The model was simulated in elevated ambient temperature conditions, for which $t_{amb} = 55°\mathrm{C}$. The highest allowable junction temperature for Component I device was given by the chip manufacturer and was equal to $t_j^{max} = 125°\mathrm{C}$. Estimation of this value can be done by using equation (4), which requires knowledge of maximum temperature on the top surface of the device t_{top}. The junction temperature, t_j can be then computed as [2]:

$$t_j = t_{top} + \left(\mathrm{R}_{\theta jc} \cdot P\right) \tag{4}$$

where $\mathrm{R}_{\theta jc}$ denotes thermal resistance coefficient between device top surface and junction. The value of this parameter was provided in the chip datasheet and, in the discussed case, was equal to $\mathrm{R}_{\theta jc} = 1.22 \dfrac{^{\circ}\mathrm{C}}{\mathrm{W}}$.

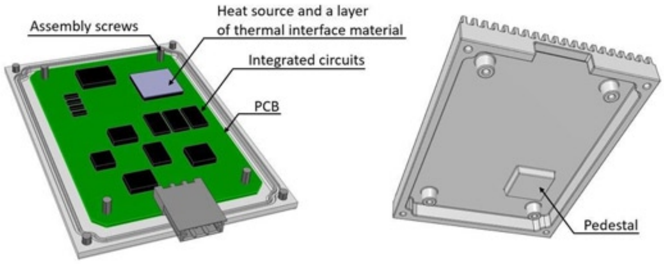


Fig. 3. Inner view of the analyzed electronic device and housing equipped with a pedestal.

4.1. Simulation results

The simulations described in this paper were carried out by means of CFD method, which allows for investigations of three-dimensional models and study the behavior of flow and heat transfer parameters in appropriate calculation space. In order to describe and solve the problem of finding the fluid flow pattern (in the discussed example the air was modelled inside and outside of the controller housing), the utilized CFD method numerically solves Navier – Stokes equations. Transport of thermal energy was considered in the calculation models by means of all three heat transfer mechanisms, that is, conduction, convection and radiation. It was assumed, that flow around the controller housing is laminar and that convective heat exchange is driven by natural convection, modelled by Boussinesq approximation. Heat radiation process was modelled by the well-known DO (Discrete Ordinates) approach, in which the radiation equation is solved along discrete number of directions established for every heat source. For accuracy, each radiation octant was divided into five angular steps, both in azimuthal and polar directions. Each simulated case was considered as a steady state problem. The numerical models prepared for the analysis consisted of ca. 4 millions of elements. Each simulation was stopped after residual values for the momentum, velocity and energy equations reached values below 10^{-4}, 10^{-4} i 10^{-8}, respectively. Additionally, temperature change observed on the top surface of the Component I heat source was negligible in every two consecutive solver iterations.

Figure 5 depicts the temperature field calculated on the surface of the PCB, Component I and other electronic devices, obtained for the standard cooling solution (pedestal above the main heat source). The temperature on the top surface of Component I reached in this case ca. 103.7°C.

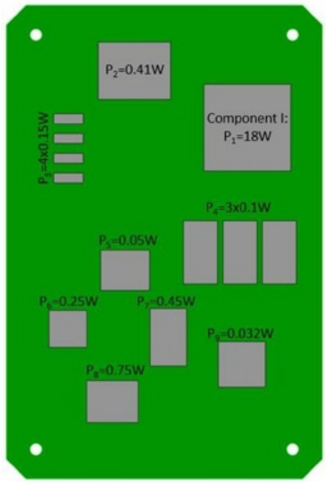


Fig. 4. Power dissipated by the components of the analyzed electronic device.

Junction temperature of this component was calculated using equation (4): $t_j = 103.7 + (1.22 \cdot 18) = 125.66$ °C. It exceeded slightly the allowable t_j^{max} limit of 125°C. However, possible variation with respect to the simulated operating conditions (e.g., accumulation of dirt on the housing surface and between the fins or strong thermal radiation from other devices), can cause further increase of t_j temperature. This, in turn, can lead to failure of the device and severely limit its operational lifetime.

In order to decrease the operating temperature of the controller components, in the subsequent design iteration, a VC was integrated into the housing of the device. This design modification was aimed at

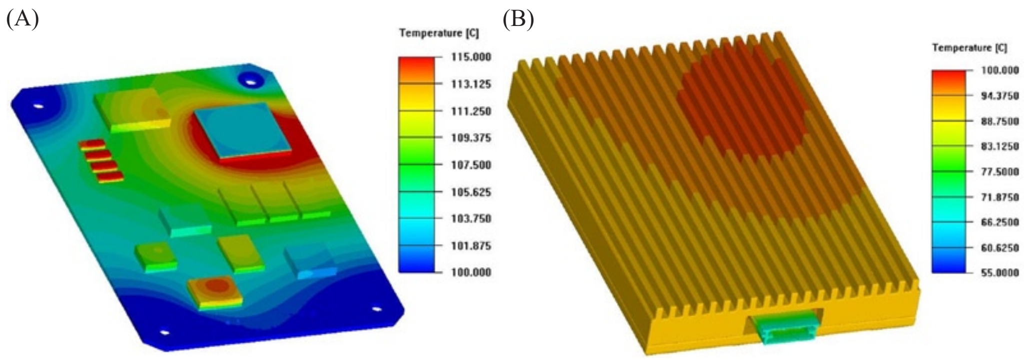


Fig. 5. Results obtained for the device equipped with a standard cooling solution. Temperature field plotted for: PCB and electronic components (A) and top side of the housing (B)

intensification of heat dissipation from the main heat source (Component I) to the fins on the top of the housing. It was assumed that demineralized water was the working fluid inside the VC. Additionally, the VC walls were made from copper, which is characterized by high heat conductivity value of $k_{Cu} = 385 \dfrac{\mathrm{W}}{\mathrm{m \cdot K}}$. The external dimensions of this component were equal to 75x90x4mm. As shown in Fig. 6, the VC was positioned inside the housing to stay in contact simultaneously with top surface of the analyzed heat source and inner wall of the aluminum enclosure. To minimize the interface thermal resistance, in both of these joints a 0.5mm layer of thermal interface material was used. Thermal conductivity for the latter material was equal to $k = 3.5 \dfrac{\mathrm{W}}{\mathrm{m - K}}$.

In order to effectively use the VC heat dissipation capability, inner top wall of the housing was modified by elimination of the pedestal.

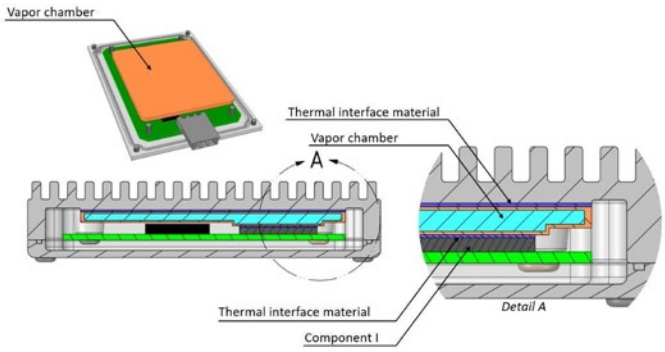


Fig. 6. Vapor chamber integrated into the housing of the analyzed device

This allowed for maximization of contact surface area between the enclosure and VC condenser region. For the sake of simulation simplification, it was assumed, that the processes of evaporation, convection, condensation and capillary transport of working fluid inside the VC can be represented as a steady state heat conduction using an isotropic material model. This approach is commonly used by simulation community. The authors of [7] and [13] presented results of separate experimental campaigns aimed at elaboration of simplified VC numerical models. In both cases, the derived VC simulation models were built from solid blocks of homogeneous material, characterized by extremely high thermal conductivity values of $8300-10000\frac{\mathrm{W}}{\mathrm{m}\cdot\mathrm{K}}$. In [13] the authors measured the VC ability to conduct the heat in various temperatures. According to the presented results, higher operating temperature led to greater equivalent thermal conductivity of the device. This can be explained as a result of increased intensity of evaporation process and more rapid transport of the vapor inside the device cavity, which take place in elevated temperature. It can be read in [13] that the measured equivalent thermal conductivity for the device subjected to experimental testing reached ca. $9750\frac{\mathrm{W}}{\mathrm{m}\cdot\mathrm{K}}$, when the device temperature was $50°\mathrm{C}$ above the ambient.

In the analysis described in this paper, the numerical model of VC integrated into the electronics housing was built using isotropic material, characterized by the equivalent heat conductivity equal to $k_{eq}=9000\frac{\mathrm{W}}{\mathrm{m}\cdot\mathrm{K}}$. Figure 7 shows the simulation results obtained for the discussed model.

The simulation results show, that integration of VC into the housing of the analyzed device causes decrease in temperature of all the components, PCB and the housing. The temperature drop and uniformity is clearly visible on the top surface of the main heat source, which stayed in contact with the VC. Maximum temperature value measured on the top surface of Component I was equal to $t_{max}^{vc}=99.95°\mathrm{C}$, which can be transformed into the junction temperature of this electronic device: $t_j^{vc}=121.91°\mathrm{C}$ (using equation (4)). Hence, the main heat source junction temperature difference in the two analyzed design alternatives reached $\Delta t_j=125.66°\mathrm{C}-121.91°\mathrm{C}=3.75°\mathrm{C}$. The obtained simulation results are combined in Tab. 2.

Figure 7 (B) depicts temperature distribution on the top side of the housing. The results show, that integration of a VC into the electronic device enclosure leads to more uniform flow of thermal energy, that is, to decreased density of heat flux, through this component.

Table 2. The results of the numerical analyses for two tested design alternatives

	Component I top surface temperature	**Component I junction temperature**
Design variant I: device housing equipped with a pedestal	103.70°C	125.66°C
Design variant II: VC integrated into the device housing	99.95°C	121.91°C

5. Expected lifetime of Component I

Silicone-based semiconductor devices operating in elevated temperature conditions undergo a number of physical and chemical degradation processes, which accelerate their failure. The most common failure mechanism that takes place in semiconductor component dies is electromigration, which consists in migration of material atoms under electric current load. The intensity of this phenomenon depends on material temperature – the higher it is, the more rapid the electromigration.

It is common to use the Arrhenius equation in order to estimate the electromigration degradation pace of semiconductor devices. However, as described in [21], if a silicone device operates in temperature above $105°\mathrm{C}$, this approach can lead to unreliable results, underestimating the failure reaction rate. The proposed in [21] procedure, which is based on empirical data, provides higher accuracy of the failure-free lifetime estimations. The procedure is based on calculation of the so-called acceleration factor AF, which is an input to equation (5) used for estimation of expected operating lifetime:

$$L_U=L_D\cdot AF \tag{5}$$

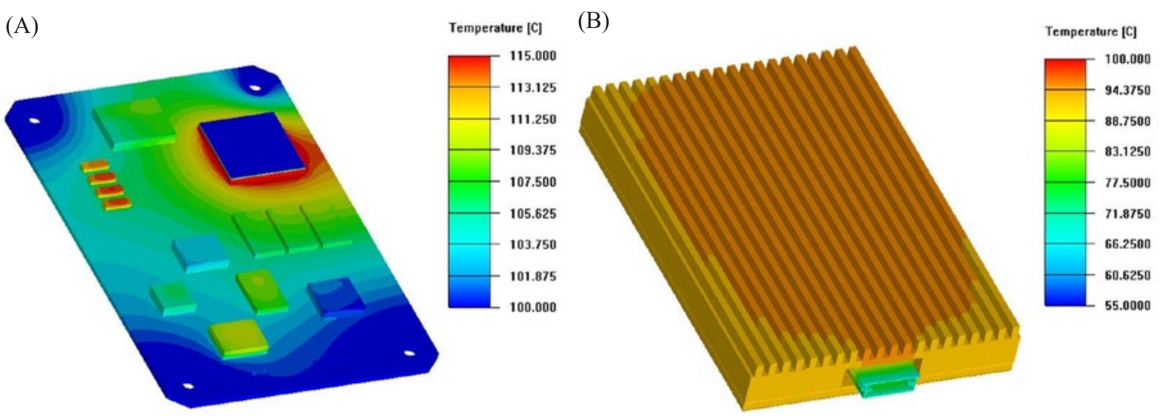


Fig. 7. Results obtained for the device equipped with a vapor chamber. Temperature field plotted for: PCB and electronic components (A) and top side of the housing (B).

where: L_D – nominal expected lifetime for junction temperature of $t_j = 105°C$, L_U – expected lifetime in specific operating temperature conditions.

According to [21], within junction temperature range of $110°C - 125°C$, the temperature dependency of the AF factor can be expressed by equation (6):

$$AF = -0.02t + 2.7 \tag{6}$$

where: $t[°C]$ denotes junction temperature in Celsius scale.

For the analyzed electronic device, the AF factor for Component I takes values of: $AF = 0.19$ and $AF^{vc} = 0.26$ for the standard cooling system and VC-based alternative, respectively. Assuming full load mission profile for this component at all times and the calculated junction temperatures of $t_j = 125.66°C$ oraz $t_j^{vc} = 121.91°C$, equation (5) leads to the conclusion, that integration of VC into the device housing will result in higher reliability of the main heat source. The change in the expected lifetime duration is equal to $0.07L_D$, therefore the device equipped with a VC is expected to operate 7% longer compared with the pedestal-based design alternative.

6. Summary

This paper presents the concept of vapor chamber integration into aluminum housing of an electronic device, operating in elevated ambient temperature of $t_{amb} = 55°C$. The effect of using VC-based cooling solution has been compared with classical, aluminum pedestal-based design. The obtained numerical simulation results demonstrate, that the use of two-phase cooling component leads to $\Delta t_j = 3.75°C$ reduction in junction temperature of the main heat source. In the assumed ambient conditions, this allows for a decrease in operating junction temperature below the limit of $t_j^{max} = 125°C$, as well as an increase in the device expected lifetime by 7%.

References

1. Aranzabal I, de Alegría I M, Delmonte N, Cova P, Kortabarria I. Comparison of the Heat Transfer Capabilities of Conventional Single-and Two-Phase Cooling Systems for an Electric Vehicle IGBT Power Module. IEEE Transactions on Power Electronics 2018; 34(5): 4185-4194, https://doi.org/10.1109/TPEL.2018.2862943.
2. Cengel Y A. Heat transfer: a practical approach, 2nd Edition. McGraw-Hill Higher Education, 2002.
3. Chainer T J, Schultz M D, Parida P R, Gaynes M A. Improving data center energy efficiency with advanced thermal management. IEEE Transactions on Components, Packaging and Manufacturing Technology 2017; 7(8): 1228-1239, https://doi.org/10.1109/TCPMT.2017.2661700.
4. Chang N, Pan S, Srinivasan K, Feng Z, Xia W, Pawlak T, Geb D. Emerging ADAS Thermal Reliability Needs and Solutions. IEEE Micro 2018; 38(1): 66-81, https://doi.org/10.1109/MM.2018.112130058.
5. Gadalla B, Schaltz E, Blaabjerg F. A survey on the reliability of power electronics in electro-mobility applications. Proceedings of: 2015 Intl Aegean Conference on Electrical Machines & Power Electronics (ACEMP), 2015 Intl Conference on Optimization of Electrical & Electronic Equipment (OPTIM) & 2015 Intl Symposium on Advanced Electromechanical Motion Systems (ELECTROMOTION) 2015; 304-310, https://doi.org/10.1109/OPTIM.2015.7427015.
6. Gu J, Pecht M. Predicting the reliability of electronic products. Proceedings of: 8th International Conference on Electronic Packaging Technology 2007; 1-8, https://doi.org/10.1109/ICEPT.2007.4441552.
7. Gurevich A, Steiner I, Huang E. Design of thermal systems based on combination of Thermoelectric and Vapor Chamber technologies. Proceedings of: 33rd Thermal Measurement, Modeling & Management Symposium (SEMI-THERM) 2017; 1-5, https://doi.org/10.1109/SEMI-THERM.2017.7896899.
8. Hager M, Gromala P, Wunderle B, Rzepka S. Affordable and Safe High Performance Vehicle Computers with Ultra-Fast On-Board Ethernet for Automated Driving. Advanced Microsystems for Automotive Applications 2018: Smart Systems for Clean, Safe and Shared Vehicles. Springer, 2019, https://doi.org/10.1007/978-3-319-99762-9_5.
9. Hammoud J, Dudley S, Apte N. Use of Heat Pipe Technology for Multi Media Thermal Management. Proceedings of: SAE 2006 World Congress & Exhibition 2006; 1-3, https://doi.org/10.4271/2006-01-0482.
10. Khazaka R, Mendizabal L, Henry D, Hanna R. Survey of high-temperature reliability of power electronics packaging components. IEEE Transactions on power Electronics 2014; 30(5): 2456-2464, https://doi.org/10.1109/TPEL.2014.2357836.
11. Ma X, Guo Y, Wang L, Ji W. Exploration of the reliability of automotive electronic power steering system using device junction electrothermal profile cycle. IEEE Access 2016; 4: 7054-7062, https://doi.org/10.1109/ACCESS.2016.2621034.
12. Mochizuki M. Latest development and application of heat pipes for electronics and automotive. Proceedings of: IEEE CPMT Symposium Japan 2017; 87-90, https://doi.org/10.1109/ICSJ.2017.8240095.
13. Niittymaki L, Biber C, Carbone M. Flexible CFD simulation model of a thin vapor chamber for mobile applications. Proceedings of: 22nd International Workshop on Thermal Investigations of ICs and Systems (THERMINIC) 2016; 152-157, https://doi.org/10.1109/THERMINIC.2016.7749044.
14. Qi F, Wang Y, Bob-Manuel C, Li H, Jones S, Li B, Chen Y, Yan Y. Advanced Cooling Solutions of High Power Automotive Module. Proceedings of: PCIM Asia 2017: International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management 2017; 1-3.
15. Ram Sundar R, Sarweshwara Sarma N S V. Performance study of motorcycle driving-beam LED headlight with different heat sink models and LED pitch. Proceedings of: 16th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm) 2017; 784-792.
16. Riches S, Johnston C. Electronics design, assembly and reliability for high temperature applications. Proceedings of: IEEE International Symposium on Circuits and Systems (ISCAS) 2015; 1158-1161, https://doi.org/10.1109/ISCAS.2015.7168844.
17. Robinson A J, Colenbrander J, Kempers R, Chen R. Solid and Vapor Chamber Integrated Heat Spreaders: Which to Choose and Why. IEEE Transactions on Components, Packaging and Manufacturing Technology 2018; 8(9): 1581-1592, https://doi.org/10.1109/TCPMT.2018.2822400.

18. Simon D, Boianceanu C, De Mey G, Ţopa V. Experimental reliability improvement of power devices operated under fast thermal cycling. IEEE Electron Device Letters 2015; 36(7): 696-698, https://doi.org/10.1109/LED.2015.2432128.
19. Sun L, Gu X H, Song P, Di Y. A generalized equivalent temperature model in a time-varying environment. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19(3): 432-440, https://doi.org/10.17531/ein.2017.3.14.
20. Tan C M, Narula U, Kapoor D. Reliability paradox for worldwide automotive electronics. Proceedings of: IEEE Annual Reliability and Maintainability Symposium (RAMS) 2017; 1-7, https://doi.org/10.1109/RAM.2017.7889654.
21. Webber A. Calculating Useful Lifetimes of Embedded Processors: Application Report. Texas Instruments Incorporated, 2017.
22. Zeng S, Sun B, Tong C. A modified model of electronic device reliability prediction. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2009; 4(44): 4-9.

Jakub KORTA

Paweł SKRUCH
AGH University of Science and Technology
Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering
Department of Automatic Control and Robotics

E-mails: jakub.korta@ieee.org, pawel.skruch@agh.edu.pl

Article citation info:
POURHASSAN MR, RAISSI S, HAFEZALKOTOB A. A simulation approach on reliability assessment of complex system subject to stochastic degradation and random shock. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2020; 22 (2): 370–379, http://dx.doi.org/10.17531/ein.2020.2.20.

Mohammad Reza POURHASSAN
Sadigh RAISSI
Ashkan HAFEZALKOTOB

A SIMULATION APPROACH ON RELIABILITY ASSESSMENT OF COMPLEX SYSTEM SUBJECT TO STOCHASTIC DEGRADATION AND RANDOM SHOCK

METODA SYMULACYJNA OCENY NIEZAWODNOŚCI ZŁOŻONEGO SYSTEMU PODLEGAJĄCEGO PROCESOM DEGRADACJI STOCHASTYCZNEJ I NARAŻONEGO NA OBCIĄŻENIA LOSOWE

Many systems are affected by different random factors and stochastic processes, significantly complicating their reliability analysis. In general, the performance of complicated systems may gradually, suddenly, or continuously be downgraded over times from perfect functioning to breakdown states or may be affected by unexpected shocks. In the literature, analytic reliability assessment examined for especial cases is restricted to applying the Exponential, Gamma, compound Poisson, and Wiener degradation processes. Consideration of the effect of non-fatal soft shock makes such assessment more challenging which has remained a research gap for general degraded stochastic processes. Through the current article, for preventing complexity of analytic calculations, we have focused on applying a simulating approach for generalization. The proposed model embeds both the stochastic degradation process as well randomly occurred shocks for two states, multi-state, and continuous degradation. Here, the user can arbitrarily set the time to failure distribution, stochastic degradation, and time to occurrence shock density function as well its severity. In order to present the validity and applicability, two case studies in a sugar plant alongside an example derived from the literature are examined. In the first case study, the simulation overestimated the system reliability by less than 5%. Also, the comparison revealed no significant difference between the analytic and the simulation result in an example taken from an article. Finally, the reliability of a complicated crystallizer system embedding both degradation and soft shock occurrence was examined in a three-component standby system.

Keywords: *system reliability; multi-state system; competing failures; stochastic degradation; random shocks; discrete event simulation.*

Prawidłowe działanie wielu systemów zależy od różnych czynników losowych i procesów stochastycznych, co znacznie komplikuje analizę niezawodności tych układów. Parametry pracy skomplikowanych systemów mogą ulegać stopniowemu, nagłemu lub stałemu obniżeniu ze stanu doskonałego funkcjonowania do stanu awaryjnego. Wpływ na nie mogą też mieć niespodziewane obciążenia. W literaturze przedmiotu, analityczną ocenę niezawodności stosuje się do badania przypadków szczególnych i ogranicza do badania degradacji w oparciu o proces wykładniczy, proces gamma, złożony proces Poissona i proces Wienera. Ocena niezawodności z uwzględnieniem wpływu obciążeń miękkich, nieprowadzących do całkowitej awarii, stanowi większe wyzwanie tworząc lukę w badaniach nad ogólnymi stochastycznymi procesami degradacji. Aby uniknąć złożonych obliczeń analitycznych, w niniejszej pracy skupiliśmy się na zastosowaniu podejścia symulacyjnego w celu uzyskania generalizacji. Proponowany model obejmuje zarówno stochastyczny proces degradacji, jak i losowo występujące obciążenia i uwzględnia przypadki degradacji systemów dwustanowych, wielostanowych oraz degradacji ciągłej. Posługując się tym modelem, użytkownik może dowolnie ustawiać rozkład czasu do uszkodzenia, degradację stochastyczną, czas do wystąpienia obciążenia, funkcję gęstości prawdopodobieństwa wystąpienia obciążenia, a także jego nasilenie. Trafność oraz możliwości zastosowania przedstawionego modelu zilustrowano na podstawie dwóch studiów przypadków dotyczących cukrowni oraz przykładu zaczerpniętego z literatury. W pierwszym studium przypadku, poziom niezawodności systemu obliczony na podstawie symulacji różnił się o mniej niż 5% od wyniku otrzymanego na drodze analitycznej. Porównanie nie ujawniło również żadnej istotnej różnicy między wynikiem analitycznym a symulacyjnym w przykładzie pochodzącym z literatury. Artykuł wieńczy analiza niezawodności złożonego układu krystalizatora, obejmująca zarówno degradację, jak i występowanie miękkich obciążeń w trójelementowym systemie krystalizatora z rezerwą.

Słowa kluczowe: *niezawodność systemu; system wielostanowy; uszkodzenia konkurujące; degradacja stochastyczna; obciążenia losowe; symulacja zdarzeń dyskretnych.*

1. Introduction

Reliability is common scientific characteristic of a system with commutability, operability, or usability upon any request to accomplish the relevant nominated tasks over time to finally evaluate the system potential or performance. In this regard, their assessment is a crucial analytic task given the huge complexity and solving the many states equation especially in the presence of stochastic degradation process and random arrival shock with unknown severity. This context has remained a research gap, which has attracted much attention in the literature by Patelli et al. [17]

In general, the system reliability is analyzed in three ways: binary or two states, multi-state, and degradation process which present system state continuously over time. Commonly, to avoid heavy calculations, reliability assessment is carried out for a few states. Such an approach employs an oversimplification in many real-life situations where the system is accomplished based on assuming a comprehensive range of states, varying from perfect functioning to complete breakdown.

Conventionally, system reliability assessment should follow tedious computations for simultaneously solving a large number of differential equations to calculate the probability of the system being in each state. Then system reliability is computed based on summation of probability of all states where the system functions well. In many real cases, a system may become degraded when being subject to random shocks with different degrees of severity. This may predispose the system to fail suddenly or accelerate their degradation process. Hence, applying analytic methods for reliability assessment is complicated especially for multi-state systems. Accordingly, in the present study, we have developed an appropriate and efficient simulation approach for this issue to develop the professional capabilities required by analysts.

Degradation refers to either performance degradation (e.g. power output of a generator) or some measure of actual degradation (e.g., toxic concentration in a chemical process or fatigue crack in a gear). Commonly, the degradation process reveals a continuous alteration of the system state over time. Once a proper degradation variable is selected, degradation data, when properly measured, could provide substantial information as there are quantitative measurements (not just at discrete points of time or their failure). Indeed, it is possible to make powerful reliability inferences from degradation data even when there are no failures.

The degradation process could be modeled using the experimental data through degradation path modeling method. Stochastic degradation process tends to model the degradation variable over time while considering the measurement error. This kind of modeling consists of two terms to present the deterministic behavior of the variable level through a linear, quadratic, exponential and other terms alongside the error term which is described by a given random term; e.g. Gamma, Logistic or Weibull distribution (Nikulin et al. [16]).

Additionally, most engineering systems suffer from catastrophic events occurring randomly and they could cause sudden breakdown or initiate other mechanisms which accelerate the failure process. Hence, the time of shock occurrences and their severity are presented by two random variables. A sophisticated review on shock modeling methods in reliability engineering has been presented by Finkelstein Maxim and Cha Ji Hwan [7].

Through current research, we focused on answering to the following research questions.

1. How to estimate reliability of a multi-component multi-state system on the based on simulation modeling?
2. What is the consequence of randomly occurred shocks?
3. How to estimate reliability of a gradually stochastic degraded system?

The rest of this paper is organized as follows. Section 2 reviews the literature on reliability assessments of binary, multi-state, and degraded processes using analytic methods and simulation models. Due to complex calculations of the analytic method, essential basis as well the proposed simulation model are presented in section 3. The validation method for the proposed simulation model presented on section 4. Section 5 discusses a real case study to clarify the proposed method in details. Finally, section 6 closes the paper with concluding remarks, advantages, and drawbacks to be covered by future research.

2. Literature review

Many research efforts have been made to assess reliability of systems. So far, a great deal of attention has been paid to system reliability analysis which deals with a binary state system describing system states using functioning or failure states via a specified random variable. Over the past few decades, reliability practitioners have been working on analyzing system reliability using more data collected during the system life time. In this way, the system state has been evaluated over time discretely or continuously through multi-state or degraded level. Some reliability experts have also focused on other sources accelerating failure process such as shock or hazardous events.

An analytic model to evaluate a degrading binary system during a fatal shock has been presented in Riascos-Ochoa et al. [19]. They fitted a phase-type distribution to inter-arrival time in case of shock occurrence. This approach helps users evaluate one single component reliability. Another research Caballé and Castro [3] proposed a model with internal degradation under a gamma process and random shocks with non-homogeneous Poisson process. In addition, they analyzed the robustness of the solution by changing the input parameters. Also, binomial shock process was evaluated in Eryilmaz [5]. They extended their model to the presence of shock dependent processes using the Markov chain. An analytic model for a single component on the presence of hard failure (shocks) and soft failures (degradation) for fault-tolerant systems was prosed in Liu et al. [14]. They also implemented a proposed model on an example to show the applicability to many systems via a model with cumulative shocks based on batch Markovian arrival process Montoro-Cazorla and Pérez-Ocón [15]. In their model, shock processes are interdependent and the system failure occurs when the number of cumulative shocks exceeds the defined threshold. A Stress-Strength model was developed in Hao et al. [9] for soft and hard failures and their interactions. The results revealed a positive correlation between shock process and degradation performance and the mutually dependent processes had direct effects on the system reliability. In Rafiee et al. [18] , a generalized mixed shock model involving fatal and non-fatal shocks was presented analytically. In this paper, three types of shock patterns were taken into account. A sensitivity analysis was applied on an example of micro-electro-mechanical system to show the application of the proposed model.

For analyzing system reliability in shock-degradation models, some methodologies have been presented. For example, Huang et al. [10] offered an analytic method for reliability assessment for a system affected by smooth degradation with the gamma process and traumatic failure caused by Poisson shock process. In their proposed method, given the two processes in that, with increase in the degradation level, the probability of traumatic failure caused by a random shock increased. A degradation model for in civil structures was presented by Wang et al. [22]. This model had two main components: non-increasing stochastics parameters and existing correlation between load processes and degradation. They also developed this model on a numerical example to analyze sensitivity of reliability on degradation and shock processes as well as the load-deterioration correlation. They found that the system reliability was very sensitive to variations of cumulative deterioration and shock. Further, two different types of dependent competing failure processes were modeled in An and Sun [1].

At first, a shock and a degradation process were considered simultaneously along with the existing interdependency between them. Secondly, multiple degradation processes with their correlation were added to the model. Finally, by extending a numerical example, sensitivity analysis was made to evaluate the effects of parameter models on the system reliability. The reliability of a system based on the presence of fatigue degradation and shock processes was evaluated by Zhang et al. [26]. They considered retardation event in their model, i.e. fatigue procedures were retarded when the shock occurred in the system. They evaluated the reliability in two case studies: in the first one, they considered shock processes as a fixed time period, while in the second, shock occurred at various time periods. The reliability of load-sharing systems with dependent shock and degradation processes was investigated by Che et al. [4] . In their assumed system, the failure time of components, the time between arrival of shocks, and their interaction were stochastic, so they used an analytical method to analyze the reliability of the system. Experimental results indicated that the reliability in load-sharing systems was lower than in simple parallel systems.

Some authors used a simulation model to model and analyze the reliability of systems. Monte Carlo simulation was applied to analyze the system reliability of a degradation-shock process model in Fan et al. [6] . In their simulation model, the shock process was influenced by the soft failures (degradation) and random shocks were categorized into three zones based on their magnitudes. In Warrington and Jones [23], a discrete event simulation with path-sets methodology was presented to generate a dynamic model for analyzing the reliability of system. Another simulation model for reliability of a self-healing network for scalable and fault-tolerant, parallel runtime environments was evaluated by Angskun et al. [2]. They used a simulation method to calculate the system under failure conditions. Gola [8] focused on the way to estimate system reliability with changing machine due to maintain the production process stability using Enterprise Dynamics software.

In Vaisi et al. [21], an availability reliability model for a two-machine robotic cell was presented for different sources of uncertainty. They implemented this structure on a multi-state transmission system. Further Juan et al. [12] presented a simulation methodology in a time-dependent building for civil engineering structures. They discovered that the simulation method could offer more advantages over other approaches, since it could measure details such as multi-state systems and discover critical components in a structure. Many researchers such as Kosicka et al. [13], Jasiulewicz-Kaczmarek and Gola [11], Zaim et al. [25] and Sobaszek et al. [20] focused on the way to increase system reliability in different aspects. Interested readers may follow some beneficial methods in

In the case of reliability assessment for a multi-state system, analytic calculation is highly complex. Most cases have focused on the exponential time to failure and time to repair distribution. Wenjie et al. [24] proposed a reliability index for a repairable multi-state component using homogenous continuous time Markov chain process. Their method was limited to a shirt time repair process. Further, they tried to balance the maintenance cost and lifetime of multi-state components in an illustrated example. The complexity of computation of analytic methods encourages researchers to use efficient simulation techniques especially for reliability assessment of multi-component multi-state systems.

As the literature suggests, almost all existing methods have been conducted based on restricted assumptions. The reliability assessment of a multi-state system which degrades randomly and suffers from competing random shock effects when all random variables could not be modeled in any free given process has been research gap so far. This paper proposes an efficient computer simulation approach in reliability assessment of a multi-state system subject to stochastic degradation process and randomly occurring non-fatal shocks. The proposed method has no restriction on applying Markov or semi-Markov process. It has also the capability to be applied for any time of occurrence of shocks, and for their relevant consequences and any random degradation process.

3. Computer simulation model basis

Basically, the reliability analysis of a multi-state system depends its component features, states and the system RBD. Suppose a multi-state system consisting of N components where each component j could have k_j different states corresponding to its performance rates, represented by Eq. 1:

$$g_j = \left\{ g_{j1}, g_{j2}, \ldots, g_{jl}, \ldots, g_{jk_j} \right\} \tag{1}$$

where, g_{ji} is the performance rate of component j in the state $i \in \left\{1, 2, ..., k_j\right\}$. Suppose that the performance rates are arranged in a descending order at different states of each component. For example, in state g_{j1}, the performance rate of component j is perfect and complete. When the state transfers from g_{j1} to g_{j2}, the performance rate will decrease for example to 90%. This descending status will continue until the state of the component reaches g_{jl}. In the remaining states $\in \{l+1, l+2, ..., k\}$, the system does not have an adequate performance level which could be called complete breakdown. So, the element j may be functioning well in the state $q \in \{1,2,...,l\}$. Figure 1 displays typical performance rates for a system at 12 different states which could be considered during the last four states.

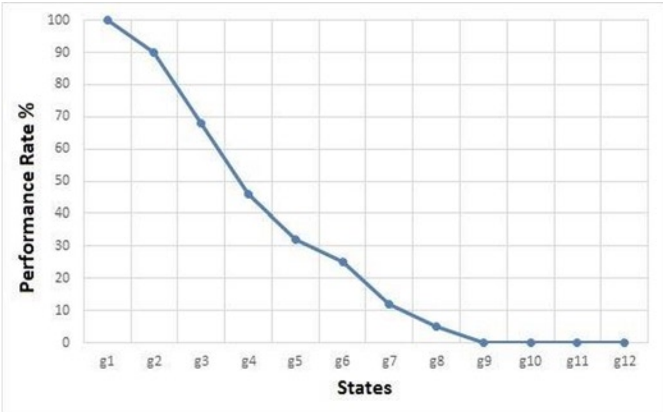


Fig. 1. A typical degrading system performance over time

The probabilities related to the different states of the component j at any instant time t can be displayed by the Eq. 2:

$$p_j(t) = \left\{ p_{j1}(t), p_{j2}(t), \ldots, p_{jl}(t), \ldots, p_{jk_j}(t) \right\} \tag{2}$$

where $p_{ji}(t) = \Pr\{ \text{Gj (t)} = g_{ji}\}$ and $\sum_{i=1}^{k_j} p_{ji}(t) = 1$.

So the reliability of each component presents by Eq. 3:

$$R_j(t) = p_{j1}(t) + p_{j2}(t) + \ldots + p_{jl}(t) = \sum_{i=1}^{l} p_{ji}(t) \tag{3}$$

Finally, in a multi-state system with N series components, the reliability is equal to Eq. 4:

$$R_s(t) = \prod_{j=1}^{N} R_j(t) \tag{4}$$

Also, for a multi-state system with N parallel component, the system reliability is calculated by Eq. 5:

$$R_s(t) = 1 - \prod_{j=1}^{N}(1 - R_j(t)) \tag{5}$$

Degradation processes act as a failure mode and are often defined by a smoothing continuous damage accumulated over time. In addition, temporal variability should be taken into account during the system degradation process. It usually may be modeled by a given deterministic curve beside an error terms follows from a given statistically density function such as Exponential, Gamma, Logistic, Weibull, or etc. In a multi-state system with N elements, the performance rate of component j in the state i may be reduced or updated due to its relevant deterioration process or shock occurrence as Eq. 6:

$$\dot{g}_{ji}(t) = \alpha_{ji}(t).\, g_{ji}(t) \tag{6}$$

where, $0 < \alpha_{ji}(t) < 1$ is the degradation rate of performance of element j in the state i at time t, and the prime accent on g denotes the updated values for the performance rate.

Suppose $X_j(t)$ represents the degradation level of component j at time t. For a continuous or discrete degraded process, each component fails if the relevant level exceeds its threshold l_j. Let T_{dj} be the failure time of degradation process; thus, the reliability of the system while only considering with degradation process can be then estimated by the fraction of the time when the system performance level is greater than the threshold level l. Equivalently, in mathematical terms for each component, the reliability is given by Eq. (7):

$$R_j(t) = \Pr\left(t < T_{dj}\right) = \Pr\left(X_j(t) > L_j\right) = \Pr(g_j(t) > l_j) \tag{7}$$

Consequently, the simulation model extracts the component reliability output through the counting ratio of the desired condition over the total runs. Note that after running the simulation model over given period of time, this equation is applied for individual estimation of only single component reliability not the system reliability. Formerly, system reliability could be calculated using reliability block diagram indicating how component reliability contributes to the success or failure of a complex system. After a few repetitions of the process for different simulated observation periods, reliability curve can be illustrated.

Shock is another common competing cause in system failures which accelerates the component degradation rate or random failure processes. The literature has pointed to two kinds of shock; fatal and non-fatal shock. The first type causes rapid disruption while the second accelerates the degradation process. Thus, the transition rate between consecutive states grows progressively.

The proposed simulation method need to following inputs.

1. The system configuration in terms of System Reliability Block Diagram (RBD). For each component, time to failure density function needed to defined necessary. Also specify the component's repair time density function if needed to calculate availability.
2. State transition matrix for each component just for multi-state system. This modulus has no need to be defined for continuous degraded systems.
3. Degradation modulus. For each component just one time dependent degradation model should be described as well a time to failure density function.
4. Random shock occurrence modulus which embeds time to shock occurrence density function and its random effects in terms of a constant or random density function.
5. System reliability/availability estimation modulus.

The first four moduli need some inferences on the system and which are commonly used for reliability analyzers and should be set as simulation model inputs. These moduli are designed individually and some of them are not necessary in all problems. Figure 2 illustrates the proposed simulation method schematically.

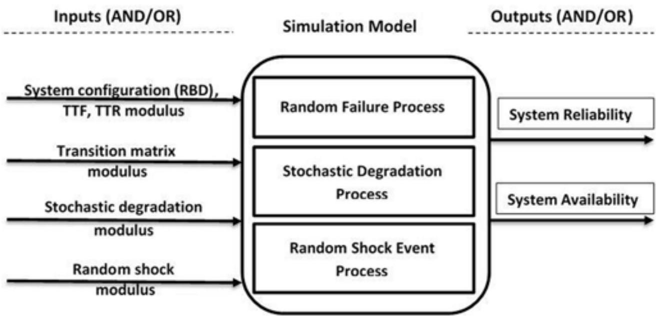


Fig. 2. A schematic view of the proposed computer simulation method

The basic output modulus accounts the system reliability through summation of the probability of existence of a system in a given set of desired states after running the system for a given replications. Here, the system configuration should be defined via the system RBD and breakdown features of some components. They are: 1) failure patterns; hence time to failure (TTF) and time to repair (TTR) density function for each component should be acknowledged experimentally. Degradation process which may be presented by a deterministic curve and randomly distributed residuals/error terms (ET). Random shocks through defining the time between occurrence/arrival (TBA) alongside its severity statistical distribution (S). Here, S represents the substitution of extremely high hazard rate models of a fatal shock; otherwise it is a repeated non-fatal shock which accelerates the degradation process. An alternative approach to apply the soft shock consequences surveyed in the 2nd case study report.

In order to simulate such a system, we applied an object-oriented approach where all components are represented by an individual object using the Enterprise Dynamics Incontrol simulation software; ED™. In this simulation, the package of almost all well-known density functions is ready to use at its library. So, TTF, TTR, TBA, ET and S could be easily interpreted via a well-known statistical density function. We applied a "Server" to model each component. Accordingly, the number of atoms in the model should be related to the number of components. Here, their relevant cycle time could be adjusted to any constant or random value. So, we set them to zero for immediate processing. In order to estimate the system reliability, it is necessary to eliminate TTR effects. In other words, the repair time should be considered as a very large value to prevent completion of the repairing process. Each simulation run takes a long period of time. When the simulation model is run, a continuous flow of "Product" entity is created. The entity flow simulates the desired component performance and may change the system state over time until one of the failure modes occurs.

At the beginning of simulation, all components are set on their first state with its maximum performance until any significant discrete event occurs over time where the entrance of any entity affects them through degradation process or shock occurrence modulus. These two individual moduli have their own network. In the basic network, the entity is allowed to go to the next Server based on the system RBD. This sequence will continue to meet the last server. Figure 2 typically presents a sample layout for the simulation of only one component which has four states. Note that this sample network does not reveal the degradation process and shock occurrence and it could cover only random breakdowns. So, no triggers from other modulus affect the basic network. Consequently, the probability of the system being on each state could be calculated by counting the number of entities transitioning into the respective states. For example, consider after a two-hour simulation, the number of entities reaching states 1, 2 and 3 is equal to 200, 150 and 80 respectively.

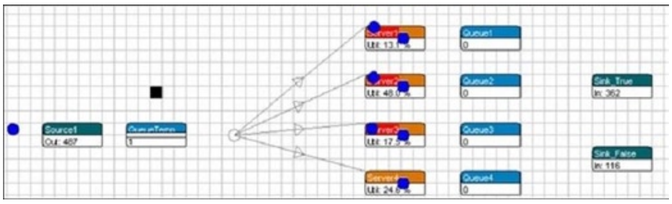


Fig. 3. An Enterprise Dynamic layout for a four-states system

Suppose the component is working both in the first two states, consequently the reliability calculated through $\frac{200+150}{200+150+80}=81\%$ for a period of 430 hours. After replicating the model, the relevant curve could as plotted as shown by Figure 4. More replication conducts to the more smooth curve.

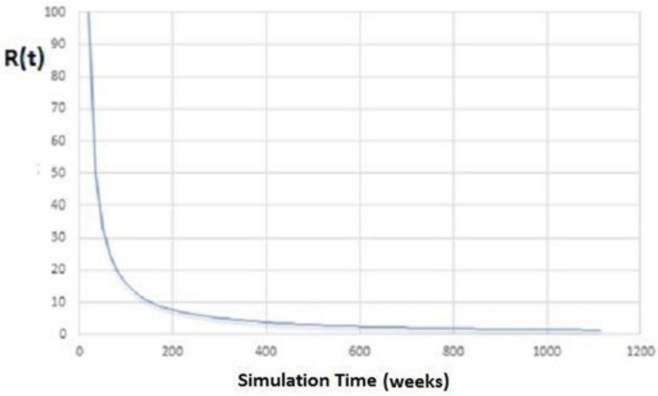


Fig. 4. A typical reliability curve

As a parallel of the basic network, two individual networks could influence each component failure by sending triggers. The first simulates the degradation process while the second models the shock occurrence. If fatal shock occurs, the relevant component immediately fails; otherwise the time to failure time drops by a constant adjustable parameter. This trick enabled us to model the accelerated breakdown process by a constant percentage. The shock network need two inputs: 1) Inter-arrival shock process TBA which could be set by any random distribution (e.g. Exponential, Gamma, Weibull, …), and 2) The shock consequence table which embeds the updated degradation rates; $\alpha_{ji}(t)$. For a fatal shock, the practitioner may set the rates to large value to immediately reveal the failure of the component.

4. Validation of the proposed simulation model

In order to examine the computer simulation model capability, we followed a two-step process. First, the validity of the simulation model was examined and compared with an analytic method. Once the validity of the simulation model was verified, it was used to estimate the system reliability in the desired complex systems. Hence, applying an analytic model is a very difficult technical task.

In order to validate the simulation model, a crystalizing shaft crack grow model was considered on a sugarloaf plant in Shahroud, Iran. A research team consisting of five skilled persons from the University studied 10 shaft samples of one type crystallizer machine for 100 consecutive days in terms of crack growth using an X-ray machine. Here, the crack initiation time was uncertain and should have been estimated. Experts believed that since the shaft under study was not a critical component and could not withstand high mechanical loads, the maximum crack length should have been less than 100 mm. Table 1 reports the values of the experimental data for the measured crack length. .

In order to examine the degradation models, we overviewed the crack length of samples over time. Figure 1 displays such actual degradation. Rottenly linear, exponential and power models act as a proper candidate for the fitness function. Since the linear and exponential models did not have an adequate fitting index in terms sum of squared error, the best fit was obtained based on the power model. The general form of such fitness function is presented by the function below:

$$D_t = b\left(t - t_0\right)^a \tag{8}$$

where, D_t denotes the crack length in time. Hence t_0 begins for the crack initiation time and like a and b parameters they should be estimated using maximum likelihood estimation method from experimental data. Table 2 presents an estimation of the power model parameters. In this table, the failure time was calculated based on the $D_{Max} = 100mm$ and the growth time was calculated according to the period of time between the failure time and crack initiation time.

Fitting a different probability distribution function to the calculated crack growth time reveals that the Weibull has a good candidate to estimate crack growth time density function. Figure 5 presents the relevant probability plot justification. Low amount of the Anderson Darling statistic; 1.657 beside high p-value; greater than 0.1 reveals that we has not any evidence to reject the Weibull distribution in the goodness of fit hypothesis testing.

Using the Minitab statistical package, the reliability plot of the shaft growth time using 95% confidence interval is illustrated in Figure 6. According to the results, the newly received shafts do not have good quality and their cracking started within the interval of 8 to 24 days reaching their maximum permissible average over a period of 110 days. Also, the reliability of these shafts has been only 51.57%

Table 1. Crack length (mm) of 10 samples of crystallizer machine shaft

Time (days)	Sample No.									
	1	2	3	4	5	6	7	8	9	10
10	13	11	13	24	8	22	16	18	12	14
20	34	46	32	71	38	60	33	48	31	35
30	45	69	40	108	54	90	42	67	40	45
40	55	88	45	139	65	120	50	84	48	50
50	63	103	52	165	75	143	55	100	53	60
60	71	118	54	195	85	165	60	112	58	63
70	77	130	60	216	92	187	61	124	60	68
80	84	144	65	243	100	210	64	135	65	70
90	89	153	67	264	106	225	70	145	70	75
100	93	165	70	285	111	245	72	158	75	80

Table 2. Degradation parameters and time to failure estimation for crack process

Parameters	Sample No.									
	1	2	c	4	5	6	7	8	9	10
t_0	13	11	13	24	8	22	16	18	12	14
a	34	46	32	71	38	60	33	48	31	35
b	45	69	40	108	54	90	42	67	40	45
Failure time	111.15	48.00	231.47	27.78	81.42	33.02	204.45	51.288	213.91	176.01
Growth time	63	103	52	165	75	143	55	100	53	60

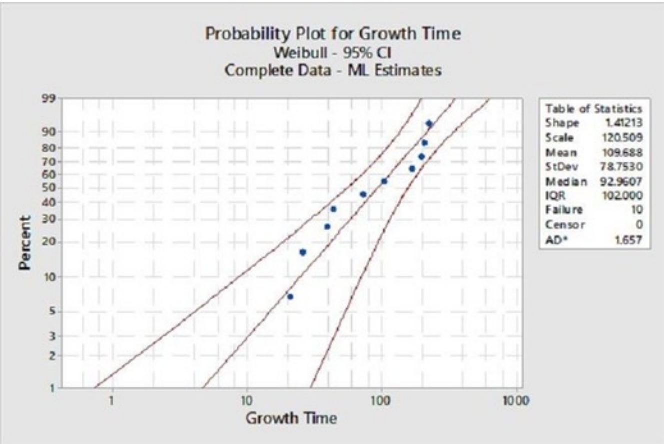


Fig. 5. The crack growth time probability plot based on the two parameter Weibull distribution

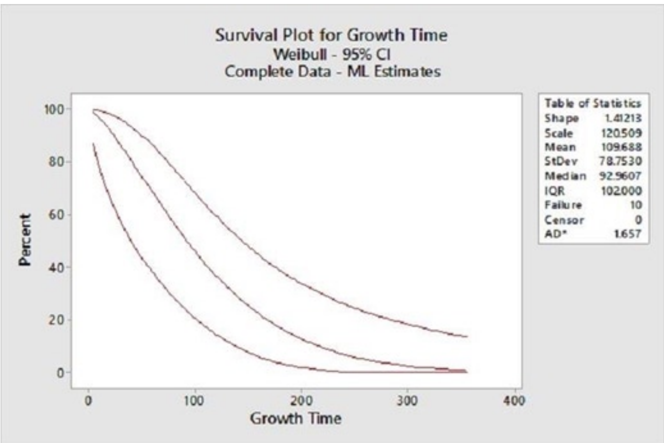


Fig. 6. Fig. 6. The 95% confidence interval for reliability of the growth time

for 90 days of operation after cracking initiation; so, this type of shaft should be replaced with higher quality types.

As mentioned earlier, applying a technical method for reliability of a case study needs many sequential statistical methods. Since the shaft breakdown tends to fail the crystallizer machine and there are a set of different machines in this department, so the reliability analysis of such departments is very time-consuming especially when the system has a complicated reliability block diagram of multiple multi-component machines engaged with different stochastic degradation processes and random soft/hard shocks. A feasible efficient solution method could be applied using a computer simulation model.

Let us now examine the system reliability estimation through simulation model for the first case study report. This case embeds one degradation process alongside a soft shock occurrence. In that case at any given time, the crack length develop from a Weibull distribution with 1.412 and 120.509 respectively for the shape and the scale parameters. Hence the mean of the degradation process is assumed to develop a power function based on the mean estimates of samples for a , b , t_0 parameters as function $D_t = 11.34(t-8.33)^{0.521}$. Defining one block showing the crystallizer in the first modulus and setting a two-state system transition diagram with time to failure density function based on the Weibull (1.412,120.504) in the second modulus while recognizing the degradation function in the fourth modulus are necessary to estimate system reliability at a given time, say 90 days. Running the simulation model under 5 runs reported the values of system reliability as: 45.1 %, 64.9%, 57.1%, 54.4%, and 49.2%. Thus, the simulation model estimated the mean shaft reliability as 54.14%, only 4.7% overestimation in relation to the analytic method.

In order to further examine the simulation model validity, we run the sample model using a two-parameter gamma degradation process. According to Huang et al. [10], the reliability of such a process is given by Eq. (9):

$$R_d(t) = 1 - \int_L^{\infty} f_{\alpha t,\beta}(x)\,dx = 1 - \frac{\Gamma(\alpha t, L\beta)}{\Gamma(\alpha t)} \tag{9}$$

Here, α denotes the shape parameter β stands for the scale parameter. Also, the threshold degradation is represented by L. By substituting the mentioned parameters by 60, 10, and 14.5 respectively, the system reliability will be 0.97, 0.53, 0.31, and 0.22 for 4, 8, 12, and 16 weeks respectively. In order to compare the simulation model under the same circumstances as with the analytic model, we set the simulation model based on the gamma random TTF while the another feature of the model was relaxed. Running the simulation model for 25 replications, the Mann-Whitney non-parametric statistical testing was applied using the SPSS statistical package. The results of this hypothesis testing are summarized below:

```
Point estimate for ETA1-ETA2 is 0.044
95.0 Percent CI for ETA1-ETA2 is (0.032,1)
W = 10285.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is signifi-
cant at 0.9432
```

The report reveals that the mean differences between the simulated and analytic values of reliability is too small (0.044). The calculated p-value for the W-statistic is 0.9432 greater than the significant level. Consequently we could conclude the there is no reason to reject the null hypothesis. Hence the validation of the proposed simulation method is acceptable.

Once again it is emphasized that the simulation output illustrates reliability instead of availability when repair time sets to a very large amount to prevent the system from returning to the working state.

5. Case study

In order to fully reveal the applicability of the proposed simulation model, we focused on a more complicated system in the same crystalizing process in the Shahroud sugarloaf. Here, this production department was equipped with three crystallizer units. Units A and B worked serially while another one (unit C) acted as a cold standby unit and could be activated instantly. The desired performance requires proper operation of two out of three units carrying out their duties successfully. Progressively, the state of each crystalizing unit changes over working time through the state space of "Good" to "Faulty", "Imperfect" and finally "Fail". The nomenclature of this crystalizing modulus is shown in Table 3.

Based on the historical data and expert judgments, the parameters of the model estimated and presented in Table 5.

Figure 8 illustrated the model layout in the ED software. This model verified conceptually through examining logical entity flow within the networks.

As the layout presents the sugar plant has four state, Hence a Server atom is considered to model each state. Also, time to failure and time to repair for such atom has capability to define transition rate or density function. We have considered each unit A, B, and C act as a Server atom with their relevant states. Every Server has three inputs. meanwhile, in the degradation state in the sub model, one entity enters the system. Based on a given randomly distributed failure, an entity activates one of the servers A, B, or C, and accelerate its failure rate.

Table 3. Nomenclature

Nomenclature	Description
A, B, C	Crystalizing units
0, 1, 2, 3	States of each crystalizing unit, respectively stands as Good, Faulty, Imperfect and Fail
S, W, R, O	Condition of each crystalizing unit, respectively stands for Standby, Working, Repair, and Operable
$S_i\left(A_{mS}, B_{mW}, C_{nW}\right)$	The i^{th} state of the system, in which elements A, B, and C are in state m and standby, state m and working, and state n and working, respectively. $(m,n \in (0,1,2,3)$ and $i \in N$ is the size of system state space, For example state S_7:(A_{0S},B_{1W},C_{2W}) means the system is working, the unit B and C are in faulty and imperfect respectively and working while unit A is in good state and standby.
γ_{30X}	Repair rate (transition from state 3 to 0 for crystalizing unit of element X; $X \in \left(A,B,C\right)$
λ_{mnX}	Rate of transition from state m to n for crystalizing unit of element X ; $m,n \in \left(0,1,2,3\right)$, $X \in \left(A,B,C\right)$

The state transition sequence for crystalizing unit of X is shown in Figure 7. Here $X \in \left(A,B,C\right)$.

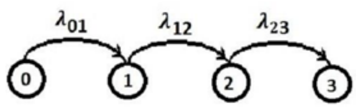


Fig. 7. State diagram for each crystalizing unit in the sugarloaf plant

Each unit is said to be working if it is in states 0, 1, or 2 and the units is said to be failed if it is in state 3. Among of all of system states, there are 54 working system states that have been listed in table 4.

The reliability becomes the probabilities that the system is in the working states and is given by:

$$R\left(t\right) = \sum_{1 \le i \le 54, i \in N} P_i\left(t\right) \tag{10}$$

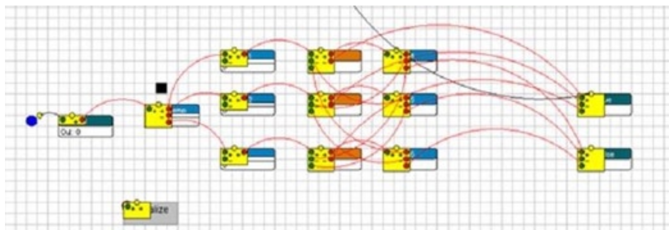


Fig. 8. The ED layout for the two four-states serially crystallization equipment with one standby

In the case of presenting a random shock, such as the Gamma distributed occurrence, for example after 100 min, a hard shock strikes the system and a unit state will change into the fail mode causing the failure of the server. Note that here the time to repair should be set to a large value to prevent system condition to return to working state after finishing the repair time. In the case of considering a small amount, the long term system availability may be calculated.

Table 4. The states in which the system acts as working

Unit A in repair or standby		Unit B in repair or standby		Unit C in repair or standby	
S_1:(A_{0S},B_{0W},C_{0W})	S_{10}:(A_{3R},B_{0W},C_{0W})	S_{19}:(A_{0W},B_{0S},C_{0W})	S_{28}:(A_{0W},B_{3R},C_{0W})	S_{37}:(A_{0W},B_{0W},C_{0S})	S_{46}:(A_{0W},B_{0W},C_{3R})
S_2:(A_{0S},B_{1W},C_{0W})	S_{11}:(A_{3R},B_{1W},C_{0W})	S_{20}:(A_{1W},B_{0S},C_{0W})	S_{29}:(A_{1W},B_{3R},C_{0W})	S_{38}:(A_{1W},B_{0W},C_{0S})	S_{47}:(A_{1W},B_{0W},C_{3R})
S_3:(A_{0S},B_{2W},C_{0W})	S_{12}:(A_{3R},B_{2W},C_{0W})	S_{21}:(A_{2W},B_{0S},C_{0W})	S_{30}:(A_{2W},B_{3R},C_{0W})	S_{39}:(A_{2W},B_{0W},C_{0S})	S_{48}:(A_{2W},B_{0W},C_{3R})
S_4:(A_{0S},B_{0W},C_{1W})	S_{13}:(A_{3R},B_{0W},C_{1W})	S_{22}:(A_{0W},B_{0S},C_{1W})	S_{31}:(A_{0W},B_{3R},C_{1W})	S_{40}:(A_{0W},B_{1W},C_{0S})	S_{49}:(A_{0W},B_{1W},C_{3R})
S_5:(A_{0S},B_{0W},C_{2W})	S_{14}:(A_{3R},B_{0W},C_{2W})	S_{23}:(A_{0W},B_{0S},C_{2W})	S_{32}:(A_{0W},B_{3R},C_{2W})	S_{41}:(A_{0W},B_{2W},C_{0S})	S_{50}:(A_{0W},B_{2W},C_{3R})
S_6:(A_{0S},B_{1W},C_{1W})	S_{15}:(A_{3R},B_{1W},C_{1W})	S_{24}:(A_{1W},B_{0S},C_{1W})	S_{33}:(A_{1W},B_{3R},C_{1W})	S_{42}:(A_{1W},B_{1W},C_{0S})	S_{51}:(A_{1W},B_{1W},C_{3R})
S_7:(A_{0S},B_{1W},C_{2W})	S_{16}:(A_{3R},B_{1W},C_{2W})	S_{25}:(A_{1W},B_{0S},C_{2W})	S_{34}:(A_{1W},B_{3R},C_{2W})	S_{43}:(A_{1W},B_{2W},C_{0S})	S_{52}:(A_{1W},B_{2W},C_{3R})
S_8:(A_{0S},B_{2W},C_{1W})	S_{17}:(A_{3R},B_{2W},C_{1W})	S_{26}:(A_{2W},B_{0S},C_{1W})	S_{35}:(A_{2W},B_{3R},C_{1W})	S_{44}:(A_{2W},B_{1W},C_{0S})	S_{53}:(A_{2W},B_{1W},C_{3R})
S_9:(A_{0S},B_{2W},C_{2W})	S_{18}:(A_{3R},B_{2W},C_{2W})	S_{27}:(A_{2W},B_{0S},C_{2W})	S_{36}:(A_{2W},B_{3R},C_{2W})	S_{45}:(A_{2W},B_{2W},C_{0S})	S_{54}:(A_{2W},B_{2W},C_{3R})

Table 5. Parameters setting of the model

Notation	Transition of States	Rate
λ_{01A}	$0 \to 1$ for unit A	0.001
λ_{12A}	$1 \to 2$ for unit A	0.002
λ_{23A}	$2 \to 3$ for unit A	0.003
λ_{01B}	$0 \to 1$ for unit B	0.002
λ_{12B}	$1 \to 2$ for unit B	0.004
λ_{23B}	$2 \to 3$ for unit B	0.009
λ_{01C}	$0 \to 1$ for unit C	0.003
λ_{12C}	$1 \to 2$ for unit C	0.006
λ_{23C}	$2 \to 3$ for unit C	0.009
γ_{30A}	$3 \to 0$ repair rates for A	0.01
γ_{30B}	$3 \to 0$ repair rates for B	0.01
γ_{30C}	$3 \to 0$ repair rates for C	0.01

Using the Experimental wizard, simulation running parameters established and the system availability chart achieved and is shown in Figure 9.

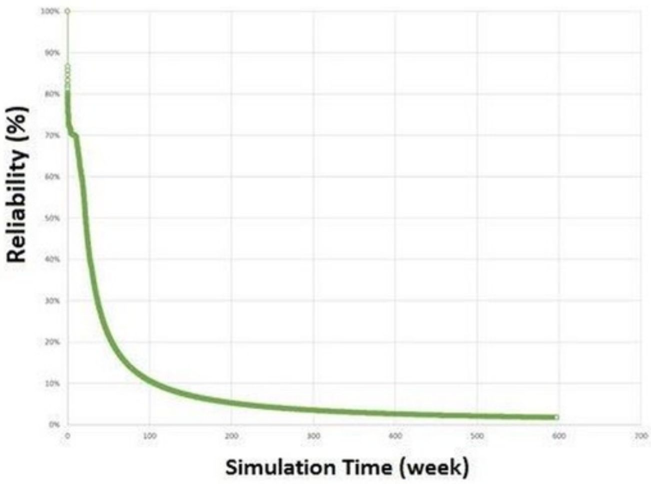


Fig. 9. Crystalizing system reliability curve

In order to look at the simulation model capability, we focused on the most critical degradation process on the crystallizations. Here their bearing displacement considered as deterioration variable and called hereinafter by D_t. Based on the historical data for a period of one cycle replacement ended to April 2018 displacement values recorded on Figure 10. As shown in the figure, 3 missing data at week 3, 19 and 31 observed. Here the maximum allowed displacement considered as 0.02 mm.

Applying curve fitting process over different alternatives patterns reveals that the crystalizing main bearing displacement sets up a parabolic curve as Eq. 11:

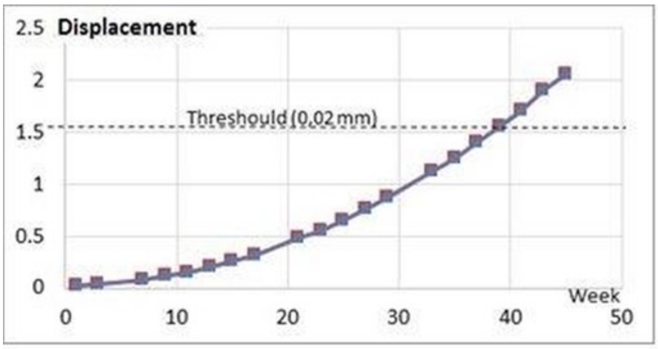


Fig. 10. Main bearing displacement in crystalizing unit (Degradation curve)

$$D_i = 0.1\,t^2 + 0.2\,t + 2.1 + e_i \tag{11}$$

where the error terms of e_i depicts model residuals that deploys from a Weibull distribution with 1 and 0.7 for its shape and scale parameters respectively. This fact shows that degradation process has Weibull distributed random process.

In order to simulate the system under such circumstance process, an extra network extended to the main simulation model, whereas entity flow in that sub-model acts as the main bearing displacement. Hence, any over amounts (displacement greater than or equal to nominated threshold) signals a breakdown event and a complete set of bearing parts including bushings, sleeve, two ended caps and four ball bearings should be replaced and after greasing calibration is required. This process simulated again 1000 times and reliability of crystalizing modulus illustrated by Figure 11. The figure also compares the reliability curves before and after considering the degradation process and reveals significant difference.

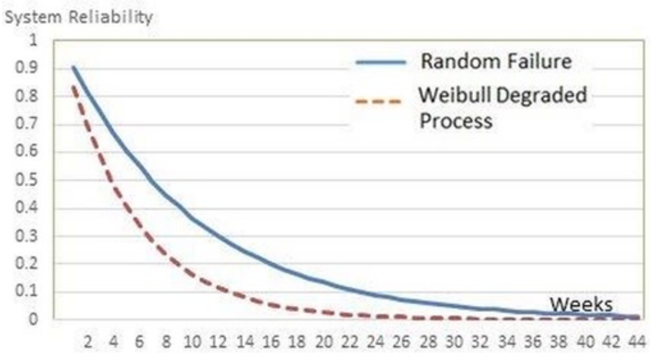


Fig. 11. Reliability comparison between before and after considering stochastic degradation

In order to examine the system reliability in the presence of random shock events due to welding operations and unusual voltage fluctuation, we applied random non-fatal shocks which would accelerate the failure process. In the simulation model, the system works as previous models up to the occurrence of a non-fatal shock; based on the expert's judgment, their mean time for arrival is on the range of [0.5 - 4] weeks with a mean of 2. Consequently, we modeled such shock arrivals by a Gamma distribution with parameters of 10 and 0.2 for the relevant shape and scale parameters, respectively. When a shock occurs, the deterioration process will rapidly be accelerated. Due to lack of reported data and by considering the experts judgment, we supposed that the failure process would accelerate by 3 times. Thus, in the simulation, we modeled the new deterioration process via Eq. 12:

$$D_t^* = 0.9t^2 + 0.6t + 2.1 + e_i \qquad (12)$$

where, the residual keeps the previous value without any changes. Figure 12 compares the reliability of the system with the non-fatal shock, and degradation process with their absence.

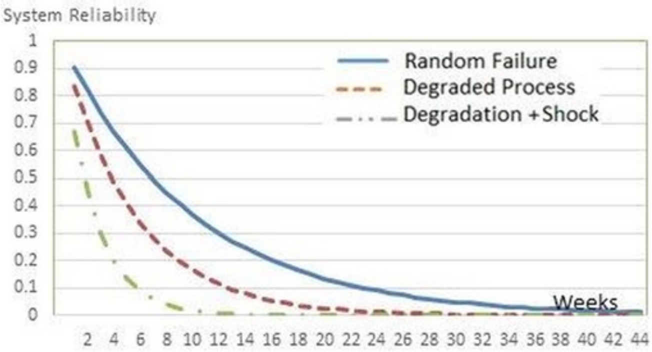


Fig. 12. Reliability comparison for crystalizing system at different failure modes

As reliability figures reveal, when the real degradation process is neglected, the system reliability will be overestimated. A similar argument can be considered when shock is also present.

6. Conclusion

The literature survey shows that system reliability estimation always accompanied by complexity in analytic methods especially when there is a great deal of uncertainties the system analysis. These uncertainties arises complicated effects and interactions in the estimations. Some usual source of uncertainty related to random time to failure, time to repair, degradation process and shock occurrences.

There is a lot of system reliability assessment methods in the literature. But almost all of them restricted to apply in the especial cases due to their relevant assumptions. For example system reliability assessment under the degradation process is a common task and sparse studies have reported on the presence of just only the Gamma continuous degradation process alongside non-fatal shock occurrences. Rottenly real cases may cover a wide variety of systems consisted of multi-component, multi-state, different type of time to failure, time to repair density function for each component, vast amount of continuous degradation functions, and their severities. Although many articles have discussed the necessity of simulating complex systems and some of its applications in specific cases have been delivered at yet, there is still a need to more address this issue.

The proposed object-oriented simulation model has a couple of advantages in comparison to the analytic and previous methods. The major advantages are as follows:

1- The method may be justified for any kind of system with complicated configurations.
2- It consists of a few individual moduli where each of them may be relaxed for more simplified cases.
3- It supports both continuous or discrete degradation deterministic or stochastic processes. Thus, it could be applied for multi-state multi-component systems.
4- The model has no dependence on special random variables such as exponential or gamma degradation. Any practitioner could set them to any other well-known density functions (e.g. Weibull, Logistic, Beta).
5- It covers the effects of all types of random fatal and non-fatal shocks with any severities.
6- The simulation model reports the system availability as well as reliability.

The proposed model has been established using ED simulation software capabilities which may be accounted as a disadvantage. Nevertheless, non-familiar simulation experts could follow a specific logic for implementation via other software applications. Modeling some auto-correlated degradation processes in this context will be remain as future research for interested researchers.

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 & System Safety 2017; 157: 129-138, https://doi.org/10.1016/j.ress.2016.08.025.
2. Angskun T, Bosilca G, Fagg G, Pje J, Dongarra JJ. Reliability analysis of self-healing network using discrete-event simulation. Seventh IEEE International Symposium on Cluster Computing and the Grid (CCGrid '07). Washington DC. USA 2007, https://doi.org/10.1109/CCGRID.2007.95.
3. Caballé N, Castro I. Analysis of the reliability and the maintenance cost for finite life cycle systems subject to degradation and shocks. Applied Mathematical Modelling 2017; 52: 731-746, https://doi.org/10.1016/j.apm.2017.08.002.
4. Che H, Zeng Sh, Guo J. Reliability Analysis of Load-Sharing Systems Subject to Dependent Degradation Processes and Random Shocks. IEEE Access 2017; 5: 23395 - 23404, https://doi.org/10.1109/ACCESS.2017.2762727.
5. Eryilmaz S. On the lifetime behavior of a discrete time shock model. Journal of Computational and Applied Mathematics 2013; 237(1): 384-388, https://doi.org/10.1016/j.cam.2012.06.008.
6. Fan M, Zeng Z, Zio E, Kang R. Modeling dependent competing failure processes with degradation-shock dependence. Reliability Engineering & System Safety 2017; 165: 422-430, https://doi.org/10.1016/j.ress.2017.05.004.
7. Finkelstein Maxim, Cha Ji Hwan. Stochastic Modeling for Reliability- Shocks. Burn-in and Heterogeneous Populations. Springer-Verlag London. 2013, https://doi.org/10.1007/978-1-4471-5028-2_8.
8. Gola A. Reliability Analysis of Reconfigurable manufacturing System Structures Using Computer Simulation Method. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21(1): 90-102, https://doi.org/10.17531/ein.2019.1.11.
9. Hao S, Yang J, Ma X, Zhao Y. Reliability modeling for mutually dependent competing failure processes due to degradation and random shocks. Applied Mathematical Modelling 2017; 51: 232-249, https://doi.org/10.1016/j.apm.2017.06.014.
10. Huang Wenping, Zhou Jinglun, Ning Juhong. A Condition Based Maintenance for System Subject to Competing Failure due to Degradation and Shock. IAENG International Journal of Applied Mathematics 2016; 46(2): 223-230.
11. Jasiulewicz-Kaczmarek M, Gola A. Maintenance 4.0 Technologies for Sustainable Manufacturing - an Overview. IFAC PapersOnLine 52-10, 2019; 91-96, https://doi.org/10.1016/j.ifacol.2019.10.005.

12. Juan AA, Ferrer A, Serrat C, Monteforte A, Faulin J. Applications of discrete-event simulation to reliability and availability assessment in civil engineering structures. Simulation Conference (WSC). Proceedings of the Winter: IEEE 2009; 11179210, https://doi.org/10.1109/WSC.2009.5429250.
13. Kosicka E, Kozlowski E, Mazurkiewicz D. Intelligent systems of forecasting the failure of machinery park and supporting fulfilment of orders of spare parts. Advances in Intelligent Systems and Computing 2018; 637: 54-63, https://doi.org/10.1007/978-3-319-64465-3_6.
14. Liu Z, Ma X, Shen L, Zhao Y. Degradation-shock-based Reliability Models for Fault-tolerant Systems. Quality and Reliability Engineering International 2015; 32(3): 949-955, https://doi.org/10.1002/qre.1805.
15. Montoro-Cazorla D, Pérez-Ocón R. A reliability system under cumulative shocks governed by a BMAP. Applied Mathematical Modelling 2015; 39(23-24): 7620-7629, https://doi.org/10.1016/j.apm.2015.03.066.
16. Nikulin M.S., Limnios N., Balakrishnan N., Kahle W., Huber-Carol C. Advances in Degradation Modeling- Applications to Reliability, Survival Analysis, and Finance. Birkhauser Boston, a part of Springer Science + Business Media LLC 2010, https://doi.org/10.1007/978-0-8176-4924-1.
17. Patelli E, Feng G, Coolen F.PA, Coolen-Maturi T. Simulation methods for system reliability using the survival signature. Reliability Engineering and System Safety 2017; 167: 327-337, https://doi.org/10.1016/j.ress.2017.06.018.
18. Rafiee K, Feng Q, Coit DW. Reliability assessment of competing risks with generalized mixed shock models. Reliability Engineering & System Safety 2016; 159: 1-11, https://doi.org/10.1016/j.ress.2016.10.006.
19. Riascos-Ochoa J, Sanchez-Silva M, Akhavan-Tabatabaei R. Reliability analysis of shock-based deterioration using phase-type distributions. Probabilistic Engineering Mechanics 2014; 38: 88-101, https://doi.org/10.1016/j.probengmech.2014.09.004.
20. Sobaszek Ł, Gola A, Kozłowski E. Application of survival function in robust scheduling of production jobs. Proceedings of the Federated Conference on Computer Science and Information Systems, IEEE, New York 2017; 575-578, https://doi.org/10.15439/2017F276.
21. Vaisi B, Farughi H, Raissi S. Two - Machine Robotic Cell Sequencing Under different Uncertainties .International Journal of Simulation Modeling 2018; 17(2): 284-294, https://doi.org/10.2507/IJSIMM17(2)434.
22. Wang C, Zhang H, Li Q. Reliability assessment of aging structures subjected to gradual and shock deteriorations. Reliability Engineering & System Safety 2017; 161: 78-86, https://doi.org/10.1016/j.ress.2017.01.014.
23. Warrington L, Jones JA. Representing complex systems within discrete event simulation for reliability assessment. Annual Reliability and Maintainability Symposium. IEEE, 2003, http//doi.org/10.1109/RAMS.2003.1182037.
24. Wenjie D, Sifeng L, Xiaoyu Y, Zhigeng F. Balancing Reliability and maintenance Cost Rate of multi state components with fault interval omission. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2019; 21 (1): 37-45, https://doi.org/10.17531/ein.2019.1.5.
25. Zaim S, Turkyılmaz A, Acar MF, Al-Turki U, Demirel O. Maintenance strategy selection using AHP and ANP algorithms: a case study. Journal of Quality in Maintenance Engineering 2012; 18(1): 16-29, https://doi.org/10.1108/13552511211226166.
26. Zhang W, Jiang Sh, Li X, Wang Z. An approach to structural reliability evaluation under fatigue degradation and shocks. Mechanical Systems and Signal Processing 2018; 113: 65-76, https://doi.org/10.1016/j.ymssp.2017.06.006.

Mohammad Reza POURHASSAN
Sadigh RAISSI
Ashkan HAFEZALKOTOB
School of Industrial Engineering
Islamic Azad University, South Tehran Branch
Tehran, Iran

E-mails: st_m_poorhassan@azad.ac.ir, Raissi@azad.ac.ir, A_hafez@azad.ac.ir

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, Poland

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. 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. Marianna Jacyna
Warsaw University of Technology, Warsaw, Poland

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

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 2018 Impact Factor is 1.806

Editorial staff:	Dariusz Mazurkiewicz, PhD, DSc (Eng), Associate Professor (Editor-in-Chief, Secretary of the Editorial Board) Tomasz Klepka, PhD, DSc (Eng), Associate Professor (Deputy Editor-in-Chief) Teresa Błachnio-Krolopp, MSc (Eng) (Editorial secretary) Andrzej Koma (Typesetting and text makeup) Krzysztof Olszewski, PhD (Eng) (Webmaster)
Publisher:	Polish Maintenance Society, Warsaw
Scientific patronage:	Polish Academy of Sciences Branch in Lublin
Address for correspondence:	"Eksploatacja i Niezawodność" – Editorial Office ul. Nadbystrzycka 36, 20-618 Lublin, Poland e-mail: office@ein.org.pl http://www.ein.org.pl/
Circulation:	550 copies

INFORMATION FOR AUTHORS

Eksploatacja i Niezawodnosc – Maintenance and Reliability – the journal of the Polish Maintenance Society, under the scientific supervision of the Polish Academy of Sciences (Branch in Lublin), published four times a year.

The scope of the Quarterly

The quarterly *Eksploatacja i Niezawodnosc – Maintenance and Reliability* publishes articles containing original results of experimental research on the durability and reliability of technical objects. We also accept papers presenting theoretical analyses supported by physical interpretation of causes or ones that have been verified empirically. *Eksploatacja i Niezawodność – Maintenance and Reliability* also publishes articles on innovative modeling approaches and research methods regarding the durability and reliability of objects.

The following research areas are particularly relevant to the journal:

1. degradation processes of mechanical and biomechanical systems,
2. diagnosis and prognosis of operational malfunctions and failures.
3. analysis of failure risk/wear,
4. reliability-and-environmental-safety engineering in the design, manufacturing and maintenance of objects,
5. management and rationalization of object maintenance,
6. risk management in the processes of operation and maintenance,
7. the human factor and human reliability in operation and maintenance systems.

Terms and Conditions of Publication

The quarterly *Eksploatacja i Niezawodnosc – Maintenance and Reliability* publishes only original papers written in English or in Polish with an English translation. Translation into English is done by the Authors after they have received information from the Editorial Office about the outcome of the review process and have introduced the necessary modifications in accordance with the suggestions of the referees! Acceptance of papers for publication is based on two independent reviews commissioned by the Editor.
The publication fee for one text is 320 EUR (all graphics in color included).

The quarterly *Eksploatacja i Niezawodnosc – Maintenance and Reliability* proceeds entirely online at submission.ein.org.pl

Technical requirements

- After receiving positive reviews and after acceptance of the paper for publication, the text must be submitted in a Microsoft Word document format.
- Drawings and photos should be additionally submitted in the form of high resolution separate graphical files in the TIFF, SVG, AI or JPG formats.
- A manuscript should include: names of authors, title, abstract, and key words that should complement the title and abstract (in Polish and in English), the text in Polish and in English with a clear division into sections (please, do not divide words in the text); tables, drawings, graphs, and photos included in the text should have descriptive two-language captions, if this can be avoided, no formulae and symbols should be inserted into text paragraphs by means of a formula editor; references (written in accordance with the required reference format); author data – first names and surnames along with scientific titles, affiliation, address, phone number, fax, and e-mail address.

The Editor reserves the right to abridge and adjust the manuscripts. All submissions should be accompanied by a submission form.

Detailed instructions to Authors, including evaluation criteria can be found on the journal's website: www.ein.org.pl

Editor contact info

Editorial Office of „Eksploatacja i Niezawodnosc – Maintenance and Reliability"
Nadbystrzycka 36, 20-618 Lublin, Poland
e-mail: office@ein.org.pl

INFORMATION FOR SUBSCRIBERS

Fees

Yearly subscription fee (four issues) is 100 złoty and includes delivery costs. Subscribers receive any additional special issues published during their year of subscription free of charge.

Orders

Subscription orders along with authorization to issue a VAT invoice without receiver's signature should be sent to the Editor's address.

Note

In accordance with the requirements of citation databases, proper citation of publications appearing in our Quarterly should include the full name of the journal in Polish and English without Polish diacritical marks, i.e.,

Eksploatacja i Niezawodnosc – Maintenance and Reliability.

No text or photograph published in „Maintenance and Reliability" can be reproduced without the Editor's written consent.

Edward Kozłowski

Lublin University of Technology
Faculty of Management
ul. Nadbystrzycka 38, 20-618 Lublin, Poland
E-mail: e.kozlovski@pollub.pl

Anna Borucka

Military University Of Technology
ul. gen. Sylwestra Kaliskiego 2,
01-476, Warsaw, Poland
E-mail: anna.borucka@wat.edu.pl

Andrzej Świderski

Motor Transport Institute
ul. Jagiellońska 80,
03-301 Warsaw, Poland
E-mail: andrzej.swiderski@its.waw.pl

ZASTOSOWANIE REGRESJI LOGISTYCZNEJ DO WYZNACZANIA MACIERZY PRAWDOPODOBIEŃSTW PRZEJŚĆ STANÓW EKSPLOATACYJNYCH W SYSTEMACH TRANSPORTOWYCH

Przedsiębiorstwa transportowe mogą być traktowane jako wyodrębniony pod względem technicznym, organizacyjnym, ekonomicznym i prawnym system transportowy. Zachowanie jakości i ciągłości realizacji zleceń przewozowych wymaga wysokiego poziomu gotowości pojazdów oraz personelu (szczególnie kierowców). Kontrolowanie i sterowanie realizowanymi zadaniami wspierane jest modelami matematycznymi, umożliwiającymi ocenę i określenie strategii dotyczącej podejmowanych działań. Wsparcie procesów zarządzania polega głównie na analizie sekwencji kolejnych, realizowanych czynności (stanów). W wielu przypadkach taki ciąg czynności jest modelowany za pomocą procesów stochastycznych, spełniających własność Markowa. Ich klasyczne zastosowanie możliwe jest tylko w przypadku, gdy warunkowe rozkłady prawdopodobieństwa przyszłych stanów są określone wyłącznie przez bieżący stan eksploatacyjny. Identyfikacja takiego procesu stochastycznego polega głównie na wyznaczeniu macierzy prawdopodobieństw przejść międzystanowych.

Niestety w wielu przypadkach analizowane ciągi czynności nie spełniają własności Markowa. Dodatkowo, na wystąpienie kolejnego stanu wpływa długość interwału czasowego pozostania systemu w określonym stanie eksploatacyjnym. W artykule przedstawiono metodę konstrukcji macierzy prawdopodobieństw przejść pomiędzy stanami eksploatacyjnymi. Wartości tej macierzy zależą od czasu przebywania obiektu w danym stanie. Celem artykułu było zaprezentowanie alternatywnej metody estymacji parametrów tej macierzy w sytuacji, gdy badany szereg nie spełnia własności Markowa. Wykorzystano w tym celu regresję logistyczną.

Słowa kluczowe: *regresja logistyczna, macierz prawdopodobieństw przejść, łańcuchy Markowa, system transportowy*

1. Wprowadzenie

Pojęcie systemu transportowego, za Grzywaczem i Burnewiczem [17] oraz Andrzejczakiem [3] postrzegane jest w niniejszym artykule jako uporządkowany układ trzech podsystemów: technicznego, organizacyjnego i ekonomiczno-prawnego, tworzący logiczną, zrównoważoną wewnętrznie całość, umożliwiającą osiągnięcie konkretnego celu. Pozwala to określić mianem systemu transportowego analizowane przedsiębiorstwo, a za cel jego funkcjonowania przyjąć realizację zadań przewozowych.

Systemy transportowe mogą być analizowane jako wielostanowe sekwencje kolejnych planowanych i nieplanowanych czynności obsługowych dokonywanych przez operatora systemu transportowego [27]. Konstrukcja modeli, które je opisują oraz pozwalają na wyznaczenie prognozy stanu użytkowanego obiektu , umożliwia planowanie strategii obsługiwania i sterowanie gotowością parku maszyn [9, 21,], pojazdów [13, 26] itp. Modelowanie funkcjonowania obiektów technicznych, za pomocą modeli deterministycznych nie zawsze jest możliwe, ponieważ na wyniki (realizacje) wpływają zaburzenia zewnętrzne (czynniki losowe), które uniemożliwiają dokładne przewidywanie kolejnych stanów. W takich przypadkach zachowanie systemów technicznych modelujemy za pomocą metod probabilistycznych, a w szczególności procesów stochastycznych. Ważną klasę procesów stochastycznych stanowią procesy Markowa. Pewne możliwości zastosowań tych procesów przedstawiono w pracach [7, 24]. Podstawą ich użycia jest spełnienie własności Markowa: przyszłość nie zależy od przeszłości, gdy znana jest teraźniejszość. W wielu analizach zakłada się a priori, że szereg czasowy spełnia tę własność i jej weryfikacja jest pomijana [55]. Jedynie nieliczni autorzy wskazują na konieczność jej sprawdzania [53] i eliminują przypadki które nie osiągnęły tego założenia [47]. Alternatywą dla systemów niespełniających własności Markowa są klasyczne metody niezawodnościowe, pozwalające na wyznaczenie empirycznych charakterystyk takich jak: strumień odnowy, funkcja odnowy, czas do następnego uszkodzenia czy intensywność strumienia odnowy [6, 14, 35] oraz obliczanie na ich podstawie głównych miar oceny systemu [15, 23, 42]. W literaturze, w ramach podobnych badań, prezentowane są różne modele [30, 43], w tym semi - Markowa [7, 24] a także wykorzystujące sztuczne sieci neuronowe [10, 33], algorytm faktoryzacji [31], drzewa błędów [52] czy modele niezawodnościowe [42, 36].

W wielu publikacjach nie jest brany pod uwagę czas pozostania w określonym stanie eksploatacyjnym. Niejednorodność interwału czasowego pomiędzy kolejnymi stanami również może powodować niespełnienie własności Markowa. W niniejszym artykule do oszacowania warunkowych prawdopodobieństw przebywania obiektu badań w poszczególnych stanach eksploatacyjnych zastosowano regresję logistyczną [49]. Regresja logistyczna opisuje związek między zmienną jakościową, a jedną lub większą liczbą zmiennych predykcyjnych [25, 46]. W literaturze regresja logistyczna stosowana jest w medycynie [5, 44], w tomografii komputerowej [46], do identyfikacji systemów technicznych [25], w obszarze finansów przedsiębiorstw [39, 54], bankowości [1, 34] i szeroko pojętych inwestycji [12, 29] oraz wykorzystywana jest do szacowania poziomu ryzyka [2, 8, 45], w badaniach społecznych, demograficznych [4, 41] i innych [25, 40]. W odniesieniu do systemów transportowych modele regresji logistycznej proponowane są przede wszystkim do oceny zagrożenia na drodze wynikającego z wypadków drogowych [18, 37], kształtowania wyborów trasy w sieci transportowej [32, 49] czy analizy wpływu wybranych czynników na realizację procesów przewozowych [38, 48].

W pracy pokazano istnienie zależności pomiędzy czasem trwania stanu eksploatacyjnego, a wartością prawdopodobieństwa przejścia do kolejnego stanu. Celem dokładnej i wyczerpującej analizy problemu w pierwszej kolejności dokonano wprowadzenia, prezentującego metody matematyczne przywoływane w artykule. W rozdziale drugim przedstawiono definicje dotyczące łańcuchów Markowa i sposób weryfikacji własności Markowa. W rozdziale trzecim zaprezentowano metodę estymacji macierzy prawdopodobieństwa przejść za pomocą regresji logistycznej. Następnie zaprezentowano przykład wdrożenia proponowanej metody z wykorzystaniem danych empirycznych dla wybranego środka transportu, realizującego zadania przewozowe w ramach systemu transportowego (przedsiębiorstwa). W końcowym etapie dokonano omówienia otrzymanych wyników, podsumowania przeprowadzonych analiz i wskazano kierunki dalszych badań.

2. Łańcuchy Markowa

Stanem obiektu określa się jego charakterystyczną cechę, właściwość techniczną, która przyporządkowuje go do danego systemu eksploatacji [50], jest to wektor, którego składowymi są wartości fizyczne, opisujące obiekt z punktu widzenia danego badania [28]. W literaturze stan obiektu technicznego definiowany jest jako rezultat jednego i tylko jednego zdarzenia w ciągu doświadczeń spośród skończonego lub przeliczalnego zbioru parami wyłączających się zdarzeń [11, 54]. Do analizy systemów technicznych wykorzystujemy narzędzia rachunku prawdopodobieństwa i statystyki matematycznej [22, 53]. Niech $(\Omega, \mathcal{F}, P)$ będzie przestrzenią probabilistyczną, N – zbiorem liczb naturalnych, S – przestrzenią stanów analizowanego zjawiska.

Definicja 1 Rodzinę $\{X_t\}_{t \in N}$ zmiennych losowych $X_t: \Omega \to S$ dla dowolnego $t \in N$ nazywamy procesem stochastycznym z czasem dyskretnym [51, 54].

W pracy analizujemy stany eksploatacyjne, w których przebywają pojazdy. Zbiór stanów eksploatacyjnych S jest to zbiór wartości procesu stochastycznego $\{X_t\}_{t \in N}$. W każdej chwili $t \in N$ obiekt znajduje się w jednym z możliwych stanów oraz $X_t(\omega) = x_t$, tzn. w przypadku zajścia zdarzenia losowego ω w momencie t, system znajduje się w stanie $x_t \in S$. W prowadzonych badaniach przyjmujemy, że zbiór stanów S jest zbiorem skończonym oraz $S = \{s_1, s_2, \ldots, s_k\}$, $k \in N$, $2 \leq k < \infty$. Wielkość $P(X_t = s_i) = p_i(t)$ oznacza prawdopodobieństwo, że system w momencie $t \in N$ znajduje się w stanie $s_i \in S$, $1 \leq i \leq k$ oraz $\sum_{i=1}^{k} p_i(t) = 1$.

Definicja 2 Proces stochastyczny $\{X_t\}_{t \in N}$ z czasem dyskretnym nazywa się łańcuchem Markowa jeżeli dla każdego $n \in N$, dowolnych momentów $t_1, t_2, \ldots, t_n \in N$ spełniających warunek $t_1 < t_2 < \cdots < t_n$ oraz dowolnych $x_1, x_2, \ldots, x_n \in S$, zachodzi równość [26, 47]:

$$P(X_{t_n} = x_n | X_{t_{n-1}} = x_{n-1}, X_{t_{n-2}} = x_{n-2}, \ldots, X_{t_1} = x_1) = P(X_{t_n} = x_n | X_{t_{n-1}} = x_{n-1}) \quad (1)$$

Z definicji łańcucha Markowa wynika, że rozkład warunkowy zmiennej losowej X_n dla danych wartości $X_{t_0}, X_{t_1}, \ldots, X_{t_{n-1}}$ zależy tylko od ostatniej znanej wartości $X_{t_{n-1}}$. Zazwyczaj zakłada się, że odstępy pomiędzy t_i i t_{i+1} są jednakowe [16]. Poniżej przyjmujemy, że $t_n = n \in N$. Jeżeli $\{X_t\}_{t \in N}$ jest niejednorodnym łańcuchem Markowa, to dla dowolnego $t \in N$ oraz $1 \leq i, j \leq k$, wielkość:

$$P\left(X_t = s_j \middle| X_{t-1} = s_i\right) = p_{ij}(t) \quad (2)$$

nazywamy prawdopodobieństwa przejścia ze stanu s_i w momencie $t-1$ do stanu s_j w momencie t. Zatem dla łańcuchów spełniających własność Markowa (1) warunkowe rozkłady prawdopodobieństwa przyszłych stanów procesu są zdeterminowane wyłącznie przez jego bieżący stan oraz moment t, bez względu na przeszłość (są warunkowo niezależne od stanów przeszłych). Macierz $P(t) = \left[p_{ij}(t)\right]_{1 \leq i,j \leq k}$ spełniającą warunek $\sum_{j=1}^{k} p_{ij}(t) = 1$ dla $t \in N$ oraz $1 \leq i \leq k$ nazywamy macierzą prawdopodobieństwa przejść w jednym kroku w chwili t [7, 47, 51].

Definicja 3 Łańcuch Markowa $\{X_t\}_{t \in N}$ jest jednorodnym łańcuchem Markowa, jeżeli prawdopodobieństwa przejścia $p_{ij}(t)$ nie zależą od momentu $t \in N$.

Zatem dla jednorodnego łańcucha Markowa $p_{ij}(t) = p_{ij}$ dla $1 \leq i,j \leq k$ oraz dowolnego momentu $t \in N$, natomiast macierz $P = \left[p_{ij}\right]_{1\leq i,j\leq k}$ spełniającą warunek $\sum_{j=1}^{k} p_{ij} = 1$, $1 \leq i \leq k$ nazywamy macierzą prawdopodobieństwa przejścia w jednym kroku. Dla jednorodnego łańcucha Markowa prawdopodobieństwa przejścia ze stanu s_i w momencie t do stanu s_j w momencie $t+n$ wyznaczamy ze wzoru [13, 24]:

$$P\left(X_{t+n} = s_j \middle| X_t = s_i\right) = p_{ij}^{(n)} \tag{3}$$

gdzie $\left[p_{ij}^{(n)}\right]_{1\leq i,j\leq k} = P^n$, $n \in N$ jest macierzą prawdopodobieństwa przejścia w n krokach.

Definicja 4 Jeżeli $\{X_t\}_{t\in N}$ jest jednorodnym łańcuchem Markowa oraz istnieje rozkład $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$, gdzie $\pi_i \geq 0$, $1 \leq i \leq k$ oraz $\sum_{i=1}^{k} \pi_i = 1$, spełniający równanie:

$$\pi P = \pi \tag{4}$$

to rozkład π nazywamy rozkładem stacjonarnym jednorodnego łańcucha Markowa.
Własność stacjonarności oznacza, że jeżeli w pewnym momencie $n \in N$ łańcuch osiągnie rozkład stacjonarny, to dla każdej kolejnej chwili większej od n rozkład pozostanie taki sam. Rozkład stacjonarny wyznaczamy rozwiązując układ równań [16, 22]:

$$\sum_{j=1}^{k} \pi_j \cdot p_{ij} = \pi_i \tag{5}$$

$$\sum_{i=1}^{k} \pi_i = 1 \tag{6}$$

oraz $\pi_i \geq 0$ dla $1 \leq i \leq k$.
Ważną rolę w badaniu procesów, przy wykorzystaniu łańcuchów Markowa, pełnią jego własności graniczne, a szczególnie granice prawdopodobieństw $p_j(n)$ oraz $p_{ij}^{(n)}$ przy $n \to \infty$, które opisują probabilistyczne zachowanie procesu po długim czasie [16, 22].

Twierdzenie 1 (ergodyczne) Niech $\{X_t\}_{t\in N}$ będzie jednorodnym łańcuchem Markowa o skończonej liczbie stanów $k < \infty$ $(k = \#S = \#\{i: s_i \in S\})$, to:

a. istnieje wektor $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$, taki że $\pi_i > 0$ dla $1 \leq i \leq k$ oraz $\sum_{i=1}^{k} \pi_i = 1$;
b. dla dowolnych $1 \leq i,j \leq k$

$\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$;

c. wektor π jest rozwiązaniem równania (6).

Poniżej przedstawiony zostanie sposób oszacowania macierzy prawdopodobieństwa przejścia dla jednorodnego łańcucha Markowa oraz sposób weryfikacji własności Markowa. Niech $\{x_t\}_{0\leq t\leq n}$ będzie realizacją łańcucha Markowa. Wielkość $n_i = \#\{t: x_t = s_i, 0 \leq t \leq n\}$ oznacza liczbę momentów, dla których system przebywał w stanie s_i dla $1 \leq i \leq k$, gdzie $\sum_{i=1}^{k} n_i = n$, natomiast wielkość $n_{ij} = \#\{t: x_t = s_i, x_{t+1} = s_j, 0 \leq t \leq n-1\}$ oznacza liczbę przejść ze stanu s_i do stanu s_j dla $1 \leq i,j \leq k$ oraz $\sum_{j=1}^{k} n_{ij} = n_i$. Przy założeniu, że spełniona jest własność Markowa, szacujemy macierz prawdopodobieństwa przejść. Estymator przejścia ze stanu s_i do stanu s_j wyznaczamy ze wzoru $\hat{p}_{ij} = \frac{n_{ij}}{n_i}$ dla $1 \leq i,j \leq k$.

W celu weryfikacji własności Markowa stosujemy test zgodności χ^2. Na poziomie istotności $\alpha, \alpha \in (0,1)$ tworzymy hipotezę roboczą:

$H_0: P(X_t = x | X_{t-1} = y, X_{t-2} = z) = P(X_t = x | X_{t-1} = y)$ (łańcuch $\{X_t\}_{t\in N}$ spełnia własność Markowa)

oraz hipotezę alternatywną:

$H_1: P(X_t = x | X_{t-1} = y, X_{t-2} = z) \neq P(X_t = x | X_{t-1} = y)$ (łańcuch $\{X_t\}_{t\in N}$ nie spełnia własności Markowa),

gdzie $x, y, z \in S$. Za miarę rozbieżności pomiędzy rozkładami $P(X_t = x | X_{t-1} = y, X_{t-2} = z)$ oraz $P(X_t = x | X_{t-1} = y)$ wybieramy statystykę testową:

$$\chi_e^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{v=1}^{k} \frac{(n_{ijv} - n_{ij}\hat{p}_{jv})^2}{n_{ij}\hat{p}_{jv}} \tag{7}$$

która ma rozkład χ^2 o k^3 stopniach swobody.
Wielkość $n_{ijv} = \#\{t: x_t = s_i, x_{t+1} = s_j, x_{t+2} = s_v, 0 \leq t \leq n-2\}$ oznacza liczbę przejść ze stanu s_i do stanu s_j, a następnie do stanu s_v dla $1 \leq i, j, v \leq k$. Z tablic dla rozkładu χ^2 o k^3 stopniach swobody wyznaczamy kwantyl rzędu $1-\alpha$, który oznaczamy jako $\chi^2(1-\alpha, k^3)$. Jeżeli $\chi_e^2 < \chi^2(1-\alpha, k^3)$, to na poziomie istotności α nie ma podstaw do odrzucenia hipotezy roboczej H_0, zatem przyjmujemy, że łańcuch $\{X_t\}_{t\in N}$ spełnia własność Markowa. Jeżeli natomiast $\chi_e^2 \geq \chi^2(1-\alpha, k^3)$, to na poziomie istotności α hipotezę roboczą H_0 odrzucamy na korzyść hipotezy alternatywnej, zatem łańcuch $\{X_t\}_{t\in N}$ nie spełnia własności Markowa.

3. Regresja logistyczna

W wielu przypadkach proces stochastyczny $\{X_t\}_{t\in N}$ nie spełnia własności Markowa. Realizacja procesu $\{X_t\}_{t\in N}$ zależy od dodatkowych czynników. W systemach transportowych, logistycznych czas pozostania w określonym stanie bezpośrednio wpływa na prawdopodobieństwo przejścia do stanów pozostałych. Poniżej autorzy zastosowali regresję logistyczną do zdefiniowania macierzy prawdopodobieństwa przejścia, która zależy od czasu przebywania obiektu w danym stanie. W rozważanym przypadku, zmienna losowa $X_t, t \in N$ opisująca stan systemu, może przyjąć k możliwych realizacji. Ponieważ rozważamy momenty, dla których stan systemu zmienia się, to jeżeli w momencie $t \in N$ system znajdował się w stanie $s_i \in S$, to w momencie $t+\tau$ system może przebywać w stanach $S\backslash\{s_i\}$ (zmienna losowa $X_{t+\tau}$ może przyjąć $k-1$ możliwych realizacji). Wyznaczenie wartości prawdopodobieństw przejść umożliwia wielomianowa regresja logistyczna [19, 25, 44, 46]. Jeden z poziomów należy przyjąć jako odniesienie. Dla każdego stanu $s_i \in S$, $1 \leq i \leq k$ wyznaczamy prawdopodobieństwa przejścia do pozostałych stanów:

$$P\left(X_{t+\tau} = s_j \middle| X_t = s_i\right) = p_{ij}(\tau) \tag{8}$$

gdzie $t, \tau \in N$ oraz $s_j \in S\backslash\{s_i\}$. Ze zbioru $S\backslash\{s_i\}$ wybieramy stan odniesienia $s_q \in S\backslash\{s_i\}$ oraz wyznaczamy logarytmy szans dla pozostałych stanów:

$$ln\, \frac{P\left(X_{t+\tau} = s_j \middle| X_t = s_i\right)}{P\left(X_{t+\tau} = s_q \middle| X_t = s_i\right)} = \beta_{ij}^0 + \beta_{ij}^1 \tau \tag{9}$$

dla $s_j \in S\backslash\{s_i, s_q\}$. Wartości parametrów strukturalnych w modelu (9) wyznaczamy korzystając z metody największej wiarygodności [20, 25, 49]. Do oceny istotności parametrów modelu stosuje się test Walda.
Prawdopodobieństwa przejścia dla stanów $s_j \in S\backslash\{s_i, s_q\}$ wyznaczamy ze wzoru:

$$p_{ij}(\tau) = \frac{e^{\beta_{ij}^0 + \beta_{ij}^1 \tau}}{1 + \sum_{\substack{1 \leq v \leq k \\ v \neq i \\ v \neq q}} e^{\beta_{iv}^0 + \beta_{iv}^1 \tau}} \tag{10}$$

natomiast dla stanu odniesienia s_q prawdopodobieństwo wynosi:

$$p_{iq}(\tau) = \frac{1}{1 + \sum_{\substack{1 \leq v \leq k \\ v \neq i \\ v \neq q}} e^{\beta_{iv}^0 + \beta_{iv}^1 \tau}} \tag{11}$$

Ze wzorów (10)-(11) otrzymujemy, że logarytm ilorazu szans dla dwóch dowolnych poziomów $s_j, s_v \in S\backslash\{s_i, s_q\}$ jest równy:

$$ln\frac{P(X_{t+\tau}=s_j|X_t=s_i)}{P(X_{t+\tau}=s_v|X_t=s_i)} = ln\frac{p_{ij}(\tau)}{p_{iv}(\tau)} = (\beta_{ij}^0 - \beta_{iv}^0) + (\beta_{ij}^1 - \beta_{iv}^1)\tau. \qquad (12)$$

4. Estymacja macierzy prawdopodobieństwa przejść dla wybranego środka transportu

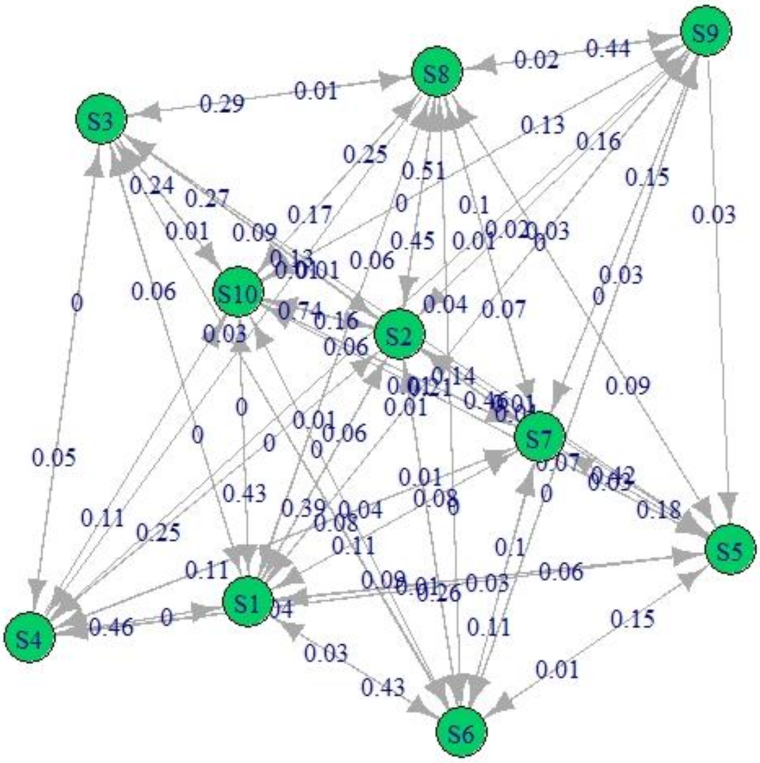
Podmiotem badania był belgijski oddział dystrybucji działający na rzecz sieci hipermarketów. Usługi transportowe realizowane są codziennie, 24 godziny na dobę, dlatego istotne jest właściwe harmonogramowanie przewozów, uwzględniające dostępność zatrudnionego personelu (szczególnie kierowców), a także gotowość pojazdów.

W badaniu wykorzystano dane pochodzące z funkcjonującego w przedsiębiorstwie systemu informatycznego do zarządzania flotą, który integruje, przetwarza i archiwizuje odczyty z zamontowanego na pojeździe nadajnika GPS, tachografu, magistrali CAN (*Controller Area Network*) oraz komputera pokładowego. Pozwala to na pozyskanie danych dotyczących kierowcy i pojazdu w czasie rzeczywistym, umożliwia śledzenie pozycji i ruchu samochodów, wizualizację lokalizacji pojazdów i naczep na mapie, monitoring czasu jazdy i odpoczynku itp. Informacje dotyczyły 69 pojazdów ciężarowych marki Iveco Stralis EEV 460. Zgromadzone dane uporządkowano oraz analizowano 10 stanów eksploatacyjnych realizowanych przez pojazdy ciężarowe. Czynności te wyszczególniono w tab. 1.

Tabela 1. Wyróżnione w badaniu stany eksploatacyjne

Lp.	Nazwa stanu eksploatacyjnego
S1	Dyspozycyjność
S2	Jazda
S3	Czynności manipulacyjne
S4	Naprawa
S5	Obsługa
S6	Parkowanie
S7	Postój
S8	Wyładunek
S9	Tankowanie
S10	Załadunek

Zaprezentowane w artykule badanie przeprowadzono dla jednego, losowo wybranego pojazdu. Dokonano sprawdzenia własności Markowa. Wykorzystano w tym celu test χ^2. Statystyka testu wyniosła $2672{,}74$, a $p - value = 2.2 * 10^{-16}$. Oznacza to, że na poziomie istotności $\alpha = 0.001$ hipotezę roboczą należy odrzucić na korzyść hipotezy alternatywnej, zatem analizowany proces stochastyczny nie spełnia własności Markowa. Niemniej jednak oszacowano (w celach porównawczych) macierz prawdopodobieństw przejść realizacji procesu $\{X_t\}_{1\le t\le n}$, $n = 6822$, którą w sposób graficzny zaprezentowano na rys.1, natomiast wartości tej macierzy przedstawiono w tab. 2



Rysunek 1. Graf przejść międzystanowych wg łańcucha Markowa

Tabela 2. Macierz prawdopodobieństw przejść dla łańcucha Markowa

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	0	0.392	0.002	0.002	0.012	0.032	0.113	0.009	0.007	0.431
S2	0.060	0	0.128	0.003	0.015	0.015	0.142	0.453	0.022	0.162
S3	0.065	0.274	0	0.003	0.009	0.029	0.085	0.294	0	0.241
S4	0.456	0.246	0.053	0	0.035	0	0.105	0	0	0.105
S5	0.056	0.416	0.011	0.034	0	0.146	0.180	0.090	0	0.067
S6	0.433	0.264	0.082	0	0.014	0	0.111	0.005	0.005	0.087
S7	0.084	0.456	0.003	0.044	0.027	0.103	0	0.074	0.001	0.208
S8	0.062	0.510	0.008	0.009	0.005	0.044	0.097	0	0.019	0.247
S9	0.035	0.163	0	0.012	0.035	0.035	0.151	0.442	0	0.128
S10	0.004	0.735	0.008	0	0.012	0.001	0.064	0.173	0.003	0

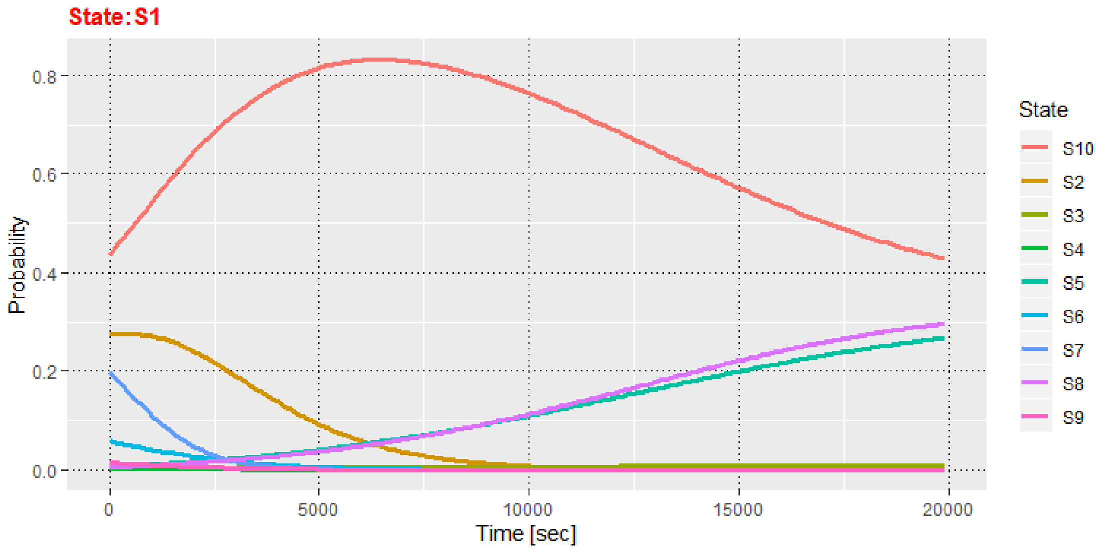
Rozwiązując równanie (4) oszacowano prawdopodobieństwa graniczne. Wartości tych prawdopodobieństw przedstawia tab. 3.

Tabela 3. Wartości granicznych prawdopodobieństw przejść dla łańcucha Markowa

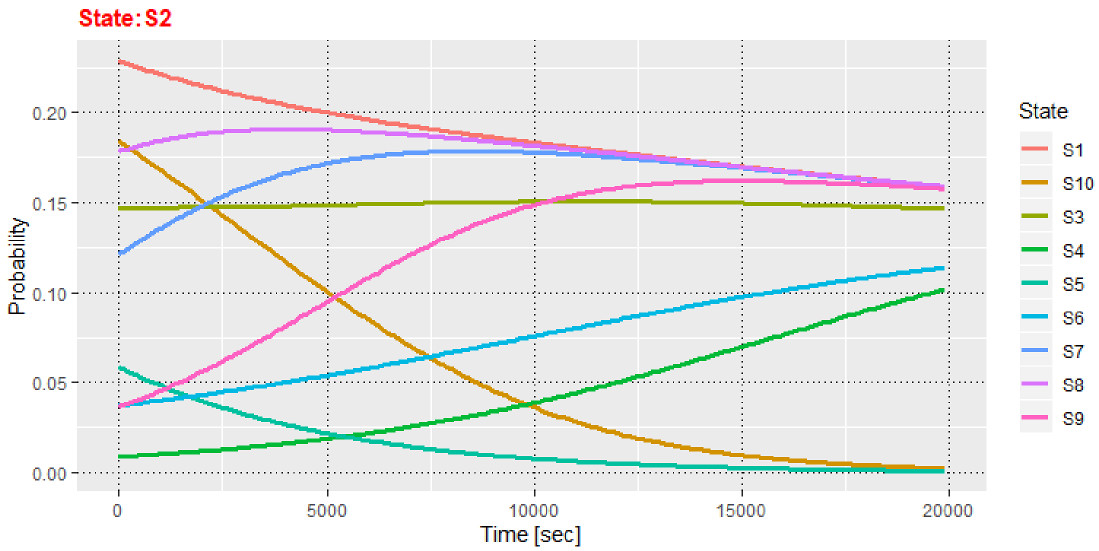
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
π_i	0.064	0.337	0.050	0.008	0.013	0.031	0.099	0.212	0.013	0.173

Ponieważ dla analizowanych danych własność Markowa nie została spełniona, dokonano oszacowania parametrów macierzy prawdopodobieństw przejść, wykorzystując w tym celu wielomianowy model regresji logistycznej. Zbadano wpływ czasu pobytu w określonym stanie eksploatacyjnym na prawdopodobieństwo przejścia do pozostałych stanów. Założono, że prawdopodobieństwo w momencie $t+\tau$ jest wartością warunkowo zależną w od stanu w jakim obiekt przebywał w chwili t oraz od długości czasu τ jego trwania, a także, że po czasie τ system do niego nie powraca. Dla każdego stanu wyznaczono 8 równań regresji logistycznej postaci (9), które opisują zależności dla dziewięciu możliwych przejść. Istotność parametrów strukturalnych zbadano za pomocą testu Walda. W wyniku zastosowania wielomianowej regresji logistycznej dla każdego ze stanów systemu otrzymano macierz prawdopodobieństwa przejścia, która zależy od czasu przebywania τ.

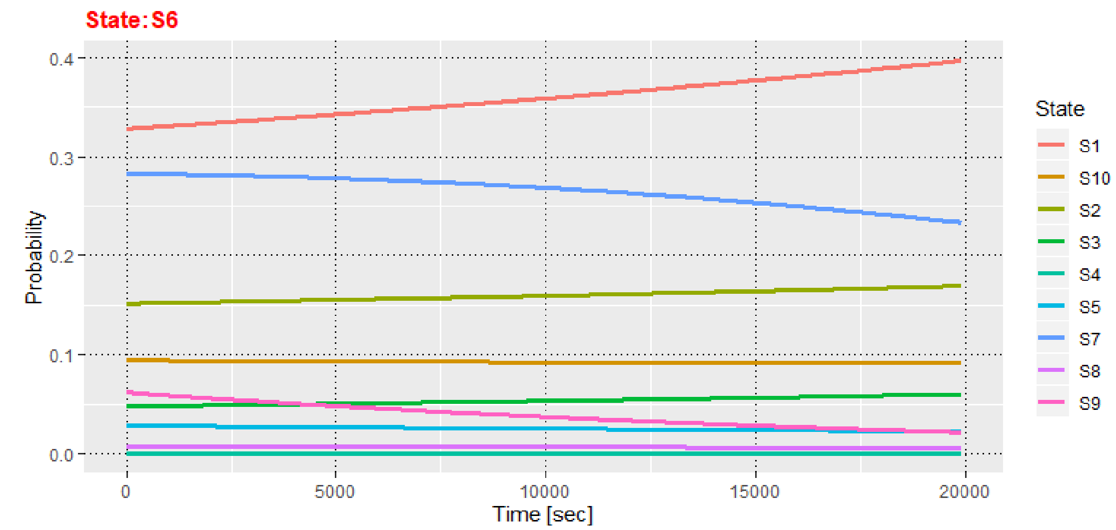
Następnie wyznaczono wykresy obrazujące zmianę prawdopodobieństwa przejścia w zależności od czasu τ pozostania w danym stanie eksploatacyjnym. Dla wybranych stanów: dyspozycyjność, jazda, parkowanie, postój, tankowanie, przedstawiono zależności prawdopodobieństwa przejść od czasu przebywania (rys. 2 – rys. 6).



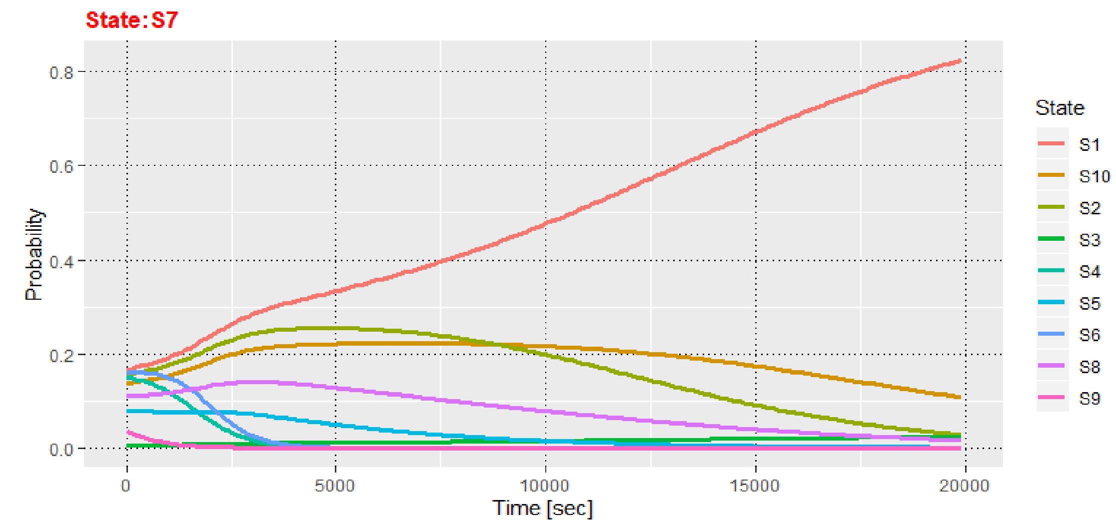
Rysunek 2. Zależność prawdopodobieństwa przejścia ze stanu Dyspozycyjność od czasu jego trwania



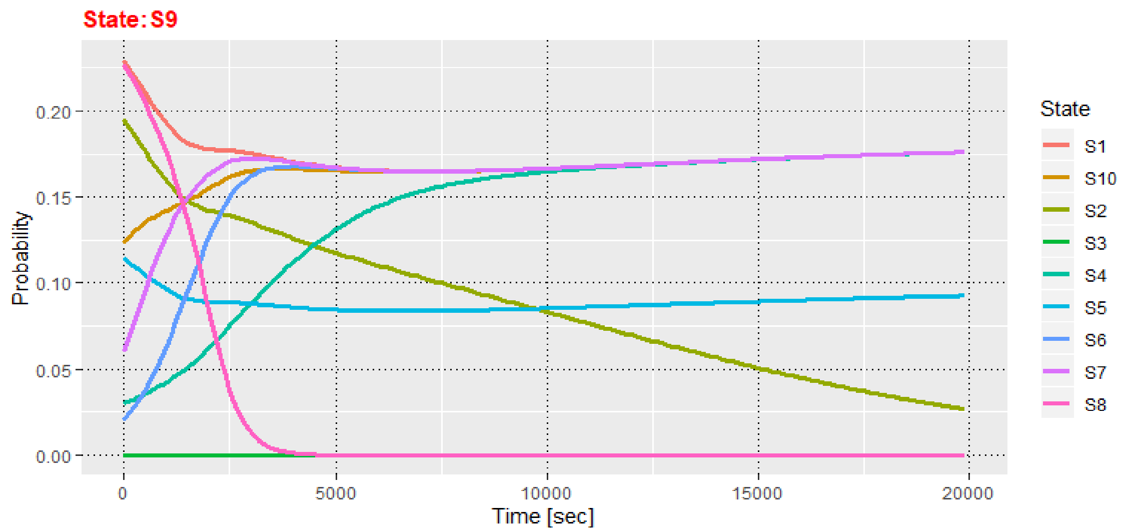
Rysunek 3. Zależność prawdopodobieństwa przejścia ze stanu Jazda od czasu jego trwania



Rysunek 4. Zależność prawdopodobieństwa przejścia ze stanu Parkowanie od czasu jego trwania



Rysunek 5. Zależność prawdopodobieństwa przejścia ze stanu Postój od czasu jego trwania



Rysunek 6. Zależność prawdopodobieństwa przejścia ze stanu Tankowanie od czasu jego trwania

Powyższe rysunki prezentują zależność wartości prawdopodobieństwa przejścia z danego stanu eksploatacyjnego do kolejnych, w zależności od czasu jaki pojazd w nim spędza. Z rysunków widzimy, że wartości tych prawdopodobieństw nie są stałe, co pokazuje niezasadność stosowania klasycznego podejścia, przy oszacowaniu macierzy prawdopodobieństw przejść jak dla łańcucha Markowa. Podejście zaproponowane przez autorów pokazuje sposób wyznaczenia macierzy przejść dla przypadku, gdy czas pozostania systemu w określonym stanie istotnie wpływa na wartości elementów tej macierzy. Zmienność wartości prawdopodobieństw przejść jest uzasadniona i oddaje specyfikę realizacji procesów transportowych, które zdeterminowane są częściowo uregulowaniami prawnymi dotyczącymi na przykład czasu pracy kierowcy, a także terminami wynikającymi z wdrożonej w przedsiębiorstwie strategii eksploatacyjnej, regulującej czasookresy napraw i przeglądów.

Przedstawione rozwiązania są pomocne dodatkowo do opracowania metody pozwalającej na ocenę gotowości systemu do realizacji zadań transportowych. Stany eksploatacyjne można sklasyfikować jako stany zdatności oraz niezdatności i wyznaczyć współczynnik gotowości technicznej jako sumę odpowiednich prawdopodobieństw stanów niezawodnościowych.

5. Zakończenie

W artykule dokonano estymacji macierzy prawdopodobieństw przejść do wyróżnionych stanów eksploatacyjnych, w których przebywał badany pojazd. Popularne w takich oszacowaniach jest zastosowanie łańcuchów Markowa, które wymaga spełnienia warunku o braku pamięci analizowanego procesu. W przedstawionym przypadku własność ta nie została spełniona. Dodatkowo pokazano, że prawdopodobieństwo przejścia do danego stanu eksploatacyjnego jest warunkowo zależne od stanu, w jakim przebywał obiekt oraz od długości czasu jaki w nim pozostawał. Dlatego do ich oszacowania zaproponowano alternatywną metodę. Wykorzystano w tym celu wielomianowy model regresji logistycznej. Otrzymano prawdopodobieństwa przejść, których wartości dla danego stanu różniły się w zależności od długości czasu pobytu pojazdu w stanie wcześniejszym.

Uzyskane wyniki porównano z wartościami otrzymanymi wg łańcucha Markowa - dla którego są one stałe - pokazując, że zastosowanie tak obliczonej macierzy prawdopodobieństw przejść, przy niespełnieniu warunku Markowa, może prowadzić do błędnych wniosków.

Zaproponowany model regresji logistycznej pozwala na krótkookresowe prognozy w zakresie realizacji procesu transportowego. Ocena prawdopodobieństwa przejścia w zależności od wcześniej realizowanych czynności stanowi wsparcie procesu harmonogramowania zadań przewozowych, a także planowania w zakresie obsługi technicznej pojazdów.

W ramach dalszych badań, zaproponowaną metodę warto rozszerzyć o wyznaczanie granicznych wartości estymatorów i ocenę prawdopodobieństw przejść dla poszczególnych stanów w długim przedziale czasu. Pozwoli to na kompleksową ocenę funkcjonowania systemu, a także na wyznaczenie poziomu gotowości do realizacji zadań transportowych. Podział stanów eksploatacyjnych na stany zdatności i niezdatności umożliwi obliczenie współczynnika gotowości technicznej jako sumy odpowiednich prawdopodobieństw granicznych stanów niezawodnościowych. Przedstawione rozwiązanie może także znaleźć zastosowanie przy modelowaniu cykli jezdnych pojazdów ciężarowych, które bezpośrednio odwzorowują rzeczywiste warunki pracy silnika, bądź też podzespołów na hamowni podwoziowej.

Bibliografia

1. Abid L, Masmoudi A, Zouari-Ghorbel S. The Consumer Loan's Payment Default Predictive Model: an Application of the Logistic Regression and the Discriminant Analysis in a Tunisian Commercial Bank. Journal of the Knowledge Economy 2018; 9(3): 948-962.
2. Adha M R, Nurrohmah S, Abdullah S. Multinomial Logistic Regression and Spline Regression for Credit Risk Modelling, Journal of Physics: Conference Series 2018; 1108.
3. Andrzejczak K. Metody prognozowania w modelowaniu eksploatacji środków transportu. Poznań: Wydawnictwo Politechniki Poznańskiej, 2013.
4. Ari E. Using Multinomial Logistic Regression to Examine the Relationship Between Children's Work Status and Demographic Characteristics, Ekonomi ve Yönetim Araştırmaları Dergisi 2016; 4(1): 77-93.
5. Bertens l C, Karel K G, Rutten F H, van Mourik Y, Hoes A W, Reitsma J B. A nomogram was developed to enhance the use of multinomial logistic regression modeling in diagnostic research. Journal of Clinical Epidemiology 2016; 71: 51-57.
6. Bobrowski D. Modele i metody matematyczne teorii niezawodności w przykładach i zadaniach. Warszawa: Wydawnictwa Naukowo – Techniczne, 1985.
7. Borucka A, Niewczas A, Hasilova K. Forecasting the readiness of special vehicles using the semi-Markov model. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (4): 662–669.
8. Calster B V, Hoorde V K, Vergouwe Y, Bobdiwala S, Condous G., Kirk E., Bourne T, Steyerberg E W. Validation and updating of risk models based on multinomial logistic regression. Diagnostic and Prognostic Research 2017; 1:2.
9. Cartella F, Lemeire J, Dimiccoli L, Sahli H. Hidden Semi-Markov Models for Predictive Maintenance, Mathematical Problems in Engineering 2015; 278120.
10. Ferdowsi A, Challita U, Saad W. Deep Learning for Reliable Mobile Edge Analytics in Intelligent Transportation Systems: An Overview, *IEEE Vehicular Technology Magazine* 2019; 14(1): 62-70.
11. Fisz M. Rachunek prawdopodobieństwa i statystyka matematyczna. Warszawa: PWN, 1967.
12. Gautam V, Vaidya R R. Evidence on the determinants of investment-cash flow sensitivity Indian Economic Review December 2018; 53(1–2): 229-244.
13. Girtler J, Ślęzak M. Application of the theory of semi-Markov processes to the development of a reliability model of an automotive vehicle, Archiwum Motoryzacji 2012; 2: 15-27.
14. Gniedenko B.W., Bielajew J.K., Sołowiew A.D. Metody matematyczne w teorii niezawodności. Warszawa: Wydawnictwa Naukowo-Techniczne, 1965.
15. Gola A. Reliability analysis of reconfigurable manufacturing system structures using computer simulation methods, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019, 21 (1): 90-102.
16. Grabski F. Teoria semi-Markowskich procesów eksploatacji obiektów technicznych. Gdynia: Zeszyty Naukowe Wyższej Szkoły Marynarki Wojennej 75A, 1982.
17. Grzywacz W., Burnewicz J. Ekonomika transportu. Gdańsk: Wydawnictwa Komunikacji i Łączności, 1989.
18. Halim H, Ramli M I, Adisasmita S A, Aly S H, Prasetijonternational J A. Relationship model between accident factors and the traffic accident severity using logistic regression model, Journal of Engineering and Science Application 2018; 4(2), 169-181.
19. Hastie T., Tibshirani R., Friedman J., The elements of statistical learning, Springer-Verlag New York Inc., 2009.

20. James G., Witten D., Hastie T., Tibshirani R., An introduction to statistical learning, Springer-Verlag GmbH, 2013.
21. Jasiulewicz-Kaczmarek M., Gola A., Maintenance 4.0 Technologies for Sustainable Manufacturing – an Overview, IFAC PapersOnLine 2019; 52-10: 91-96.
22. Jaźwiński J, Grabski F. Niektóre problemy modelowania systemów transportowych. Radom: Instytut Technologii Eksploatacji, 2003.
23. Jia X, Cui L, Xing L. New insights into reliability problems for supply chains management based on conventional reliability model. Eksploatacja i niezawodnosc – Maintenance and Reliability 2018; 20 (3): 465–470.
24. Knopik L, Migawa K. Semi-Markov system model for minimal repair maintenance. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 256–260.
25. Kozłowski E, Mazurkiewicz D, Żabiński T, Prucnal S, Sęp J. Assessment model of cutting tool condition for real-time supervision system. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (4): 679–685.
26. Landowski B. Example of applying markov decision process o model vehicle maintenance process. Journal of KONES Powertrain and Transport 2013;20(4): 209-218
27. Lewitowicz J, Zieja M, Szelmanowski A, Pazur A. New Measures of Operational Readiness for Multi-states Avionics Integrated Systems with Reduced Efficiency. In: Hanus R., Mazur D., Kreischer C. (eds) Methods and Techniques of Signal Processing in Physical Measurements. Lecture Notes in Electrical Engineering, 2019; 548: 165-180.
28. Lewitowicz J., Kustroń K. Podstawy eksploatacji statków powietrznych. Warszawa: Wydawnictwo ITWL, 2003.
29. Linhares F S, Costa F M D, Beiruth A X. Earnings management and investment efficiency. Revista Brasileira de Gestão de Negócios 2018; 20(2): 295-310.
30. Macián V, Tormos B, Herrero J. Maintenance management balanced scorecard approach for urban transport fleets. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (2): 226–236.
31. 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.
32. Murat Y S, Uludag N. Route choice modelling in urban transportation networks using fuzzy logic and logistic regression methods, Journal of Scientific & Industrial Research; 67(01): 19-27.
33. Nabian M A, Meidani H. Deep Learning for Accelerated Seismic Reliability Analysis of Transportation Networks, Computer-Aided Civil and Infrastructure Engineering 2018; 33: 443–458.
34. Nguyen T. Understanding Municipal Bond Ratings Using Ordered Multinomial Logistic Regression. Senior Independent Study Theses 2018; 8171.
35. Nicholson A, Schmöcker J, Bell M, Iida Y. Assessing Transport Reliability: Malevolence and User Knowledge. In The Network Reliability of Transport, Emerald Group Publishing Limited 2003; 1-22.
36. Niewczas A, Rymarz J, Debicka E. Stages of operating vehicles with respect to operational efficiency using city buses as an example. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (1): 21-27.
37. Nowakowska M. Spatial and temporal aspects of prior and likelihood data choices for Bayesian models in road traffic safety analyses. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (1): 68–75.
38. Odeck J, Bråthen S. A meta-analysis of DEA and SFA studies of the technical efficiency of seaports: A comparison of fixed and random-effects regression models. Transportation Research Part A: Policy and Practice 2012; 46(10): 1574-1585.

39. Oliva F L. A maturity model for enterprise risk management, International Journal of Production Economics 2016; 173: 66-79.
40. Osuji M N, Mejeha R O, Simonyan J B, Gbolagun A S. Determinants of Enterprise Preference of Cassava-based farmers by Gender in Southeast, Nigeria. Research Journal of Food and Nutrition 2018; 2(4): PP 45-52.
41. Park M, Wu A D. Investigating differential options functioning using multinomial logistic regression. International Journal of Quantitative Research in Education 2017; 4(1-2): 94 - 119.
42. Paś J, Klimczak T. Selected issues of the reliability and operational assessment of a fire alarm system. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (4): 553–561.
43. Paś J, Rosiński A. Selected issues regarding the reliability-operational assessment of electronic transport systems with regard to electromagnetic interference. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19 (3): 375–381.
44. Peter D, Hart A. Multinomial Logistic Regression Examination of TV Time and Two Different Measures of Obesity in U.S. Adults. American Journal of Public Health Research 2018; 6(5): 222-226.
45. Raco A, Pesce A, Fraschetti F, D'Andrea G, Polli F M, Acqui M, Frati A. Risk of Postoperative Performance Status Worsening after Resection of Lesions Involving the Motor Pathway: A Multinomial Logistic Regression Model. Journal of Neurological Surgery, Part A: Central European Neurosurgery 2018; 79(06): 453-463.
46. Rymarczyk T.; Kozłowski E.; Kłosowski G.; Niderla K. Logistic Regression for Machine Learning in Process Tomography. Sensors 2019, 19, 3400.
47. Sanusi W, Jemain A A, Zin W Z, Zahari M. The Drought Characteristics Using the First-Order Homogeneous Markov Chain of Monthly Rainfall Data in Peninsular Malaysia, Water Resources Management 2015; 29(5): 1523–1539.
48. Schwartz J, Guasch J L, Wilmsmeier G, Stokenberga A. Logistics, Transport and Food Prices in LAC: Policy Guidance for Improving Efficiency and Reducing Costs. Latin America and Caribbean Region Environment and Water Resources occasional paper series 2009; 2.
49. Shao C, Xia J C, Lin T G, Goulias K G, Chen C. Logistic regression models for the nearest train station choice: A comparison of captive and non-captive stations, Case Studies on Transport Policy 2015; 3(4): 382-391.
50. Sowa A. Stan obiektu jako wieloznaczne pojęcie we współczesnej eksploatacji technicznej. Czasopismo Techniczne. Mechanika 2012; 109(7-M): 269-278.
51. Stewart W.J. Probability, Markov Chains, Queues, and Simulation, 2009.
52. Szkoda M, Kaczor G. Reliability and availability assessment of diesel locomotive using fault tree analysis. Archives of Transport 2016; 40(4): 65-75.
53. Wang Z, Huang X, Liang Y J. Oil-gas reservoir lithofacies stochastic modeling based on one- to three-dimensional Markov chains. Journal of Central South University 2018; 25(6): 1399–1408.
54. Zhu Y, Xie C, Sun B, Wang G-J, Yan X-G. Predicting China's SME Credit Risk in Supply Chain Financing by Logistic Regression, Artificial Neural Network and Hybrid Models. Sustainability 2016; 8(5): 1-17.
55. Ziółkowski J., Borucka A. Markov model in logistic management of enterprise. Journal of Konbin 2016; 38(1): 271-290.

dr hab. inż. Piotr Nikończuk, dr hab. inż. Włodzimierz Rosochacki prof. ZUT

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
Wydział Techniki Morskiej i Transportu
Al. Piastów 41, 71-065 Szczecin, Poland
E-mail: piotr.nikonczuk@zut.edu.pl, wlodzimierz.rosochacki@zut.edu.pl

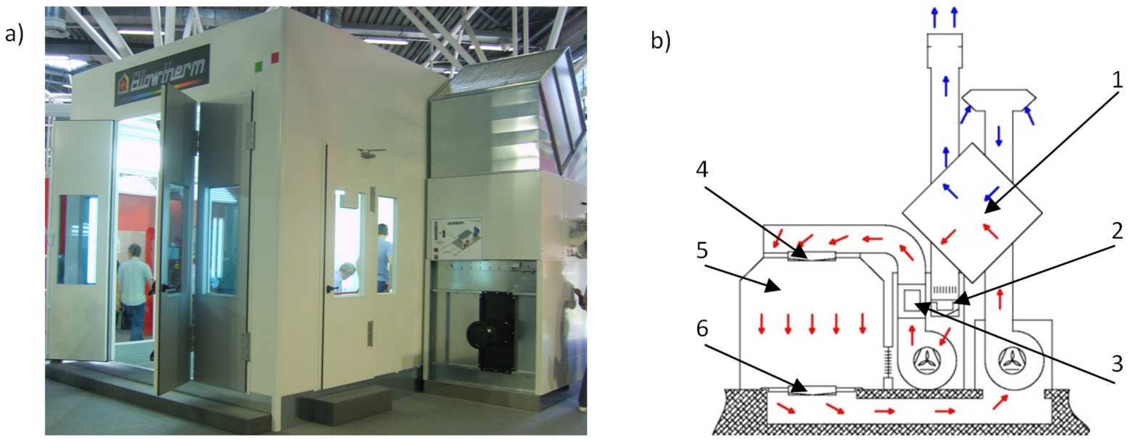
Koncepcja miary niezawodności rekuperatora kabiny lakierniczej

***Słowa kluczowe**: niezawodność, kabina lakiernicza, rekuperator, osady lakiernicze*

Streszczenie: Odkładające się na lamelach rekuperatora osady lakiernicze powodują stopniowe zmniejszanie przekroju poprzecznego kanałów rekuperatora. Skutkiem tego procesu są wzrosty oporów przepływu powietrza oraz oporu termicznego przy wymianie ciepła. Oba zjawiska wpływają negatywnie na niezawodność urządzenia.
W artykule przedstawiono koncepcję miary niezawodności rekuperatora. W tym celu sformułowano podstawowe wymaganie niezawodnościowe (nieuszkadzalność) oraz zdefiniowano uszkodzenia utożsamiając je z utratą zapasu strumienia powietrza oraz zapasu efektywności wymiany ciepła. Na tym tle określono cechy zdatności urządzenia, granice ich obszarów oraz krytyczny czas utraty zdatności rekuperatora.

1. Wprowadzenie

Rekuperator jest urządzeniem technicznym stosowanym w systemach wentylacyjnych, również w komorach lakierniczych. Ze względu na wymagania technologiczne związane z lakierowaniem istotne jest, aby proces ten był prowadzony w odpowiednich warunkach, określonych przede wszystkim właściwą temperaturą i czystością powietrza [30]. Celem stosowania rekuperatora jest odzyskiwanie części ciepła odpadowego pochodzącego ze zużytego powietrza wywiewanego z przestrzeni roboczej kabiny lakierniczej oraz wstępne ogrzanie pobieranego z zewnątrz powietrza świeżego. Wewnątrz rekuperatora znajdują się na przemian kanały ciepłego i zimnego powietrza, oddzielone od siebie cienkimi, aluminiowymi lamelami. W rekuperatorze krzyżowym strumienie ciepłego i zimnego powietrza przepływają względem siebie pod kątem prostym. Za pośrednictwem lamel wymieniane jest ciepło pomiędzy strumieniami powietrza. Na rysunku 1 przedstawiono kabinę lakierniczą z rekuperatorem krzyżowym oraz schemat cyrkulacji powietrza podczas pracy kabiny w trybie lakierowania. Świeże powietrze pobierane z zewnątrz jest wstępnie ogrzewane w rekuperatorze (1) następnie po oczyszczeniu w filtrze zgrubnym (2) ogrzewane jest do zadanej temperatury przez palnik z wymiennikiem ciepła (3). Ogrzane powietrze jest ostatecznie oczyszczane w filtrze nawiewnym (4) oraz nawiewane do przestrzeni roboczej kabiny (5). W przestrzeni roboczej odbywa się proces aplikacji lakierów, podczas którego powstaje mgła lakiernicza. We mgle lakierniczej unoszą się lotne związki organiczne (LZO) oraz drobiny lakieru, które nie znalazły się na lakierowanej powierzchni. Powietrze przechodząc przez przestrzeń roboczą kabiny zabiera ze sobą mgłę lakierniczą i opuszcza kabinę za pośrednictwem filtra typu *paint stop* (6), którego zadaniem jest zatrzymywanie drobin lakieru. Następnie oczyszczone powietrze za pośrednictwem kanału wywiewnego trafia do rekuperatora (1) gdzie częściowo oddaje ciepło do czerpanego, świeżego powietrza.



Rys. 1. Kabina lakiernicza z rekuperatorem a) obiekt rzeczywisty b) Schemat obiegu powietrza, 1- rekuperator krzyżowy, 2 – filtr wstępny, 3 – palnik z wymiennikiem ciepła, 4 – filtr nawiewny, 5 – przestrzeń robocza kabiny lakierniczej, 6 – filtr typu *paint stop*

Eksploatacji rekuperatora w kabinie lakierniczej towarzyszy proces sedymentacji drobin lakieru na lamelach rekuperatora w kanałach ciepłego powietrza. Prowadzone są badania i modelowanie zanieczyszczania wymienników ciepła [6]. Model powstawania osadów lakierniczych dla rekuperatora krzyżowego przedstawiono w pracy [14]. Bezpośrednim skutkiem tego zjawiska jest zmniejszanie się przekroju poprzecznego kanałów ciepłego powietrza w wymienniku ciepła. Zostało to szerzej opisane m.in. w pracy [12]. W rezultacie prowadzi to do wzrostu oporów przepływu powietrza oraz oporu termicznego w procesie wymiany ciepła. Wzrost oporu termicznego skutkuje obniżeniem efektywności odzyskiwania ciepła w rekuperatorze, natomiast zmniejszenie wolumenu wymienianego powietrza prowadzi do powstawania zagrożenia wybuchu. Proces sedymentacji drobin lakieru na lamelach rekuperatora ma więc charakter destrukcyjny, mający istotny wpływ na niezawodność jego działania.

Wymogi bezpieczeństwa aplikacji lakierów zostały poruszone w publikacji [18], określane i aktualizowane są w odpowiednich regulacjach Unii Europejskiej [29] oraz pozostałe światowe akty prawne, między innymi w Australii [32], w Stanach Zjednoczonych [34] oraz w Nowej Zelandii [33]. Wymogi zawarte w lokalnych regulacjach zostały zaprezentowane w przewodniku dobrych praktyk dla prac lakierniczych opracowanym przez National Air Filtration Association [30]. Podobne opracowanie zostało wydane w Wielkiej Brytanii [31]. Wpływ regulacji na branżę lakiernictwa samochodowego przedstawiono w opracowaniu [35]. W kabinach lakierniczych zidentyfikowano dwa główne zagrożenia: zatrucia lakiernika oraz powstawania mieszanki wybuchowej. Metodę określania ryzyka wybuchu dla kabin stosowanych do farb proszkowych opisano w pracy [25].

Obniżenie tempa wzrostu osadów lakierniczych jest możliwe między innymi poprzez poprawę skuteczności oczyszczania powietrza usuwanego z kabiny lakierniczej z drobin mgły lakierniczej. Efektywność filtrów typu *paint stop* uzależniona jest od ich rodzaju [4] oraz rozmiaru drobin lakieru unoszonych we mgle lakierniczej. Analizę efektywności oczyszczania powietrza z drobin lakieru w funkcji ich wielkości przedstawiono w pracy [1]. Wyniki porównawcze efektywność filtrów przedstawiono w pracy [4] natomiast szerszy zakres prac badawczy został zawarty w podsumowaniu zakończonego projektu badawczego [3]. Rozmiar drobin uzależniony jest od materiału lakierniczego oraz parametrów aplikacji. Analizę wielkości drobin lakieru przedstawiono w publikacji [20] natomiast znacznie rozszerzone wyniki zawarto w pracy [21]. Oddzielną analizę powstawania mgły lakierniczej oraz oczyszczania powietrza dla technologii bez udziału sprężonego powietrza (*airless spray painting*) przedstawiono w pracy [23].

Problematyka oczyszczania powietrza usuwanego z lakierni jest wciąż aktualna, przegląd stanu wiedzy zaprezentowano w pracy [22]. Prowadzone są prace nad nowymi rozwiązaniami filtracji powietrza w kabinach lakierniczych [7]. Rozważane są techniki oczyszczania grawitacyjnego na mokro [10] oraz metody bez stosowania wymiennych filtrów (*medialess dynamic filtration*) [26]. Do tej pory nie opracowano technologii oczyszczania powietrza, która zapewnia całkowite usuwanie drobin mgły lakierniczej. Dla usuwania lotnych związków organicznych rozważane są technologie biofiltracji [8] zaproponowano między innymi zastosowanie biologicznej filtracji strumieniowej [24] lub filtrów grzybowych [16].

Konsekwencją sedymentacji drobin lakieru na lamelach rekuperatora jest, jak już wcześniej zauważono, pogorszenie poziomu niezawodności urządzenia. Obecne dokumentacje techniczne kabin lakierniczych wyposażonych w rekuperatory krzyżowe nie zawierają wytycznych dotyczących okresowych inspekcji stanu rekuperatora oraz częstości jego oczyszczania. Zidentyfikowanie głównych cech zdatności rekuperatora oraz określenie wartości progowych dla obszarów zdatności ułatwi wyznaczenie częstości inspekcji i oczyszczania rekuperatora, w celu zapewnienia bezpieczeństwa aplikacji lakierów. W niniejszej pracy przedstawia się propozycję miar niezawodności rekuperatora oraz oszacowania krytycznego czasu zdatności rekuperatora $t_{kr,}$ po którym nastąpi jego uszkodzenie, przyjmując założenie, że istotnym wymaganiem niezawodnościowym jest jego nieuszkadzalność. Czas krytyczny zdatności rekuperatora t_{kr} jest również wyznacznikiem częstości przeglądów i prac konserwacyjnych rekuperatora.

2. Cechy zdatności

W dalszych rozważaniach przyjęto założenie, że na niezawodność rekuperatora istotny wpływ ma proces odkładania się osadów lakierniczych. Jest to zasadnicze założenie upraszczające rzeczywistość, polegające na pominięciu innych, mniej istotnych procesów mogących prowadzić do innych form uszkodzeń (np. mechanicznych). Opisany powyżej wpływ osadów lakierniczych na parametry pracy rekuperatora pozwala na przyjęcie do dalszych analiz dwóch cech określających jego zdatność. Są nimi: zapas efektywności wymiany ciepła oraz zapas strumienia powietrza. Utratę zapasu w odniesieniu do każdej z tych cech utożsamia się z wystąpieniem uszkodzenia i przejściem urządzenia w stan zawodności.

Mając na uwadze charakter zjawiska prowadzącego do uszkodzenia rekuperatora, w celu określenia jego miary niezawodności za poprawne i użyteczne cechy zdatności przyjęto *zapas efektywności wymiany ciepła* oraz *zapas spadku ciśnienia*.

2.1 Zapas efektywności wymiany ciepła

Cecha odnosi się do uszkodzenia utożsamianego ze stanem rekuperatora, w którym osiągnięta zostaje graniczna wartość współczynnika przewodności cieplnej k_{gr}. Cecha ta dotyczy aspektów ekonomicznych związanych z rekuperacją ciepła odpadowego.

Efektywność wymiany ciepła w rekuperatorze związana jest ze strumieniem ciepła $\dot{\boldsymbol{Q}}(t)$[1], który jest zależny od współczynnika przewodności cieplnej $\boldsymbol{k(t)}$ oraz różnicy

[1] W niniejszej pracy symbole zmiennych losowych pisane są grubą czcionką.

temperatur strumieni powietrza po obu stronach lameli rekuperatora ΔT (przy założeniu, że różnica ta jest zdeterminowana i niezmienna w czasie):

$$\dot{\boldsymbol{Q}}(t) = \boldsymbol{k}(t)\Delta T \tag{1}$$

Zapas efektywności wymiany ciepła $\boldsymbol{h}_1(t)$ określa się jako:

$$\boldsymbol{h}_1(t) = \dot{\boldsymbol{Q}}(t) - \dot{Q}_{gr} \tag{2}$$

gdzie:

$\dot{Q}_{gr}$ – graniczna wartość zapasu efektywności ciepła.

Mając na uwadze (1), zależność opisująca zapas efektywności wymiany ciepła $\boldsymbol{h}_1(t)$ przyjmuje postać:

$$\boldsymbol{h}_1(t) = (\boldsymbol{k}(t) - k_{gr})\Delta T \tag{3}$$

Wartość współczynnika przewodności cieplnej $\boldsymbol{k}(t)$ określa się dla rekuperatora z uwzględnieniem wpływu narastających na powierzchni lameli warstw osadów lakierniczych. Współczynnik ten jest zmienną losową, gdyż narastanie na powierzchni lameli warstw osadów lakierniczych jest zjawiskiem losowym. Dla dowolnej chwili τ i punktu powierzchni lameli opisanego współrzędnymi (x_o, y_o), realizację procesu $\{\boldsymbol{k}(x,y,t)\}$ opisuje się następującą zależnością [17]:

$$k(x_o, y_o, \tau) = \frac{1}{\frac{1}{\alpha_1} + \frac{\delta_R}{\lambda_{TR}} + \frac{\delta_S(x_o, y_o, \tau)}{\lambda_{TS}} + \frac{1}{\alpha_2}} \tag{4}$$

gdzie:

α_1, α_2 – współczynnik przejmowania ciepła powietrza [W/(m^{2}K)]; z założenia wielkość zdeterminowana,

δ_R – grubość lameli rekuperatora [m]; z założenia wielkość zdeterminowana,

$\delta_S(x_o,y_o,\tau)$ – grubość warstwy narastających osadów lakierniczych w punkcie powierzchni lameli określonym współrzędnymi (x_o,y_o) [m]; realizacja procesu stochastycznego $\{\boldsymbol{\delta}_S(x,y,t)\}$ w chwili τ,

λ_{TR} – przewodność cieplna lameli rekuperatora [W/(mK)]; z założenia wielkość zdeterminowana,

λ_{TS} – przewodność cieplna osadów lakierniczych [W/(mK)]; z założenia wielkość zdeterminowana.

Do wyznaczenie wartości cechy zdatności $\boldsymbol{h}_1(t)$ w chwili τ realizację współczynnika $\boldsymbol{k}(t)$ określa się ze wzoru:

$$k(\tau) = \frac{1}{xy}\int_0^y\int_0^x k(x,y,\tau)dydx \tag{5}$$

2.2 Zapas spadku ciśnienia

Cecha odnosi się do uszkodzenia utożsamianego ze stanem, w którym zostaje osiągnięta graniczna wartość spadku ciśnienia strumienia powietrza ΔP_{gr} w kanałach rekuperatora. Przepływające przez kanał wentylacyjny powietrze pokonuje opór tarcia występujący na ściankach kanału wentylacyjnego. Opór ten wraz ze zmniejszającym się w czasie przekrojem kanału powoduje spadek ciśnienia na całej jego długości prowadząc do zmniejszenia objętościowego strumienia powietrza. Obniżenie wolumenu wymienianego powietrza skutkuje niebezpieczeństwem powstawania mieszanki wybuchowej w przestrzeni roboczej kabiny. Prowadzi to również do groźby zatrucia pracującego wewnątrz lakiernika [32].

Zapas spadku ciśnienia $\boldsymbol{h}_2(t)$ wyraża się zależnością:

$$\boldsymbol{h}_2(t) = \Delta P_{gr} - \Delta \boldsymbol{P}(t) \tag{6}$$

gdzie:
$\Delta \boldsymbol{P}(t)$ – chwilowy spadek ciśnienia strumienia powietrza.

Spadek ciśnienia powietrza w kanale wentylacyjnym jest procesem stochastycznym $\{\Delta \boldsymbol{P}(t)\}$, którego realizacje zależą od długości kanału l oraz losowo zmiennego w czasie jednostkowego współczynnika oporu $\boldsymbol{r}(t)$:

$$\Delta P(t) = r(t)l \tag{7}$$

Współczynnik oporu $\boldsymbol{r}(t)$ zwany również jednostkowym spadkiem ciśnienia [17] jest zależny od wielu parametrów w tym dwóch, które zmieniają swoje wartości w czasie:

$$\boldsymbol{r}(t) = \frac{\boldsymbol{\lambda_F}(t)\varsigma w^2}{2\boldsymbol{d}(t)} \tag{8}$$

gdzie:
$\boldsymbol{\lambda_F}(t)$ – bezwymiarowy współczynnik tarcia; wielkość losowa,
ς – gęstość powietrza [kg/m^3]; z założenia wielkość zdeterminowana,
w – średnia prędkość przepływu powietrza [m/s]; z założenia wielkość zdeterminowana
$\boldsymbol{d}(t)$ – średnica równoważna przekroju kanału [m]; wielkość losowa.

Średnicę równoważną dla przekroju poprzecznego kanału w ogólnej postaci wyznacza się korzystając z zależności [17]:

$$d = \frac{2ab}{a+b} \tag{9}$$

gdzie a i b oznaczają wymiary boków prostokątnego przekroju poprzecznego kanału.

Uwzględniając losowo zmienną w czasie grubość narastających osadów lakierniczych $\boldsymbol{\delta_S}(t)$ zależność opisująca realizacje średnicy równoważnej $\boldsymbol{d}(t)$ w chwili τ przyjmuje postać:

$$d(\tau) = \frac{2ab - 4(a+b)\delta_S(\tau) + 8\delta_S{}^2(\tau)}{a + b - 4\delta_S(\tau)} \tag{10}$$

Bezwymiarowy współczynnik tarcia $\lambda_F(t)$ zmienia swoją wartość w czasie ze względu na zależność od liczby Reynoldsa. Dla przepływu turbulentnego współczynnik tarcia opisany jest następującą zależnością:

$$\lambda_F(t) = \frac{0,3164}{\sqrt[4]{\boldsymbol{Re}(t)}} \tag{11}$$

Liczba Reynoldsa $\boldsymbol{Re}(t)$ zależy od średnicy równoważnej przekroju poprzecznego w kanale wentylacyjnym $\boldsymbol{d}(t)$. W chwili τ jej realizację określa się na podstawie zależności:

$$\boldsymbol{Re}(\tau) = \frac{wd(\tau)}{\upsilon} \tag{12}$$

gdzie υ oznacza kinematyczny współczynnik lepkości [m^2/s].

3. Granice obszarów zdatności

Zdefiniowane powyżej cechy zdatności stwarzają podstawę wyznaczenia poniższych granic obszarów zdatności

- dla cechy *zapas współczynnika przewodności cieplnej*:

gdy rekuperator zdatny (brak uszkodzenia)

$$\boldsymbol{h}_1\left(\dot{\boldsymbol{Q}}(t), \dot{Q}_{gr}\right) > 0 \tag{13}$$

gdy rekuperator niezdatny

$$\boldsymbol{h}_1\left(\dot{\boldsymbol{Q}}(t), \dot{Q}_{gr}\right) \leq 0 \tag{14}$$

- dla cechy *zapas spadku ciśnienia*:

gdy rekuperator zdatny (brak uszkodzenia)

$$\boldsymbol{h}_2\left(\Delta\boldsymbol{P}(t), \Delta P_{gr}\right) > 0 \tag{15}$$

gdy rekuperator niezdatny

$$\boldsymbol{h}_2\left(\Delta\boldsymbol{P}(t), \Delta P_{gr}\right) \leq 0 \tag{16}$$

4. Miara niezawodności

Przyjmuje się, że podstawowym wymaganiem niezawodnościowym stawianym konstrukcji rekuperatora jest jego funkcjonowanie bez uszkodzeń w określonym przedziale czasu. Podejście takie znajduje uzasadnienie w funkcji, którą pełni to urządzenie. Osiągnięcie opisanych powyżej stanów krytycznych utożsamianych w niniejszej pracy z uszkodzeniem jest jednoznaczne z nieakceptowanym pogorszeniem funkcjonalności urządzenia w istotny sposób rzutujący na bezpieczeństwo i koszt eksploatacji kabiny lakierniczej oraz jakość procesu aplikacji lakierów.

Mając powyższe na względzie przyjmuje się, że miarą niezawodności, która dobrze charakteryzuje sformułowane powyżej wymaganie niezawodnościowe, jest prawdopodobieństwo jego spełnienia w analizowanym okresie:

$$R(t) = P\left(\left(\boldsymbol{h}_1(t) > 0\right) \cap \left(\boldsymbol{h}_2(t) > 0\right)\right) \tag{17}$$

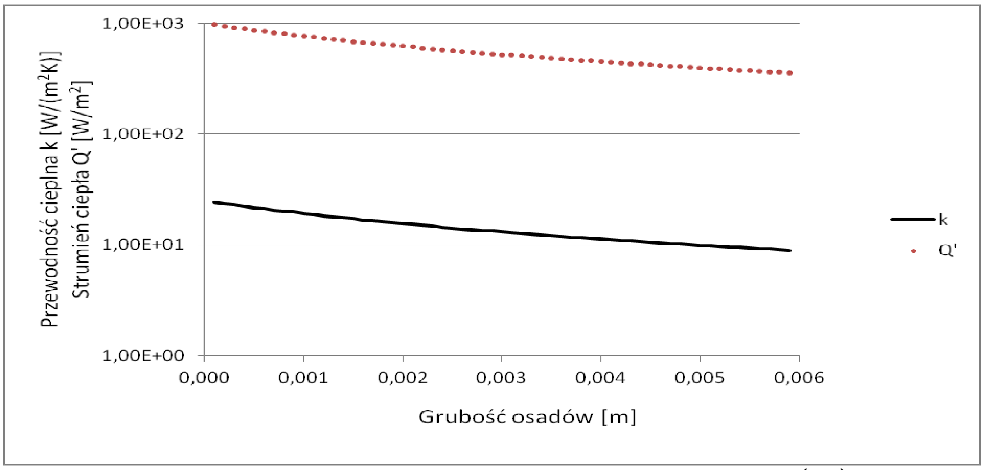
Określeniu podlega więc prawdopodobieństwo pracy rekuperatora z zachowaniem obu analizowanych tu cech zdatności. Należy zauważyć, że zależność (17) określa prawdopodobieństwo zaistnienia dwóch zdarzeń zależnych. Zależność ta wynika z powiązania obu cech zdatności z procesem narastania osadów lakierniczych, reprezentowanym tu przez proces stochastyczny $\{\boldsymbol{\delta}_S(t)\}$.

5. Analiza cech zdatności

Zakłada się, że przedstawione powyżej cechy zdatności zależą od stałych parametrów konstrukcyjnych rekuperatora oraz zmiennej w czasie grubości osadów lakierniczych $\boldsymbol{\delta}_S(t)$.

Przeprowadzono wstępną analizę cech zdatności na przykładzie rekuperatora dedykowanego dla kabin lakierniczych, znajdującego się w dokumentacji ofertowej jednego z przedsiębiorców działających na rynku branży lakierniczej [28]. Rekuperator był przedmiotem badań prezentowanych w pracy [13].

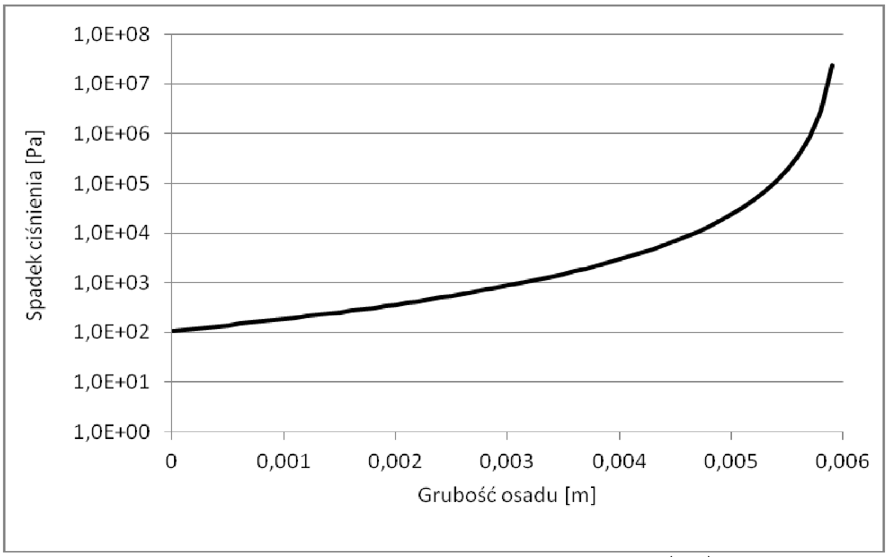
Dla przedstawionego rekuperatora przeanalizowano wpływ grubości osadów lakierniczych na zmiany sformułowanych cech zdatności. Na rysunku 2 przedstawiono zmiany współczynnika przewodności cieplnej $k(\delta_S)$ według równania (4) oraz odpowiadające im zmiany strumienia ciepła $\dot{Q}(\delta_S)$ według równania (1). Mają one wpływ na wartość cechy zdatności $\boldsymbol{h}_1(t)$. Do obliczeń przyjęto współczynnik przewodności cieplnej osadów lakierniczych λ_{TS} = 0.082±0.003 [W/(mK)], metodykę pomiarów i opracowanie wyników wraz z analizą błędów dla przewodności cieplnej osadów opisano w pracy [15]. Dla obliczeń według równia (4) przyjęto następujące wartości: przewodność cieplna aluminium λ_{TR} = 200 [W/(mK)] oraz jednakowe wartości współczynnika przejmowania ciepła dla powietrza po obu stronach lameli $\alpha_1 = \alpha_2 = 50$ [W/(m^{2}K)], za dokumentacją [28] określono grubość lameli rekuperatora δ_R = 2e-4 [m]. Wartości strumienia ciepła $\dot{Q}(\delta_S)$ obliczono dla różnicy temperatur ΔT = 40 [K].



Rys. 2 Obliczone zmiany współczynnika przewodności cieplnej $k(\delta_S)$ oraz strumienia ciepła $\dot{Q}(\delta_S)$

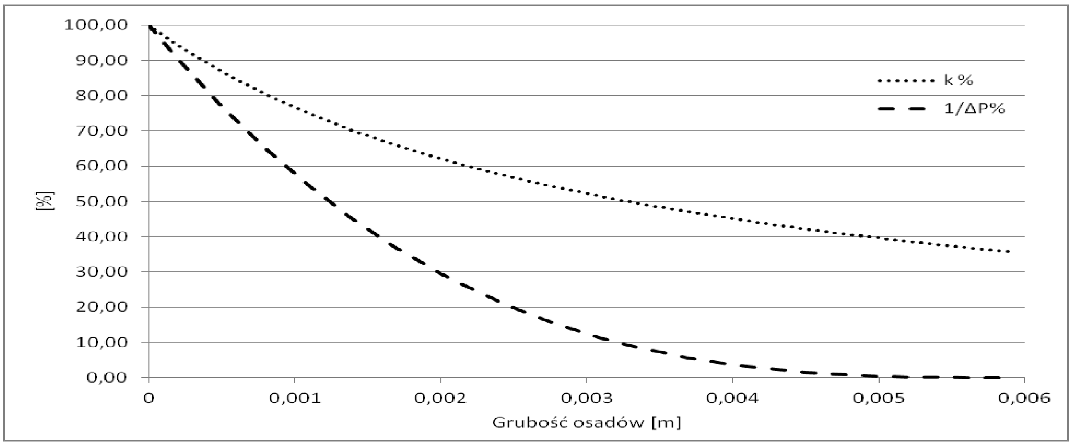
Na rysunku 3 przedstawiono wzrost spadku ciśnienia w funkcji grubości warstwy osadów lakierniczych $\Delta P(\delta_S)$. Spadek ciśnienia jest związany z cechą zdatności $\boldsymbol{h}_2(t)$. Obliczenia przeprowadzono według równania (7). Przyjęto następujące wartości zmiennych: gęstość powietrza ς = 1.2 [kg/m^3], średnia prędkość przepływu powietrza w = 5.56 [m/s], kinematyczny współczynnik lepkości υ = 1.5e-5 [m^2/s]. Na podstawie dokumentacji [28] określono następujące parametry kanałów rekuperatora: boki przekroju prostokątnego kanału

a = 1.2e-2 [m], b = 1 [m] długość kanału l = 1 [m], liczba kanałów w rekuperatorze oddzielnie dla ciepłego i zimnego powietrza n = 60. Osady lakiernicze odkładają się wyłącznie w kanałach z usuwanym z kabiny ciepłym powietrzem. Obliczenia spadku ciśnienia przeprowadzono dla pojedynczego kanału ciepłego powietrza przy założeniu, że we wszystkich przekrojach kanałów jest jednolity rozkład prędkości powietrza oraz ma miejsce przepływ burzliwy. Dla średnicy równoważnej $\boldsymbol{d(t)}$ opisanej równaniem (10) przyjęto jednorodną, uśrednioną wartość grubości osadów lakierniczych $\boldsymbol{\delta_S}(t)$.



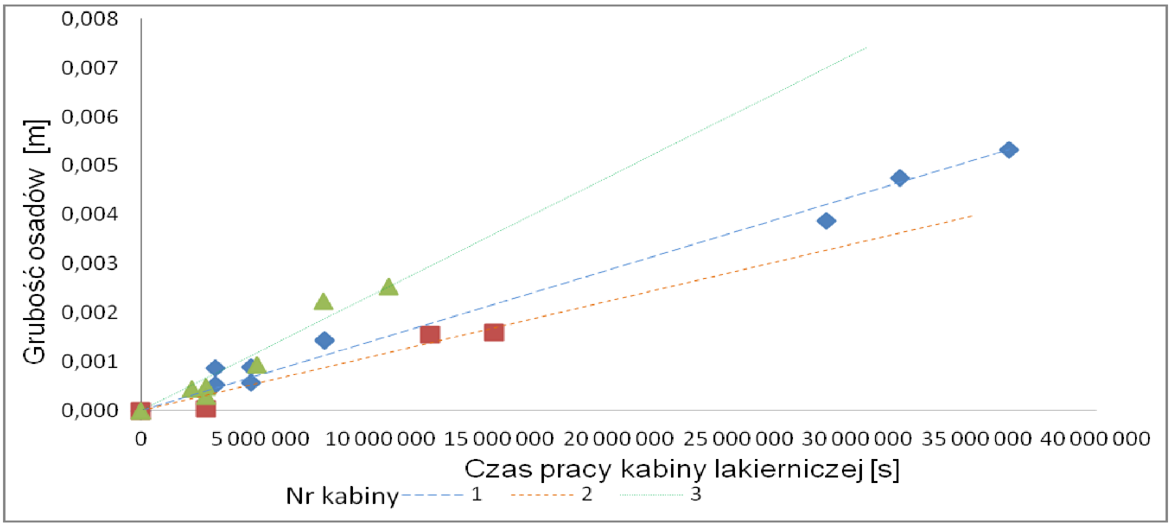
Rys. 3 Obliczony spadek ciśnienia $\Delta P(\delta_S)$

Na rysunku 4 przedstawiono procentowe zmiany efektywności wymiany ciepła oraz odwrotności spadku ciśnienia w funkcji grubości osadów lakierniczych. Na wykresie przedstawiono odwrotność spadku ciśnienia $1/\Delta P(\delta_S)$ w celu poprawy przejrzystości w porównaniu z procentowymi zmianami strumienia ciepła $\dot{Q}(\delta_S)$. Punkty początkowe wynoszące 100% dla obu parametrów wskazują ich wartości dla stanu czystych, niepokrytych osadami lakierniczymi lamel rekuperatora. Analiza prezentowanego wykresu wskazuje na znacznie większy wpływ procesu narastania osadów na zmianę spadku ciśnienia, a w rezultacie na zmianę wartości cechy $\boldsymbol{h}_2(t)$.



Rys. 4 Procentowe zmiany strumienia ciepła $\dot{Q}(\delta_S)$ i odwrotności ciśnienia $1/\Delta P(\delta_S)$ w zależności od grubości osadów

Tempo wzrostu osadów lakierniczych uzależnione jest od wielu parametrów i jest procesem o zmiennej dynamice. Na rysunku 5 przedstawiono wyniki pomiarów grubości osadów lakierniczych w trzech kabinach lakierniczych. Przedstawione na wykresie punkty reprezentują wartości średnie z pomiarów po danym okresie czasu pracy kabiny. Linie trendu przedstawiają uśrednione wartości tempa wzrostu osadów w każdej z kabin. Metodyka przeprowadzenia badań oraz ich warunki zostały opisane w pracy [13]. Pomiary prowadzono w kabinach lakierniczych niewyposażonych w rekuperatory. Punkty pomiarowe ze względów technicznych były zlokalizowane w każdej kabinie na pokrywie przepustnicy powietrza w kanale wyrzutni. Jest to miejsce, gdzie zwyczajowo instaluje się rekuperator (rysunek 1). Wyniki pomiarów były podstawą do opracowania modelu symulacyjnego odkładania się osadów na lamelach rekuperatora, model numeryczny oraz wyniki symulacji przedstawiono w pracy [14]. W modelu przyjęto, że we wszystkich przekrojach kanałów rekuperatora jest jednakowa prędkość powietrza oraz przepływ ma charakter burzliwy.

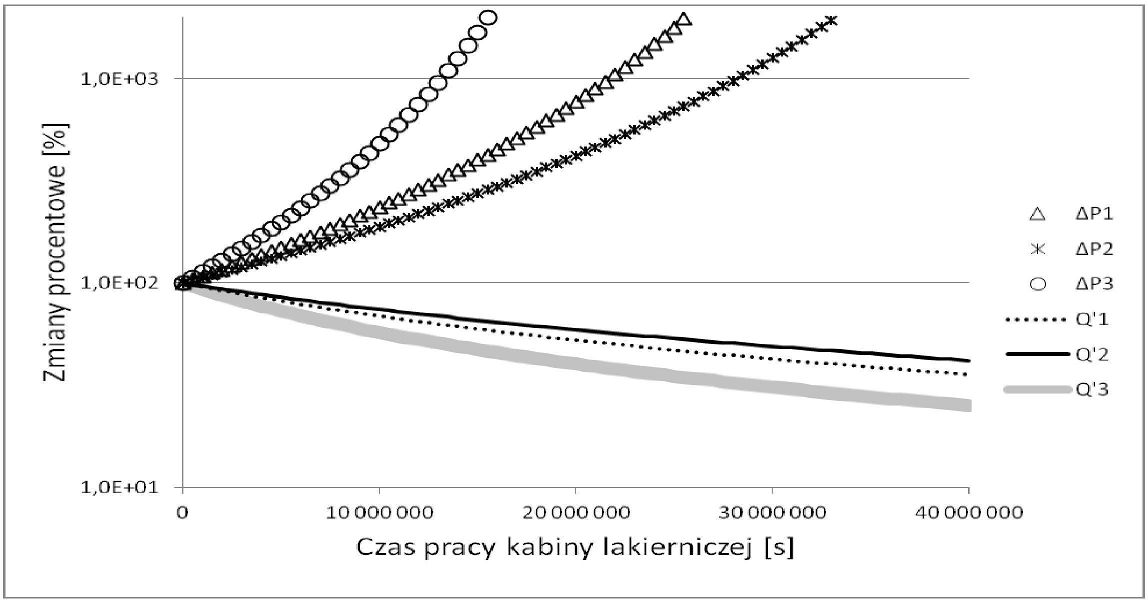


Rys. 5 Tempo wzrostu osadów lakierniczych w trzech kabinach lakierniczych [13]

Analiza wyników prezentowanych na rysunku 5 wskazuje na silnie losowy charakter procesu narastania osadów. Przykładowo, linie trendu wyników pomiarów grubości osadów w kabinach nr 2 i 3 wskazują na ponad dwukrotnie większe tempo wzrostu w kabinie nr 3 niż w kabinie nr 2. Wiąże się to między innymi z losowym wpływem na ten proces takich czynników jak: sumaryczne udziały czasów pracy w trybach lakierowania oraz suszenia w całkowitym czasie eksploatacji kabiny lakierniczej, parametry nastaw oraz efektywność transferu pistoletu lakierniczego, efektywność wywiewnego filtra powietrza w kabinie, umiejętności lakiernika, kształt i rozmiary lakierowanych obiektów.

Na podstawie linii trendu przedstawionych na rysunku 5 opracowano dla poszczególnych kabin procentowe zmiany spadku ciśnienia oraz względnych współczynników przewodności cieplnej w dziedzinie czasu. Wyniki przedstawiono na rysunku 6. Na wykresach dla kabin 1, 2 i 3 odpowiednio oznaczono spadki ciśnienia jako $\Delta P1$, $\Delta P2$ i $\Delta P3$ oraz strumień ciepła jako $Q'1$, $Q'2$ i $Q'3$. Obliczenia przeprowadzono z użyciem równań przedstawionych w rozdziale 2.

Ze względu na wykładniczy wzrost spadku ciśnienia uwidoczniony na rysunku 3, na wykresie na rysunku 6 przedstawiono fragmenty krzywych, które nie przekraczają wartości zmian 2000 [%]. Wartości te są wynikami obliczeń teoretycznych, które nie będą osiągalne w normalnej eksploatacji kabiny lakierniczej z rekuperatorem.



Rys. 6 Procentowe zmiany spadku ciśnienia (*ΔP*1, *ΔP*2, *ΔP*3) oraz strumienia ciepła (*Q'*1, *Q'*2, *Q'*3) w poszczególnych kabinach

Porównując dynamikę procentowych zmian współczynnika przewodności cieplnej oraz spadku ciśnienia zauważalny jest istotny wzrost procentowej zmiany spadku ciśnienia w stosunku do procentowej zmiany współczynnika przewodności cieplnej.

Opisane w pracy [13] badania procesu eksploatacji kabin lakierniczych, obserwacje oraz wywiady z eksploratorami kabin lakierniczych dają podstawę do oszacowania wartości, które proponuje się w analizowanych tu cechach zdatności uznać za graniczne. W odniesieniu do cechy $\boldsymbol{h}_1(t)$ utożsamianej z zapasem efektywności wymiany ciepła proponuje się wstępnie przyjąć za wartość graniczną

$$\dot{Q}_{gr} = 0.5\dot{Q}_N = 500[\mathrm{W/m^2}] \tag{18}$$

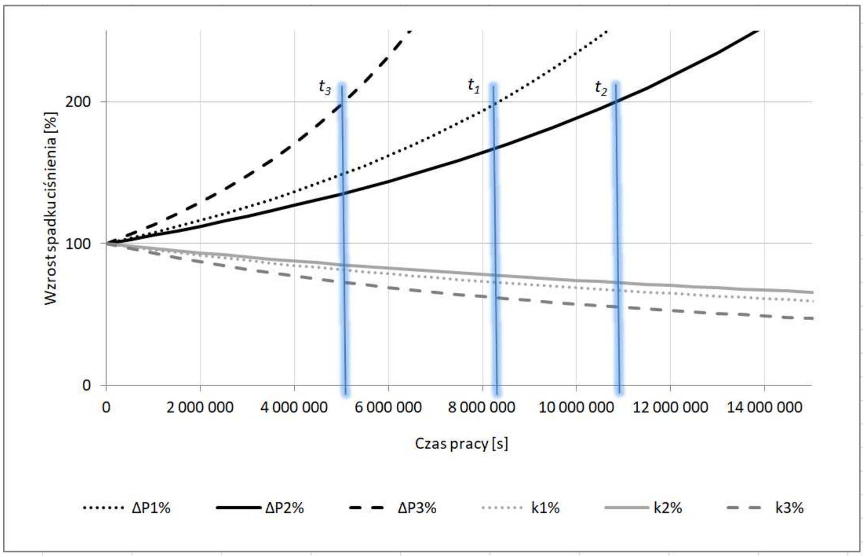
przy czym $\dot{Q}_N$ oznacza nominalną wartość strumienia ciepła dla nowego niezabrudzonego osadami rekuperatora. Wartość nominalną strumienia ciepła można odczytać z rysunku 2 $\dot{Q}_N$ = 1000 [W/m^2]. Przy takiej wartości strumienia ciepła efektywność energetyczna rekuperatora osiąga połowę wartości nominalnej, co obniża o połowę szacowane korzyści ekonomiczne użytkownika kabiny lakierniczej. Uznano, że połowa uzyskiwanych oszczędności tytułem odzyskanego ciepła stanowi granicę opłacalności kosztów inwestycyjnych związanych z zakupem i instalacją rekuperatora.

W odniesieniu do cechy $\boldsymbol{h}_2(t)$ za wartość graniczną spadku ciśnienia ΔP_{gr} proponuje się dwukrotność nominalnego spadku ciśnienia ΔP_N

$$\Delta P_{gr} = 2\Delta P_N = 216[\mathrm{Pa}] \tag{19}$$

Za nominalną wartość ΔP_N przyjęto spadek ciśnienia na rekuperatorze w stanie czystym, gdy lamele rekuperatora nie są pokryte osadami lakierniczymi. Na rysunku 3 przedstawiono obliczone według równania (7) zmiany spadku ciśnienia w zależności od grubości osadów. Wartość początkowa spadku ciśnienia dla grubości osadu δ_s = 0 [mm] wynosi ΔP_N = 108 [Pa]. Całkowity spadek ciśnienia w kanałach wentylacyjnych kabiny lakierniczej jest indywidualny dla każdej kabiny. Związany jest on z wieloma parametrami a przede wszystkim z długością oraz przekrojami kanałów, liczbą i rodzajem kształtek w instalacjach wentylacyjnych, konstrukcją wymiennika ciepła do ogrzewania powietrza, rodzajami oraz stanem czystości filtrów powietrza a także rekuperatorem. Na tej podstawie uznano, iż dwukrotność nominalnego spadku ciśnienia na rekuperatorze stanowi dla niego wartość krytyczną.

Dla przyjętych powyżej wartości krytycznych oraz na podstawie szybszych zmian spadku ciśnienia w funkcji grubości osadów, za wiodącą cechę w szacowaniu czasu utraty zdatności eksploatacyjnej przyjęto cechę $\boldsymbol{h}_2(t)$.



Rys. 7. Szacowane czasy zdwojenia spadku ciśnienia

Na rysunku 7 przedstawiono szacowane czasy, w których dla poszczególnych kabin zostały osiągnięte wartości graniczne ΔP_{gr}. Czasy, w których nastąpiło zdwojenie spadku ciśnienia dla poszczególnych kabin oznaczono kolejno jako t_1, t_2 i t_3 adekwatnie do numeru kabiny. Czasy zdwojenia dla poszczególnych kabin odpowiednio wynoszą:

t_1 = 8346234 [s]

t_2 = 10837105 [s]

t_3 = 5061598 [s]

Wartości te są zróżnicowane, szczególnie zauważalne jest, że czas t_2 jest prawie dwukrotnie większy od czasu t_3. Różnorodność wartości jest skutkiem silnej losowości procesu narastania osadów lakierniczych.

Na podstawie powyższych obliczeń, jako krytyczny czas utraty zdatności rekuperatora przyjęto wartość t_3, jest to najkrótszy okres, w którym nie nastąpi uszkodzenie

$$t_{kr} = t_3 = 5061598\ [s]$$

6. Podsumowanie

W proponowanym modelu niezawodności rekuperatora w kabinie lakierniczej wyróżniono dwie cechy zdatności: *zapas efektywności wymiany ciepła* $\boldsymbol{h}_1(t)$ oraz *zapas spadku ciśnienia* $\boldsymbol{h}_2(t)$. Cechy zdatności $\boldsymbol{h}_1(t)$ oraz $\boldsymbol{h}_2(t)$ mają zróżnicowaną zmienność w zależności od grubości osadów lakierniczych $\boldsymbol{\delta}(t)$. Jak wskazuje analiza procesu narastania osadów ma on charakter losowy. Cecha $\boldsymbol{h}_1(t)$ związana jest ze spadkiem efektywności odzysku ciepła w rekuperatorze. Ma ona charakter ekonomiczny. Natomiast cecha zdatności $\boldsymbol{h}_2(t)$ związana jest ze wzrostem oporów przepływu powietrza przez rekuperator. Obniżenie wolumenu wymiany powietrza w przestrzeni roboczej kabiny lakierniczej może prowadzić do zwiększenia koncentracji mgły lakierniczej oraz LZO. Skutkiem tego może być zatrucie lakiernika lub powstawanie mieszanki wybuchowej. W związku z tym cecha zdatności rekuperatora $\boldsymbol{h}_2(t)$ związana jest z bezpieczeństwem eksploatacji kabiny lakierniczej. W publikacji wskazano na znacznie szybsze tempo zmian cechy zdatności $\boldsymbol{h}_2(t)$ w porównaniu do cechy $\boldsymbol{h}_1(t)$ w dłuższym okresie eksploatacji. Ostatecznie cechę $\boldsymbol{h}_2(t)$ uznano jako dominującą, która ma znaczący wpływ na wyznaczenie okresowości inspekcji i oczyszczania rekuperatora.

Oszacowano czas pracy rekuperatora w stanie nieuszkodzonym na t_{kr} = 5061598 sekund co jest równoważne 1406 godzinom pracy kabiny lakierniczej. Po upływie tego czasu należy przeprowadzić inspekcję oraz oczyszczanie rekuperatora z osadów lakierniczych. Uzyskane wyniki dotyczą trzech kabin lakierniczych i stanowią wstępne wartości. Ze względu na niejednorodną prędkość wzrostu osadów lakierniczych, trudno szacować dokładne czasy osiągnięcia wartości granicznych przez cechy zdatności. Pozyskanie i uporządkowanie wyników pomiarów tempa wzrostu osadów w wielu kabinach lakierniczych ułatwi określenie średnich oraz krytycznych przedziałów czasu eksploatacji kabin lakierniczych w jakich należy się spodziewać uszkodzenia rekuperatora.

Proponowane cechy zdatności przedstawiają ujęcie, przy założeniu, że na niezawodność rekuperatora istotny wpływ ma grubość osadów lakierniczych z pominięciem innych możliwych uszkodzeń. Podobne podejście przedstawiono w pracy [19]. Stan uszkodzenia rekuperatora skutkuje utratą zdatność eksploatacyjnej całej kabiny lakierniczej. Wskazane cechy zdatności rekuperatora stają się również cechami zdatności kabiny lakierniczej, nie są one jednak jedynymi cechami wskazującymi na poziom zdatności kabiny. Uśredniając tempo wzrostu osadów lakierniczych można utworzyć indywidualny model kabiny lakierniczej jako systemu wieloelementowego opisanego w pracy [27]. W pracy [11] zaprezentowano również metodę szybkiej oceny niezawodności złożonego systemu technicznego, gdzie elementy składowe mają różne czasy odnawiania. Współczesne technologie przemysłu 4.0 polegające na połączeniu urządzeń przemysłowych z internetem oraz składowaniem danych w sieci niosą możliwości automatycznej akwizycji i składowania w chmurze wyników pomiarów parametrów pracy kabin lakierniczych. Wybrane parametry w sposób pośredni mogą wskazywać na stan czystości rekuperatora. Analiza zebranych wyników w chmurze umożliwia zdalne określenie zdatności rekuperatora [5]. Analiza zdatności może zostać przeprowadzona metodą uczenia maszynowego [2] lub z zastosowaniem logiki zbiorów rozmytych [9].

Literatura:

1. Chan T.L., D'Arcy J.B., Schreck R.M. High-solids paint overspray aerosols in a spray painting booth: particle size analysis and scrubber efficiency. American Industrial Hygiene Association Journal 1986 ;7: 411–417. DOI.org/10.1080/15298668691389964.

2. Chen X, Zhang Z, Zhang Z. Real-time equipment condition assessment for a class-imbalanced dataset based on heterogeneous ensemble learning. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21(1): 68–80, http://dx.doi.org/10.17531/ein.2019.1.9.

3. Darvin C.H. Fractional penetration of paint overspray arrestors. Project summary, United States Environmental Protection Agency, April 1997, https://permanent.access. gpo.gov/ websites/epagov/www.epa.gov/ORD/WebPubs/projsum/600sr97011.pdf.

4. Darvin C. H., Cox L. S. & Smith D. Comparative Study of Spray Booth Filter System Efficiency, Journal of the Air & Waste Management Association, 1999; 49(3): 339:344,

5. Fang Y, Tao W, Tee KF. A new computational method for structural reliability with big data. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21(1): 159–163, http://dx.doi.org/10.17531/ein.2019.1.18

6. Hassan Al-Haj Ibrahim Fouling in Heat Exchangers, MATLAB - A Fundamental Tool for Scientific Computing and Engineering Applications -2012; 3, Vasilios N. Katsikis, IntechOpen, DOI: 10.5772/46462. Available from: https://www.intechopen.com/books/matlab-a-fundamental-tool-for-scientific-computing-and-engineering-applications-volume-3/fouling-in-heat-exchangers

7. Howarth J. L., Anand S. C., Design, development and characterization of a novel and innovative exhaust filter media for the global automotive industry. Textile Research Journal DOI: 10.1177/0040517515619349.

8. Kazenski S.L., Kinney K.A. Biofiltration of paint spray booth emissions: Packing media considerations and VOC interactions, in: Proceedings of the 2000 Annual Meeting and Exhibition of Air and Waste Management Association, June 2000, Salt Lake City, A & WMA, 1–14.

9. Khaniyev T, Baskir MB, Gokpinar F, Mirzayev F. Statistical distributions and reliability functions with type-2 fuzzy parameters. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21(2): 268–274, http://dx.doi.org/10.17531/ein.2019.2.11.

10. Kim H.T., Jung C.H., Oh S.N., Lee K.W. Particle removal efficiency of gravitational wet scrubber considering diffusion, interception, and impaction, Journal of Environmental Engineering Science 2001; 2: 125–136.

11. Młynarski S, Pilch R, Smolnik M, Szybka J, Wiązania G. A Method for rapid evaluation of k-out-of-n systems reliability. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (1): 170–176, http://dx.doi.org/10.17531/ein.2019.1.20.

12. Nikończuk P. Study of Heat Recovery in Spray Booths, Metal Finishing 2013; 111(6): 37-39.

13. Nikończuk P. Preliminary analysis of heat recovery efficiency decrease in paint spray booths. Transactions of the Institute of Metal Finishing 2014; 92(5): 235-237.

14. Nikończuk P. Preliminary modeling of overspray particles sedimentation at heat recovery unit in spray booth. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (3): 387–393, http://dx.doi.org/10.17531/ein.2018.3.6.

15. Nikończuk P., Dobrzyńska R. Preliminary Measurements of Overspray Sediment's Thermal Conductivity. Ochrona przed Korozją 2018; 61 (2): 40-42, DOI:10.15199/ 40.2018.2.3.

16. Qi B., Moe W.M., Kinney K.A. Treatment of paint spray booth off-gases in a fungal biofilter. Journal of Environmental Engineering 2005; 2: 180–189.

17. Recnagel – Sprenger – Hönmann - Schramek, Ogrzewanie i klimatyzacja, Poradnik, Wydanie 1, EWFE, Gdańsk 1994.

18. Ron J. Ventilation and paint filtration requirements for spray booths. Metal Finishing 2007; 105 (6), 82–84

19. Rosochacki W. Basis of reliability model for offshore structures bearing. Scientific Journals Maritime University of Szczecin, 2009 vol. 19 (91): 92-96.

20. Sabty-Daily R.A., Hinds W.C., Froines J. Size distribution of chromate paint aerosol generated in a bench-scale spray booth. The Annals of Occupational Hygiene 2005; (1): 33–45.

21. Salaimeh A. A., Adornato A., Akafuah N. K., Saito K. Assessment of Spray Envelope Using Droplet Breakup Regime and Mechanisms. ICLASS 2015, 13th Triennial International Conference on Liquid Atomization and Spray Systems, Tainan, Taiwan, August 23-2 2015.

22. Schweizer R. Technological Advancements in Overspray Collectors. Metal Finishing, 2010; 108(4): 38-45.

23. Settles G.S., Miller J.D., Hartranft T.J., Brandt A.D. Visualization and collection of overspray from airless spray painting. In: 8th International Symposium on Flow Visualization, Sorrento, Italy 1998.

24. Webster T.S., Tongna A.P., Guarini W.J., McKnight L. Treatment of volatile organic compound emissions from a paint spray booth application using biological trickling filtration. In: 2000 USC-TRG Conference on Biofiltration (an air pollution control technology), Los Angeles, California, USA, October 2000, University of Southern California, Reynolds Group, 41–50.

25. Wiechuła B. M. The assessment of explosion risk in cabins during spraying coating powder. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2005; 7 (25): 3-15.

26. Wright S. R., Bond J. H. Hazardous waste Minimization of paint overspray via medialess dynamic particle filtration. Final Report 25 Apr 97 -25 Jan 98, INNOVATECH, INC.

27. Zhang Y, Ma Y, Ouya ng 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.

28. Archiwum firmy Techniki systemów Aplikacyjnych Sp. z o.o.

29. Coating plants – Spray booths for application of organic liquid coating materials – Safety requirements, European Standard, Final Draft, EN 12215, June 2004.

30. Guideline for Spray Finishing Particulate Recommended Practice, Virginia Beach VA, National Air Filtration Association, 2012, http://www.nafahq.org/wp-content/uploads/ Spray FinishingParticulateFiltrationa.pdf, dostęp: 11.05.2015.

31. Guidance on the application of the dangerous substances and explosive atmospheres regulations (DSEAR) 2002 to motor vehicle repair (MVR), Health and Safety Executive, Information Document HSE, 2002, 803/71, http://www.hse.gov.uk/foi/internalops/ ocs/800-899/803_71/803_71id.htm#app2, dostęp: 04.12.2015.

32. Occupational safety and health act 1984 and occupational safety and health regulations 1996, Code of practice spray painting, West Perth, Australia WorkSafe Western Australia Commission, June 2009, https://www.commerce.wa.gov.au/sites/default/files/atoms/files/ code_spray_ painting.pdf, dostęp: 24.03.2016.

33. Spray coating a guide to the spray coating regulations 1962, Wellington, New Zealand, Occupational Safety and Health Service, Department of Labour, 1962, July 2013, http:// legislation. govt.nz/regulation/public/1962/0054/latest/whole.html#DLM16295.

34. United States Environmental Protection Agency (USEPA), National emission standards for hazardous air pollutants: Surface coating of miscellaneous metal parts and products; Final rule, Federal Register 69 (1), January 2, 2004, 130–192, https://www.gpo.gov/ fdsys/pkg/FR-2004-01-02/pdf/03-21917.pdf.

35. United States Environmental Protection Agency (USEPA), Regulatory impact analysis for the proposed automobile and light duty truck coating NESHAP: Final Report, EPA-452/R-01-013, North Carolina, USEPA, Research Triangle Park, N.C., 2002.

Artur Wymysłowski, Krystian Jankowski

Wroclaw University of Science and Technology, Faculty of Microsystem Electronics and Photonics, ul. Janiszewskiego 11/17, 50-372 Wrocław, Poland
E-mail: artur.wymyslowski@pwr.edu.pl

Badania wytrzymałości połączeń lutowanych stosowanych w montażu w mikroelektronice

Strength analysis of soldered joints used in microelectronics packaging

Słowa kluczowe: *mikroelektronika, stopy lutownicze, niezawodność, kumulacja uszkodzeń*

Key words: *microelectronics, solder alloys, reliability, damage accumulation*

Streszczenie: Celem badań był problem kumulacji uszkodzeń dla stopów lutowniczych stosowanych w montażu w mikroelektronice w wyniku zmęczenia i pełzania na skutek złożonego profilu obciążeń. Wybrane rodzaje uszkodzeń przyczyniają się do ograniczenia czasu życia współczesnych urządzeń elektronicznych. Aktualnie prowadzi się badania z wykorzystaniem jednego rodzaju uszkodzeń i często pomijany jest problem ich wzajemnej interakcji. Uwzględnienie problemu wzajemnej interakcji pozwoliłoby na bardziej precyzyjne prognozowanie bezawaryjnego czasu pracy współczesnych urządzeń elektronicznych i/lub przyspieszenie testów niezawodnościowych. W ramach zrealizowanych badań przeprowadzono analizę wytrzymałości połączeń lutowanych dla stopu lutowniczego $Sn_{63}Pb_{37}$ z wykorzystaniem metody Hot Bump Pull. Wyniki przedstawionych badań obejmują: analizę wytrzymałości, analizę statystyczną oraz problem kumulacji uszkodzeń w wyniku złożonego profilu obciążeń.

Abstract: The aim of the research was the problem of damage accumulation for solder alloys used in microelectronics packaging due to creep and fatigue as a result of a combined profile of loading conditions. The selected failure modes affect the lifetime of contemporary electronic equipment. So far the research activities are focused on a single failure mode and the problem of their interaction is often omitted. Taking into account the failure modes interaction would allow more precise lifetime prediction of the contemporary electronic equipment and/or would allow for reduction of time required for reliability tests. Within the taken research framework the reliability analysis of solder joints was conducted for the $Sn_{63}Pb_{37}$ solder alloy using the Hot Bump Pull method. The results of the presented research contain: reliability tests, statistical analysis and the problem of a damage accumulation due to a combined profile of loading conditions.

1. Wprowadzenie.

Elektronika jest jedną z najszybciej rozwijających się dziedzin wiedzy i inżynierii we współczesnym świecie. W związku z ciągłym dążeniem to miniaturyzacji i integracji większość komponentów elektronicznych jest projektowana i produkowana w tzw. skali mikro. Z tego też powodu często używany jest termin mikroelektronika. Komponenty mikroelektroniczne stanowią integralną część praktycznie każdego elektronicznego urządzenia przemysłowego czy domowego. Niestety, podobnie jak inne urządzenia, także komponenty mikroelektroniczne charakteryzują się ograniczonym czasem życia. Jednym z podstawowych problemów dotyczących ich niezawodności są połączenia. W montażu w mikroelektronice (z ang. microelectronics packaging [17]) stosuje się połączenia lutowane, klejone i zgrzewane, z czego kluczowe są połączenia lutowane [13,15,27]. Większość uszkodzeń połączeń lutowanych następuje w wyniku obciążeń termomechanicznych, a ich bezpośrednią przyczyną są naprężenia będące wynikiem niedopasowania współczynników rozszerzalności cieplnej materiałów [17,35,40]. Szacuje się, że około 65% uszkodzeń w montażu w mikroelektronice jest związanych z problemami termomechanicznymi [2,38].

Niezawodność definiowana jest jako właściwość obiektów, określająca ich prawidłowe działanie w zadanych warunkach środowiskowych, w zadanym przedziale czasu. Matematyczny opis niezawodności pozwala na ocenę prawdopodobieństwa wystąpienia uszkodzenia w ustalonych warunkach eksploatacji. Jednym z tradycyjnych sposobów prognozowania niezawodności połączeń w montażu w mikroelektronice jest analiza teoretyczna tzw. złącza bimateriałowego. Złącze bimateriałowe stanowi strukturę dwóch materiałów o różnych właściwościach termomechanicznych. Prognozowanie wytrzymałości wymaga znajomości rozkładu naprężeń w samym złączu oraz jego najbliższym otoczeniu. Złącze może ulec uszkodzeniu w wyniku pękania lub rozwarstwienia. Rozwiązanie analityczne dla struktury ze złączem bimateriałowym zostało po raz pierwszy przedstawione w roku 1925 przez Timoshenko [30]. Dotyczyło ono odkształcenia struktury, jak i maksymalnego naprężenia, jakie występuje w obszarze złącza. Natomiast rozwiązanie analityczne opisujące rozkład naprężeń w obszarze samego złącza zostało przedstawione w roku 1989 przez Suhira [29]. Analiza teoretyczna stanu naprężenia w obszarze złącza bimateriałowego jest trudna i wymaga przyjęcia szeregu uproszczeń, np. dotyczących geometrii złącza czy liniowego modelu materiałowego. Z tego też powodu, we współczesnych zastosowaniach inżynierskich, preferuje się metody numeryczne oparte na symulacjach z wykorzystaniem metody MES [39,41]. Metody te pozwalają na analizę rzeczywistej geometrii złącza oraz nieliniowych modeli materiałowych. Dodatkowo, w połączeniu z wiedzą o wytrzymałości materiałów, metody numeryczne pozwalają na uwzględnienie w ramach analizy tzw. kryteriów uszkodzenia w celu prognozowania uszkodzeń. Kryteria uszkodzeń są ściśle związane z rodzajem uszkodzenia oraz materiałem. W mikroelektronice uszkodzenia rozpoznawane są w wyniku badań niezawodnościowych, a odpowiadające im kryteria wyznaczane są na drodze doświadczalnej. Problem ten został m.in. opisany w opublikowanej w roku 2018 normie IPC/JEDECC-9301 "Numerical Analysis Guidelines for Microelectronics Packaging Design and Reliability" [24].

W wyniku dynamicznego rozwoju przemysłu elektronicznego oraz szybkiego postępu technologii mikroelektronicznych istotnym problemem stało się precyzyjne badanie niezawodności ze względu na postępującą integrację i miniaturyzację komponentów składowych. Przykładem mogą być pola lutownicze, których rozmiary są mniejsze niż $100\ \mu m$, a tym samym rozmiary połączeń są porównywalne z rozmiarami ziaren. Biorąc pod uwagę skalę połączeń lutowanych oraz różnorodne obciążenia, jakim poddawane są komponenty mikroelektroniczne, okazało się, że stosowane w skali makro metody prognozowania niezawodności połączeń lutowanych są niewystarczające w przypadku skali mikro [36]. Z tego też powodu koniecznym jest opracowania, z jednej strony zaawansowanych technik pomiarowych, typowych dla technologii mikroelektronicznych, natomiast z drugiej strony zaawansowanych metod prognozowania uszkodzeń w wyniku obciążeń termomechanicznych [25]. Wymaga to jednak wiedzy z zakresu zachowania materiałów w funkcji naprężenia, temperatury i czasu, tj. reologii. Jednym z typowych problemów reologii jest analiza zjawiska zmęczenia i pełzania [20]. Należy podkreślić, że w wielu dziedzinach inżynierii, zarówno zmęczenie, jak i pełzanie są traktowane jako główny rodzaj uszkodzeń. Wielu badaczy podkreśla także istotny udział ich interakcji, szczególnie w warunkach podwyższonej temperatury [10,34]. W przypadku połączeń lutowanych, stosowanych w montażu w mikroelektronice, uszkodzenie na skutek pełzania i/lub zmęczenia materiału w wyniku obciążeń termomechanicznych odgrywa podstawową rolę w celu poprawnego prognozowaniu niezawodności [4,6,16].

2. Montaż i niezawodność połączeń lutowanych w mikroelektronice.

Pojęcie montażu w mikroelektronice związane jest z szeregiem czynności i etapów technologicznych prowadzących do wykonania funkcjonalnego urządzenia. Celem montażu jest zapewnienie dobrego połączenia elektrycznego, odpowiednich właściwości mechanicznych oraz transportu ciepła. W tym celu stosowane są różne techniki połączeń, takie jak zgrzewanie, lutowanie, klejenie. Wynikiem tej różnorodności jest m.in. podział technik montażu na poziomy [8,31]:

- zerowy: realizowany jest w fabryce i dotyczy montażu elementów składowych znajdujących się na powierzchni płytki półprzewodnikowej,
- pierwszy: wykonywanie połączeń wewnątrz układów scalonych oraz montaż struktury półprzewodnikowej w obudowie,
- drugi: montaż układów scalonych, biernych i czynnych komponentów elektronicznych na płytkach obwodów drukowanych,
- trzeci: montaż modułów w postaci pojedynczych płytek obwodów drukowanych oraz innych podzespołów w bloki funkcjonalne tworzące kompletne urządzenie.

Wszystkie rodzaje połączeń stosowane na poszczególnych poziomach muszą odznaczać się odpowiednią wytrzymałością w celu zapewnienia długotrwałego i sprawnego funkcjonowania urządzenia. W ramach przeprowadzonych badań podjęto się analizy problemu niezawodności połączeń lutowanych stosowanych na drugim poziomie montażu [12].

Celem pracy było przeprowadzenie badań wytrzymałości połączeń lutowanych stosowanych w montażu w mikroelektronice w wyniku złożonego profilu obciążeń, co pozwoliło na analizę typowych rodzajów uszkodzeń, tj. pełzania i zmęczenia oraz ich interakcji. Zastosowana metoda pomiarowa oraz otrzymane wyniki, zdaniem autorów, pozwolą na dokładniejsze prognozowanie wytrzymałości połączeń lutowanych w skali mikro, a tym samym na poprawę niezawodności komponentów mikroelektronicznych. Przedstawione wyniki badań dotyczą połączeń lutowanych wykonanych z wykorzystaniem tradycyjnego stopu lutowniczego $Sn_{63}Pb_{37}$. Należy jednak podkreślić, że dyrektywa unijna RoHS (z ang. Restriction of Hazardous Substances) ograniczyła, począwszy od roku 2006, możliwość wprowadzania do obrotu na terenie Unii Europejskiej sprzętu elektronicznego zawierającego materiały szkodliwe, np. ołowiowe stopy lutownicze. Niemniej, ta sama dyrektywa, zawiera także zbiór wyjątków, gdy zastąpienie danego pierwiastka jest trudne lub niemożliwe oraz w celach naukowo-badawczych. Tradycyjny ołowiowy stop lutowniczy $Sn_{63}Pb_{37}$ charakteryzuje się wyjątkowymi właściwościami termomechanicznymi i pozwala na trwałe i spójne łączenie elementów. Z tego powodu jest on nadal wykorzystywany w aplikacjach specjalistycznych, np. sprzęt medyczny, militarny, itp. Ponadto, ze względu na dużą wytrzymałość połączeń lutowanych wykonanych z wykorzystaniem tego stopu, jest on używany w celach porównawczych w ramach prowadzonych badań naukowych. Wspomniana dyrektywa nie obejmuje także zastosowań amatorskich.

Testy niezawodnościowe charakteryzują się tym, że są długotrwałe i kosztowne, głównie z tego powodu, że wyniki badań wymagają analizy statystycznej. Niewątpliwie zaletą takiego postępowania jest możliwość oszacowania bezawaryjnego czasu działania urządzenia, natomiast do wad można zaliczyć to, że:

- analiza statystyczna wymaga przeprowadzenie wielu testów jednostkowych, co wiąże się z długim czasem ich trwania, tj. od kilku do nawet kilkunastu miesięcy dla przypadku komponentów mikroelektronicznych,
- analiza pojedynczego czynnika uszkodzeń prowadzi do błędnych prognoz, ponieważ rzeczywiste warunki eksploatacji charakteryzują się obecnością występowania kilku rodzajów uszkodzeń i/lub ich interakcji.

Obecnie niewiele jest prac naukowych zawierających opis i wyniki badań wytrzymałościowych połączeń lutowanych w skali mikro na skutek interakcji uszkodzeń w wyniku złożonego profilu obciążeń [4,6,11]. Z tego powodu przedstawione wyniki wpisują się w ramy aktualnych badań dotyczących problemu występowania kilku typów uszkodzeń z uwzględnieniem skali analizy. Kluczowym problemem był pomiar niewielkich (rzędu ułamków mikrometra lub pojedynczych mikrometrów) przemieszczeń zachodzących w obrębie próbki i połączenia. Cyklicznie zmienna siła, przyłożona do połączenia, pozwalała na obserwację zarówno odkształceń sprężystych, jak i odkształceń niesprężystych.

2.1. Stopy lutownicze i połączenia lutowane w mikroelektronice.

Jak wspomniano wcześniej odnośnie do dyrektywy unijnej RoHS, stosowane obecnie w mikroelektronice stopy lutownicze uwarunkowane są głównie przez przepisy prawa stanowiące o zmniejszeniu ilości substancji niebezpiecznych, przenikających do środowiska. Niestety, zastąpienie tradycyjnych stopów ołowiowych stopami bezołowiowymi, wymaga zmian technologicznych, np. podwyższenie temperatury lutowania, przystosowanie linii montażowych do lutowania w wyższej temperaturze, dostosowanie montowanych elementów do montażu w wyższej temperaturze, wprowadzenie nowych topników umożliwiających odpowiednie zwilżanie lutowanych powierzchni stopami lutowniczymi o wyższym napięciu powierzchniowym itp. Należy jednak podkreślić, że tradycyjny stop lutowniczy cyna-ołów ($Sn_{63}Pb_{37}$) charakteryzuje się bardzo dobrymi parametrami termomechanicznymi. W przypadku stopów bezołowiowych stosuje się głównie stopy na bazie cyny, srebra i miedzi, w skrócie określane terminem SAC (SnAgCu), w różnych ilościach procentowych tych pierwiastków, np. SAC305 ($Sn_{96.5}Ag_{3.0}Cu_{0.5}$) [3,9].

Mechaniczne zachowanie połączenia lutowanego zależy od wielu czynników: mikrostruktury stopu, w tym rodzaju i zawartości związków międzymetalicznych, wielkości połączenia lub próbki, prędkości schładzania po utworzeniu połączenia, czy też od procesu starzenia. Innymi czynnikami istotnymi jest profil obciążeń termomechanicznych, rozrzut właściwości termomechanicznych materiałów, itp. Wieloletnie badania stopów i połączeń lutowanych w mikroelektronice przyczyniły się do powstania uproszczonych modeli charakteryzujących ich mechaniczne zachowanie oraz metod analizy z użyciem mechaniki klasycznej, czy też technik numerycznych. Kluczowym elementem tych badań są modele materiałowe stopów, które odpowiadają za matematyczny opis zachowania, w wyniku obciążeń termomechanicznych. W tym celu korzysta się zarówno z modeli materiałowych prostych, jak i złożonych. Istnieje również inne kryterium klasyfikacji, jakim jest liniowość, tj. modele materiałowe liniowe, np. model sprężysty oraz modele materiałowe nieliniowe, np. model plastyczny, lepki [33,35].

2.2. Niezawodność połączeń lutowanych w mikroelektronice.

Poprawne prognozowanie niezawodności połączeń lutowanych w montażu w mikroelektronice wymaga stosowania różnych technik i metod doświadczalnych, narzędzi numerycznych, jak również uwzględnienia różnych zjawisk fizycznych. Należy podkreślić, że w przypadku połączeń lutowanych w mikroelektronice ocenę niezawodności prowadzi się zazwyczaj dla wybranego pojedynczego rodzaju uszkodzenia, tj. głównie zjawiska zmęczenia lub pełzania [1,19,23]. Warto także wspomnieć o często pomijanym problemie w przypadku analizy niezawodności, którym jest występowanie tzw. naprężeń wbudowanych. Naprężenia wbudowane mogą być wynikiem przeprowadzonych procesów technologicznych. W celu zmniejszenia wpływu naprężeń wbudowanych, procesy technologiczne projektuje się tak, aby uwzględnić zjawisko relaksacji naprężeń, np. w wyniku powolnego schładzania lub dodatkowego procesu wygrzewania [40].

2.2.1. Zjawisko zmęczenia.

Zjawisko zmęczenia stopu lutowniczego opisuje wpływ cyklicznie zmiennych w czasie obciążeń termomechanicznych na wytrzymałość połączenia lutowanego. W konsekwencji dochodzi do odkształcenia, a następnie całkowitego zniszczenia lub znaczącego uszkodzenia połączenia. Cechą charakterystyczną zniszczenia zmęczeniowego jest to, że do uszkodzenia może dojść, przy naprężeniach znacznie niższych niż wynikałoby to z wytrzymałości materiału. Wytrzymałość zmęczeniowa mierzona jest w cyklach obciążeniowych, które mogą zostać przeliczone na czas. Istnieje szereg czynników mających wpływ na liczbę cykli: rodzaj obciążeń, ich wartość, sekwencja i czas trwania. Ponadto, w wypadku zjawiska zmęczenia wyróżnia się dwa rodzaje testów:

- zmęczenie niskocyklowe (najczęściej<1000), związane jest z zastosowaniem dużych wartości naprężeń, co prowadzi do powstania znacznych odkształceń niesprężystych podczas każdego cyklu, a tym samym krótkiej żywotności badanego materiału,
- zmęczenie wysokocyklowe (najczęściej>1000), związane jest z użyciem niewielkich wartości naprężeń, w wyniku czego odkształcenia ograniczają się do zakresu odkształceń sprężystych, co powoduje zwiększenie żywotności badanego materiału.

Przykładem popularnych testów, stosowanych w montażu w mikroelektronice, są badania niskocyklowe z wykorzystaniem empirycznego uogólnionego modelu Coffina-Mansona, który łączy rozpraszanie energii niesprężystej ΔW w połączeniu lutowanym z jego wytrzymałością zmęczeniową N_f:

$$\Delta \mathrm{W} = \mathrm{C} \cdot N_f^a \tag{1}$$

gdzie współczynniki C i a są stałymi materiałowymi, które szacuje się w wyniku badań doświadczalnych, natomiast wartość energii niesprężystej ΔW szacuje się na podstawie krzywej histerezy odkształcenie-naprężenie w wybranym cyklu zmęczeniowym [22,32].

2.2.2. Zjawisko pełzania.

Zjawisko pełzania stopu lutowniczego, nazywane również "zimnym przepływem", występuje w przypadku stałych lub zmiennych w czasie obciążeń termomechanicznych, czego wynikiem jest trwałe odkształcenie, określane także terminem odkształcenia pełzaniowego. Pełzanie może, ale nie musi prowadzić do zniszczenia materiału. Jeżeli połączenie nie jest w stanie zakumulować energii powstałej na skutek obciążenia, wówczas może ona zostać przekazana do pozostałych elementów połączenia lub może pojawić się odkształcenie nieodwracalne, np. pęknięcie. Zjawisko pełzania dotyczy praktycznie wszystkich materiałów, lecz w większości przypadków proces ten przebiega bardzo powoli w funkcji czasu i zależy od temperatury i naprężenia:

$$\varepsilon = f(t, T, \sigma) \tag{2}$$

gdzie ε to odkształcenie pełzaniowe, t to czas, T to temperatura, σ to naprężenie. W mechanice, np. dla metali i ich stopów, zjawisko pełzania obserwuje się, gdy wartość naprężenie przekracza granicę plastyczności lub poniżej tej wartości, tzw. pełzanie dyfuzyjne, gdy wartość temperatury homologicznej jest większa od 0,4 [7]:

$$T_h = \frac{T_o}{T_t} > 0{,}4 \tag{3}$$

gdzie T_h to temperatura homologiczna, T_o to temperatura otoczenia, T_t to temperatura topnienia materiału. Wartość temperatury topnienia dla stopu $Sn_{63}Pb_{37}$ wynosi 456 K, czyli wartość temperatury homologicznej dla stopu w temperaturze pokojowej wynosi ponad 0,6. Dla przypadku struktur bimateriałowych, np. połączenie lutowane, czynnikiem aktywującym proces pełzania jest naprężenie, będące wynikiem zmiany temperatury otoczenia i różnicy współczynników rozszerzalności cieplnej połączonych materiałów. Granica plastyczności dla badanego stopu $Sn_{63}Pb_{37}$ wynosi około 40 MPa, a wartość naprężenia dla połączenia lutowanego można oszacować z zależności:

$$\sigma = \Delta \alpha \, \Delta T \frac{a}{h} \tag{4}$$

gdzie σ to naprężenie, $\Delta\alpha$ to różnica współczynników rozszerzalności cieplnej połączonych materiałów, ΔT to zmiana temperatury, a to odległość od neutralnego punktu połączonych materiałów, h to grubość połączenia. Zjawisko pełzania prowadzi do zmiany parametrów badanego połączenia lutowanego w wyniku mechanizmów zachodzących we wnętrzu materiału, np. dyfuzji na granicy ziaren, pojawienie się dyslokacji, formowania pęknięć lub pustych przestrzeni [14]. Należy dodatkowo wspomnieć o dodatkowym czynniku mającym istotny wpływ na zjawisko pełzania dla połączeń lutowanych w elektronice, jakim jest wzrost lokalnej temperatury złącza w wyniku przepływającego prądu.

2.3. Problem kumulacji uszkodzeń.

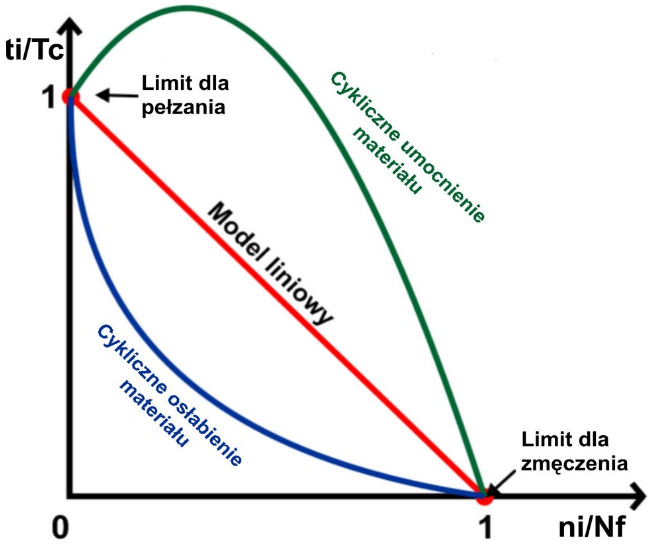
Jednym z istotnych problemów związanych z badaniem i analizą wytrzymałości połączeń lutowanych w montażu w mikroelektronice jest występowanie kilku rodzajów uszkodzeń równocześnie lub ich wzajemnej interakcji. Z tego powodu konieczne jest uwzględnienie problemu, które nosi nazwę kumulacji uszkodzeń. Jednym z najprostszych sposobów uwzględnienia zjawiska kumulacji uszkodzeń w praktyce inżynierskiej jest metoda zaproponowana przez Palmgrena-Minera [21]. Metoda ta polega na sumowaniu uszkodzeń cząstkowych, dla różnych cykli zmęczeniowych, i pozwala na wyznaczenie obciążenia zastępczego o stałej amplitudzie, które jest równoważne pod względem uszkodzenia obiektu dla przypadku obciążenia o zmiennej amplitudzie [28]:

$$\sum_{i=0}^{k} \frac{n_i}{N_i} = 1 \tag{5}$$

gdzie n_i jest liczbą cykli dla zadanej wartości obciążenia o zmiennej amplitudzie, N_i jest trwałością zmęczeniową dla zadanej wartości amplitudy obciążenia, natomiast k jest liczbą zadanych wartości amplitudy obciążenia. Niestety hipoteza ta zakład wiele uproszczeń, co powoduje, że wynik prognozowania niezawodności jest obarczony błędem. Dodatkowo hipoteza ta odnosi się do zjawisk makroskopowych, a nie uwzględnia zjawisk zachodzących w skali mikro, charakterystycznych dla montażu w mikroelektronice [11]. W przypadku jednoczesnego występowania dwóch różnych rodzajów uszkodzeń, czyli zmęczenia i pełzania, można skorzystać z liniowej zasady superpozycji. Zasada ta może być zapisana w postaci formuły matematycznej jako sumowania części ułamkowych dla obu zjawisk, przy zadanej amplitudzie obciążenia [18]:

$$\sum_{i=1}^{k} \frac{n_i}{N_f} + \sum_{i=1}^{k} \frac{\Delta t_i}{T_c} = 1 \tag{6}$$

gdzie Δt_i jest przedziałem czasu dla amplitudy obciążenia utrzymywanej na stałym poziomie, a T_c jest czasem do zniszczenia obiektu przy tej amplitudzie. Pierwszy człon równania jest częścią ułamkową dla cyklu zmęczeniowego, a drugi człon odpowiada za zjawisko pełzania. Podobnie jak dla hipotezy Palmgrena -Minera, zakłada się, że do uszkodzenia dojdzie, gdy wartość skumulowanego uszkodzenie będzie równa 1. W takim wypadku rozwiązaniem równania jest linia prosta, która reprezentuje skumulowaną niezawodność obiektu. Niestety, model ten zawiera wiele uproszczeń, a wyniki uzyskane przy jego pomocy zależą m.in. od właściwości materiału. Ponadto, hipoteza ta zakłada, że okresy Δt_i powstałe na skutek sił ściskających lub rozciągających mają podobny wpływ i nie uwzględniają problemu umocnienia czy osłabienia materiału. Nieliniowe modele kumulacji uszkodzeń dotyczą materiałów, których parametry mechaniczne czy fizyko-chemiczne zmieniają się w czasie trwania testów niezawodnościowych, co wymaga dodatkowych badań. Z tego też powodu model liniowy jest często wykorzystywany w zastosowaniach inżynierskich. Na rysunku 1 przedstawiono hipotezę liniowej kumulacji uszkodzeń oraz problem umocnienia i osłabienia materiału dla przypadku modeli nieliniowych [12].



Rys.1. Liniowy model kumulacji uszkodzeń oraz problem umocnienia i osłabienia materiału dla modeli nielinowych.

2.4. Analiza statystyczna.

Ocena niezawodności sprowadza się najczęściej do analizy przyczyn uszkodzeń oraz oceny prawdopodobieństwa ich wystąpienia w zadanych warunkach. W tym celu konieczna jest znajomość typowych obciążeń, mechanizmów uszkodzeń oraz kryteriów ich wystąpienia. Niestety, wyniki badań niezawodnościowych mają charakter losowy, w związku z czym ilościowy opis niezawodności wymaga zastosowania charakterystyk probabilistycznych oraz rozkładów prawdopodobieństwa. W teorii niezawodności, do opisu czasu poprawnej pracy obiektu, tzw. czasu zdatności, stosuje się rozkłady prawdopodobieństwa uszkodzeń w funkcji czasu *f(t)*. W dziedzinie montażu w mikroelektronice czas poprawnej pracy obiektu *t* opisuje się najczęściej dwuparametrowym rozkładem Weibulla:

$$f(t)=\frac{\beta}{\lambda}\cdot\left(\frac{t}{\lambda}\right)^{\beta-1}\cdot e^{-\left(\frac{t}{\lambda}\right)^{\beta}} \tag{7}$$

gdzie β jest parametrem kształtu, λ jest parametrem skali. Wartość prawdopodobieństwa, że obiekt ulegnie uszkodzeniu, do czasu *t* opisuje dystrybuanta *F(t)*:

$$F(t)=1-e^{-\left(\frac{t}{\lambda}\right)^{\beta}} \tag{8}$$

Celem badań niezawodnościowych jest oszacowanie wartości liczbowej parametrów rozkładu oraz wyznaczenie wybranych wskaźników niezawodności. W praktyce wygodnie jest posługiwać się pojęciem zdatności charakterystycznej obiektu odpowiadającej liczbowo wartości parametru skali λ niezależnie od wartości parametru kształtu β. Zdatność charakterystyczną wyznacza się przy założeniu, że $t=\lambda$, zatem:

$$F(\mathrm{t}=\lambda) \quad 0.632 \tag{9}$$

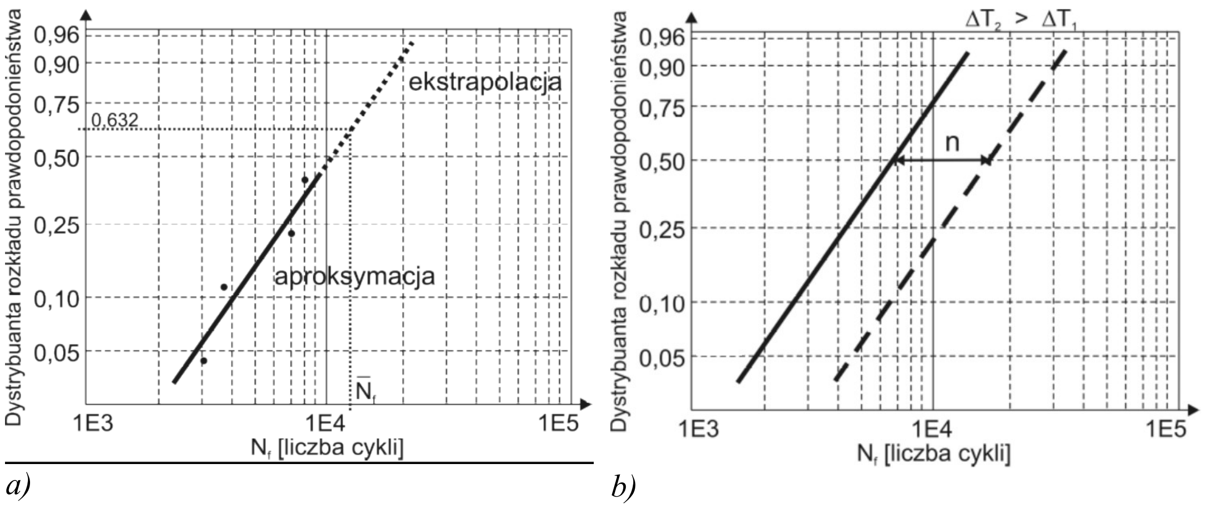
Często korzysta się także ze wskaźnika liczbowego opisującego niezawodności obiektów w połączeniu z jego wytrzymałością. Przykładowo zdatność charakterystyczną w odniesieniu do trwałości zmęczeniowej można wyznaczyć przy założeniu, że:

$$F(\bar{N}_f) \quad 0{,}632 \tag{10}$$

W praktyce zdatność charakterystyczną λ wyznacza się metodą graficzną, co pozwala na ograniczenie liczby testów doświadczalnych. Metoda ta pozwala dodatkowo na:

- wyznaczania kryterium uszkodzenia dla potrzeb projektowania numerycznego, tzn. wyznaczona z eksperymentu zdatność charakterystyczna jest przyjmowana w projektowaniu numerycznym jako kryterium uszkodzenia, co z kolei pozwala na zbudowanie empirycznych modeli uszkodzenia, np. modelu Coffina-Mansona,
- realizacji testów przyspieszonych, np. termicznych; w tym przypadku korzysta się z przesunięcia wzdłuż osi odciętych na wykresie Weibulla w zależności od amplitudy obciążenia, co pozwala na istotne skrócenie czasu realizacji testów niezawodnościowych.

Na rysunku 2 przedstawiono przykładowy wykres oraz graficzną metodę analizy wyników badań niezawodnościowych, w tym metodę wyznaczania wartości zdatności charakterystycznej i współczynnika przyspieszenia uszkodzenia *n* [37].



Rys.2. Przykład graficznej metody analizy wyników na podstawie rozkładu Weibulla: (a) aproksymacja i ekstrapolacja, (b) przyspieszone testy termiczne.

3. Opis zrealizowanych badań.

Jedną z najnowszych metod, która pozwala na badanie zachowań stopów lutowniczych w wyniku złożonego profilu obciążeń, jest metoda Hot Bump Pull HBP opracowana przez firmę Nordson Dage. Metoda ta polega na badaniu wytrzymałości połączeń uzyskanych na skutek zatopienia igły testowej w stopie lutowniczym i pozwala na badanie zjawiska pełzania i zmęczenia w skali mikro. Urządzenie umożliwia zastosowanie kilku rodzajów testów wytrzymałościowych: niszczący, zmęczeniowy, pełzaniowy oraz zmęczeniowo-pełzaniowy. W ramach przeprowadzonych badań skorzystano z testu zmęczeniowo-pełzaniowego, co pozwoliło na opracowanie odpowiedniej metody pomiarowej i algorytmu identyfikacji modelu prognozowanie niezawodności połączeń lutowanych w skali mikro w wyniku zmęczenia i pełzania oraz ich interakcji [5].

Badanie wytrzymałości różnych połączeń możliwe jest dzięki zastosowaniu rozwiązania w postaci wymiennych głowic testujących. Pozwalają one na badanie połączeń drutowych, kulkowych czy też klinowych z wykorzystaniem metody rozciągania, ściskania, ścinania lub zginania. Natomiast, dołączone do urządzenia oprogramowanie daje możliwość ustawienia szeregu parametrów testów, tj.:

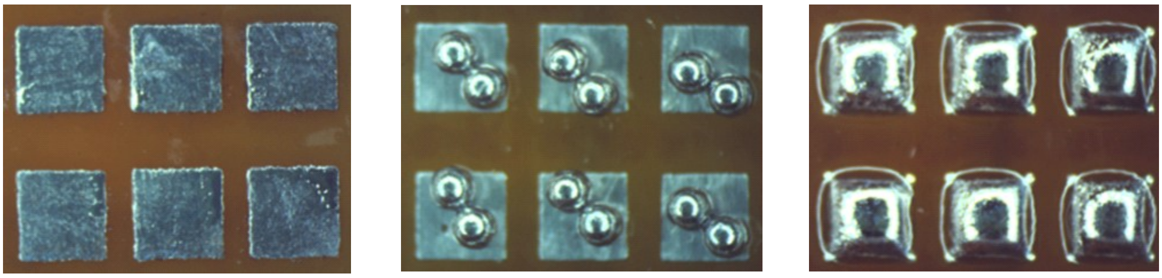
- maksymalnej i minimalnej wartości siły,
- czasu oczekiwania na poziomie maksymalnego i minimalnego narażenia,
- szybkość czasu narostu i opadania założonego profilu obciążenia,
- ustalanie i wybór zadanego profilu temperaturowego.

Zastosowana przez autorów technika pomiarowa polegała na badaniu wytrzymałości kulek stopu lutowniczego w technologii BGA za pomocą igły testowej wykonanej z miedzi [12]. Należy podkreślić, że wszystkie badania wykonano dla stopu lutowniczego $Sn_{63}Pb_{37}$, co stanowiło pierwszy etap zaplanowanych badań dotyczących niezawodności połączeń lutowanych w montażu w mikroelektronice.

3.1. Sposób wykonania próbek testowych, połączeń lutowanych i realizacji badań wytrzymałościowych.

Przed wykonaniem badań konieczne było zaprojektowanie i wykonanie próbek do testów w postaci połączenia lutowanego pomiędzy stopem lutowniczym a igłą testową. Na rysunku przedstawiono kolejne etapy wykonywanie próbek testowych:

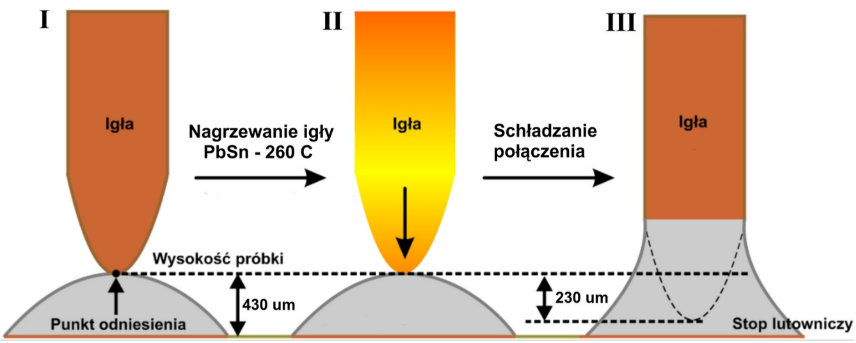
- podłoże próbek testowych w postaci wytrawionych na płytce szklano-epoksydowej (FR4) kwadratów miedzianych pokrytych cienką warstwą złota,
- podłoża pokryte cienką warstwą topnika, z umieszczonymi na powierzchni 2 kulkami stopu lutowniczego stosowanego w technologii BGA,
- próbki testowe powstałe na skutek ogrzania podłoża z kulkami stopu lutowniczego po ogrzaniu do temperatury przekraczającej temperaturę topnienia stopu lutowniczego.



Rys.3. Kolejne etapy wykonania próbek testowych.

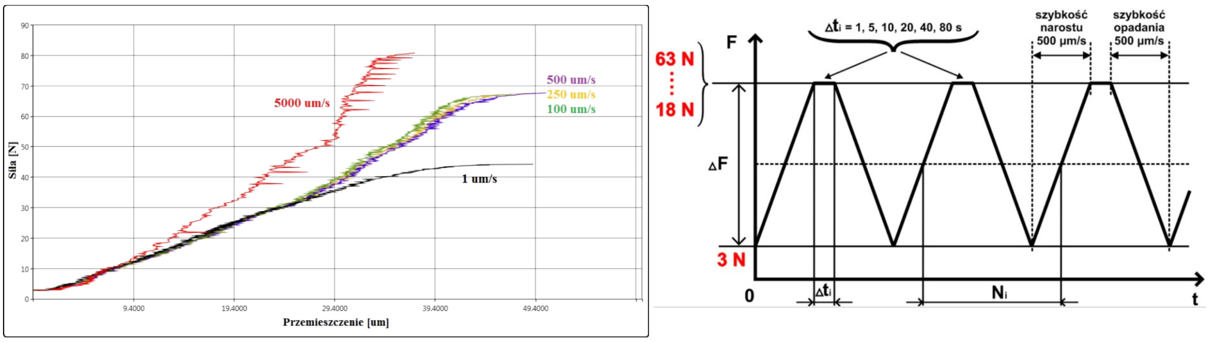
Po uzyskaniu odpowiednich próbek testowych i umieszczeniu ich w stoliku testowym wykonano w następnej kolejności połączenia. Na rysunku 4 pokazano kolejne etapy wykonywania połączenia lutowanego pomiędzy próbką testową a igłą miedzianą:

- w etapie I, po nałożeniu niewielkiej ilości topnika na igłę testową, ustawiono głębokość, do jakiej igła była opuszczona - połowa średniej wysokości próbki,
- w etapie II nagrzewano igłę, a następnie wtapiano ją w próbkę testową w tzw. czasie rozpływowym, w celu wykonania połączenia,
- w etapie III wykonane połączenie powoli schładzano, w celu zapobiegania, jak wcześniej napisano, powstawaniu niepożądanych naprężeń wbudowanych.



Rys.4. Etapy wykonywania połączenia pomiędzy igłą i próbką: (I) ustalenie punktu odniesienia i głębokości zatapiania igły, (II) zatapianie igły, (III) schładzanie próbki testowej.

W pierwszej kolejności wykonano testy wytrzymałościowe, które polegały na ciągnięciu igły z zadaną siłą i prędkością, aż do zniszczenia połączenia. W ten sposób wyznaczono zakres sił i prędkości przemieszczania igły testowej w celu wykonania testów wytrzymałościowych. Testy wykonano w temperaturze pokojowej. Na rysunku 5a przedstawiono wyniki przeprowadzonych testów. Otrzymane wyniki pokazują, że zbyt duża prędkość przemieszczania igły, w granicach *5000 μm/s*, powoduje uzyskanie wyższych wartości siły powodującej zniszczenie połączenia, co sugeruje, że stop zachowuje się jak materiał kruchy. Natomiast zbyt mała wartość prędkości przemieszczania igły, rzędu *1 μm/s*, charakteryzuje się występowaniem zjawiska pełzania, co jest niepożądane dla przypadku testów zmęczeniowych. Z tego powodu, w celu realizacji testów wytrzymałościowych, zdecydowano się na prędkość przemieszczania igły o wartości *500 μm/s*, która z jednej strony pozwalała na uniknięcie zjawiska utwardzanie materiału pod wpływem zbyt szybkiego przemieszczania igły, a z drugiej strony pozwalała na skrócenia czasu testowania. Skrócenie czasu testowania jest istotne w badaniach wymagających analizy statystycznej, co wiąże się z koniecznością realizacji dużej ilości testów jednostkowych.



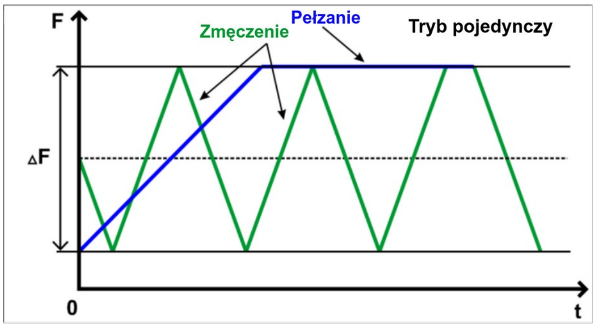
a) b)

Rys. 5. Wyniki testów wytrzymałościowych dla różnych prędkości przemieszczania igły testowej (a) oraz przyjęty do badań profil obciążenia dla testów wytrzymałościowych (b).

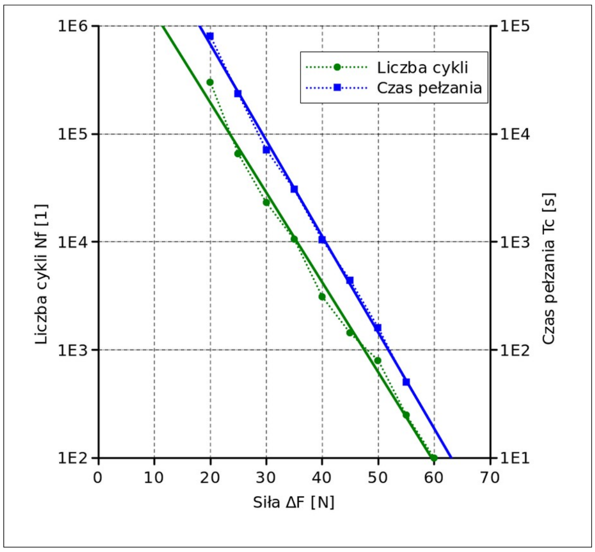
W drugiej kolejności wykonano testy wytrzymałościowe, a przyjęty profil obciążeń i założone parametry testów przedstawiono na rysunku 5b. Profil ten składa się z części zmęczeniowej oraz pełzaniowej. Część pełzaniowa zależała od tzw. czasu przytrzymania Δt_i, który zmieniał się w zakresie od *1* do *80 s*. Amplituda siły ΔF zmieniała się w zakresie od *18* do *63 N*. Przyjęto jednocześnie wartość początkowa siły *F* równą *3 N*.

3.2. Testy wytrzymałościowe.

W pierwszej kolejności wykonano testy wytrzymałościowe dla pojedynczego rodzaju uszkodzenia, tj. kolejno dla zmęczenia i pełzania. Badania te pozwoliły na określenie maksymalnej liczby cykli *N* oraz czasu pełzania *t* do zniszczenia połączenie przy zadanej amplitudzie zmian siły ΔF. Profil założonego obciążenia i uzyskane wyniki przedstawiono na rysunku 6.



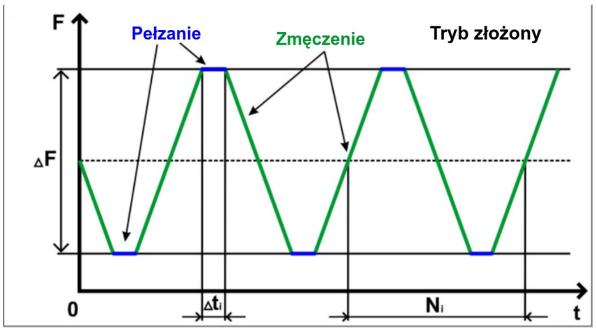
a)



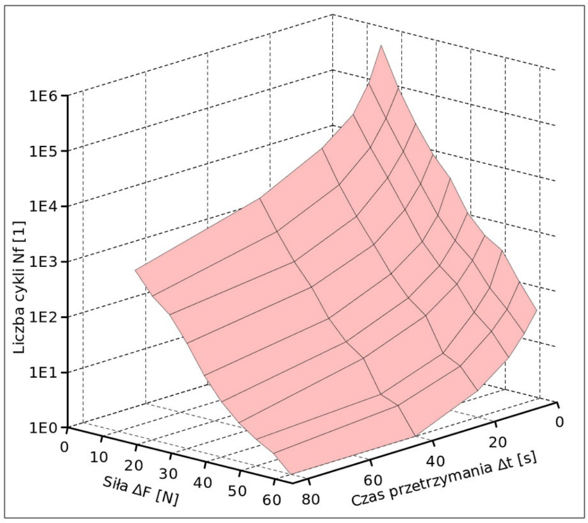
b)

Rys. 6. Profil obciążenia (a) oraz wyniki testów kolejno dla wybranego pojedynczego rodzaju uszkodzenia, tj. pełzania i zmęczenia (b).

Zgodnie z uzyskanymi wynikami, oba rodzaje uszkodzenia tj. zmęczenia i pełzania, zależą wykładniczo od wartości amplitudy siły ΔF. Uzyskane wyniki pozwoliły dobrać parametry dla testów zmęczeniowo-pełzaniowych, tj. głównie wartości czasów przetrzymania Δt_i. Na rysunku 7a przedstawiono profil obciążenia złożonego i wyniki testów wytrzymałościowych.



a)



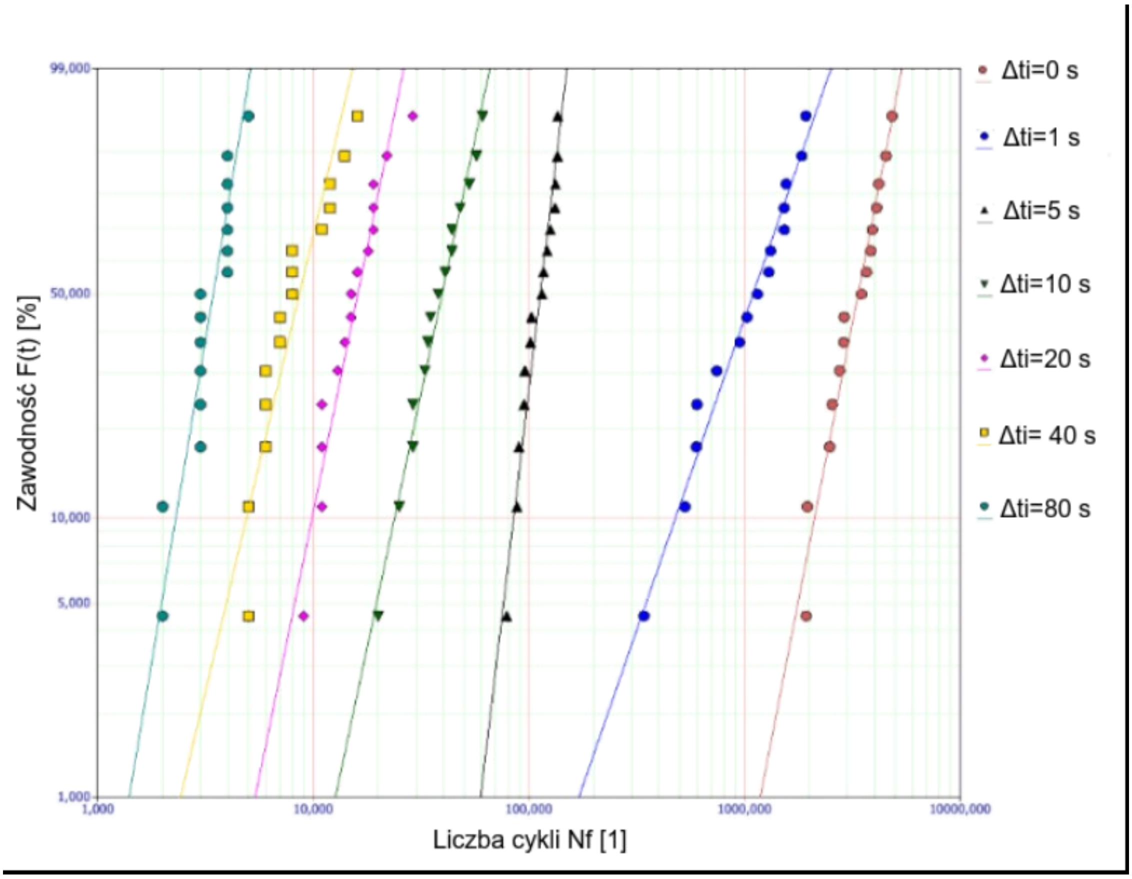
b)

Rys. 7. Profil obciążenia złożonego (a) oraz wyniki testów pełzaniowo-zmęczeniowych (b).

Zgodnie z uzyskanymi wynikami (Rys. 7b) można stwierdzić, że liczba cykli zmęczeniowych dla uszkodzenia pełzaniowo-zmęczeniowego zależy nie tylko od amplitudy siły obciążania ΔF, ale także od interakcji obu rodzajów uszkodzenia, tj. zmęczenia i pełzania. Przy czym, udział pełzanie w znaczący sposób zmniejsza liczbę cykli zmęczeniowych N_f.

3.3. Badania statystyczne.

Badania niezawodności mają charakter losowy i z tego powodu, jak napisano wcześniej, wymagają analizy statystycznej. W tym celu najczęściej korzysta się z dwuparametrowego rozkładu Weibulla. Niestety szczegółowa analiza statystyczna Weibulla wymaga dużej ilości testów jednostkowych. Na rysunku 8 przedstawiono wyniki dla wybranej wartości amplitudy siły obciążenia ΔF równej *40 N* – łączna liczba testów jednostkowych wynosiła w tym przypadku ponad 100 [26].

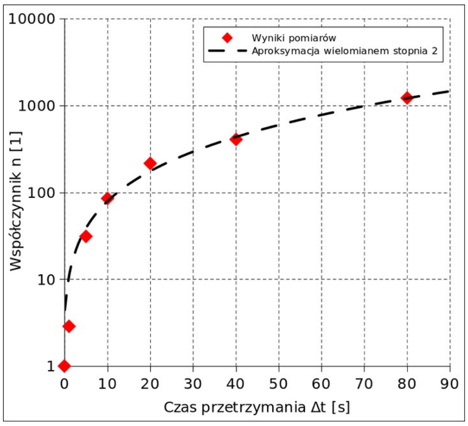


Rys. 8. Analiza statystyczna Weibulla dla wyników testów niezawodnościowych w trybie pojedynczym oraz złożonym dla wartości amplitudy siły ΔF równej 40 N.

Na podstawie uzyskanych wyników analizy statystycznej wyznaczono współczynniki dwuparametrowego rozkładu Weibulla, tzn. współczynnik kształtu β oraz współczynnik skali λ. Wyniki analizy przedstawiono w tabeli na rysunku 9a. Oszacowano także współczynnik przyspieszenia uszkodzenia n, opisującego zmniejszenie liczby cykli zmęczeniowych N_f w wyniku udziału czasu przytrzymania Δt_i. Wyniki przedstawiono na rysunku 9b.

	Zmęczenie - cykle [1]	Czas przetrzymania Δt_i [s]						Pełzanie - czas [s]
		1	*5*	*10*	*20*	*40*	*80*	
β	4,05	2,28	6,64	3,72	3,86	3,33	4,72	3,16
λ	3673	1283	118	43	17	9	3	271

a)

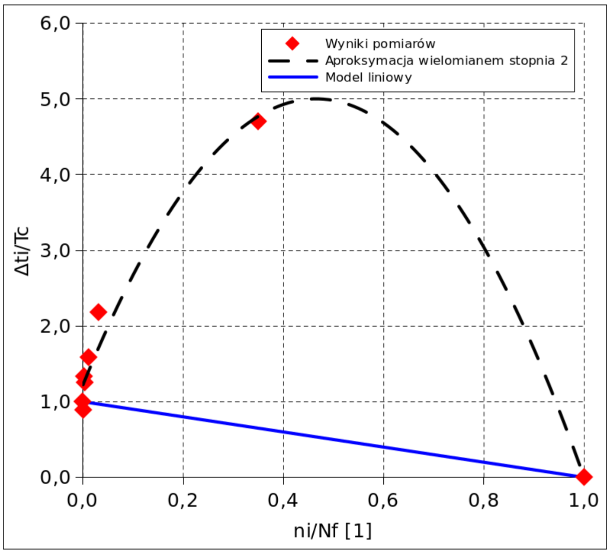


b)

Rys.9. Wyniki analizy statystycznej dla amplitudy siły ΔF równej 40 N: a) współczynniki rozkładu Weibulla, tj. współczynnik kształtu β oraz współczynnik skali λ, b) wartości współczynnika przyspieszenia uszkodzenia n.

3.4. Model kumulacji uszkodzeń.

Jak wspomniano wcześniej, celem badań była analiza wytrzymałości z wykorzystaniem testów pełzaniowo-zmęczeniowych dla połączeń lutowanych w skali mikro, typowych dla montażu w mikroelektronice. Na rysunku 10 przedstawiono model kumulacji uszkodzeń uzyskany na podstawie wyników z przeprowadzonych testów doświadczalnych i analizy statystycznej.



Rys.10. Model kumulacji uszkodzeń uzyskany na podstawie wyników testów doświadczalnych i analizy statystycznej.

Otrzymane wyniki badań sugerują, że stop $Sn_{63}Pb_{37}$ analizowany pod kątem występowania interakcji analizowanych uszkodzeń odznacza się cyklicznym utwardzaniem materiału, co znacznie skraca trwałość połączenia. Materiał z każdym cyklem zmęczeniowym staje się bardziej kruchy i tym samym podatny na powstanie pęknięć, które prowadzą do zniszczenia połączenia lutowanego. Zgodnie z wynikiem przedstawionym na rysunku 10 można stwierdzić, że uzyskany model kumulacji uszkodzeń ma charakter nieliniowy i znacznie odbiega od krzywej teoretycznej opisującej model liniowy.

4. Podsumowanie.

Artykuł zawiera wyniki pierwszego etapu badań dotyczących analizy wytrzymałości połączeń lutowanych stosowanych w montażu w mikroelektronice w wyniku złożonego profilu obciążeń. Zaproponowany profil obciążeń pozwolił na analizę typowych uszkodzeń dla stopów lutowniczych, tj. zmęczenia i pełzania oraz dodatkowo interakcji obu tych uszkodzeń. W ramach przeprowadzonych badań skorzystano z testu zmęczeniowo-pełzaniowego oraz metody Hot Bump Pull HBP opracowanej przez firmę Nordson Dage. W ramach przedstawionych badań wykonano analizę wytrzymałości połączeń lutowanych dla tradycyjnego stopu lutowniczego $Sn_{63}Pb_{37}$, w temperaturze pokojowej, a otrzymane wyniki przedstawiono w postaci:

- analizy wytrzymałości dla pojedynczych uszkodzeń oraz ich interakcji,
- analizy statystycznej dla złożonego profilu obciążeń,
- porównania liniowego modelu kumulacji uszkodzeń z modelem otrzymanym w wyniku przeprowadzonych testów wytrzymałościowych.

Wnioski, które można sformułować na podstawie zaprezentowanych wyników, to:

- zaproponowana metodologia badań wytrzymałości stopów lutowniczych pozwala na uwzględnienia zjawiska interakcji typowych rodzajów uszkodzeń, tj. pełzania i zmęczenia,
- zjawisko pełzania dla badanego stopu lutowniczego w skali mikro odgrywa istotną rolę w testach zmęczeniowo-pełzaniowych już w temperaturze pokojowej, w której przeprowadzone zostały testy wytrzymałościowe,
- popularny wśród inżynierów liniowy model kumulacji uszkodzeń nie jest odpowiedni w celu prognozowania wytrzymałości połączeń lutowanych w mikroelektronice, dla testów zmęczeniowo-pełzaniowych.

Należy podkreślić, że analiza pojedynczego czynnika uszkodzeń prowadzi do błędnych prognoz, ponieważ rzeczywiste warunki eksploatacji komponentów elektronicznych charakteryzują się obecnością występowania kilku rodzajów uszkodzeń, czego przyczyną jest złożony profil obciążeń środowiskowych. Biorąc pod uwagę powyższe, kolejne etapy badań powinny obejmować:

- zastosowanie innego modelu kumulacji uszkodzeń dla połączeń lutowanych w mikroelektronice dla testów zmęczeniowo-pełzaniowych, tj. modelu nieliniowego lub stosunkowo prostego do zastosowania w praktyce inżynierskiej modelu dwuliniowego [18],
- przeprowadzenie badań porównawczych dla stopów bezołowiowych oraz innych temperatur otoczenia,
- opracowanie kryterium oceny wytrzymałości, co pozwoliłoby na wiarygodne prognozowanie wytrzymałości połączeń lutowanych w montażu w mikroelektronice w wyniku zmęczenia i pełzania na skutek złożonego profilu obciążeń.

Na zakończenie można stwierdzić, że zaproponowana metodologia badań oraz przedstawione wyniki analizy wytrzymałości połączeń lutowanych w skali mikro pozwolą na precyzyjne prognozowanie niezawodności współczesnych komponentów mikroelektronicznych i/lub realizację przyspieszonych badań niezawodnościowych.

Literatura

[1] Arra M, Shangguan D, Ristolainen E. Characterization of mechanical performance of Sn/Ag/Cu solder joints with different component lead coatings. Soldering and Surface Mount Technology 2004; 16(1): 35-43.

[2] Bisschop J. Failure mechanisms in plastic package IC’s. Proceedings of 2nd EuroSimE Conference 2002; 328-331.
[3] Bukat K, Hackiewicz H. Lutowanie bezołowiowe, Warszawa: Wydawnictwo BTC, 2007.
[4] Chicot D, Tilkin K, Jankowski K, Wymysłowski A. Reliability analysis of solder joints due to creep and fatigue in microelectronic packaging using microindentation technique, Microelectronics Reliability 2013; 53(5): 761–766.
[5] DAGE UK Hot Bump Pull/Hot Pin Pull, 2011.
[6] Dasgupta A, Oyan C, Barker D, Pecht M. Solder Creep-Fatigue Analyses by an Energy Partitioning, Journal of Electronic Packaging 1991; 114(2): 152-160.
[7] Dowling NE, Kampe SL, Kral MV. Mechanical Behavior of Materials, Pearson Prentice Hall, 2018.
[8] Felba J. Montaż w Elektronice, Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 2010.
[9] Ganesan S. Lead-free Electronics, John Wiley & Sons Inc., 2006.
[10] Holdsworth SR. Component assessment data requirements from creep-fatigue tests, ASTM International 2011; 1–14.
[11] Jankowski K, Wymysłowski A. Life prediction of lead alloy based on multi-failure criteria. 18th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems 2017; 1-5.
[12] Jankowski K, Wymysłowski A, Chicot D. Combined loading and failure analysis of lead-free solder joints due to creep and fatigue phenomena, Soldering & Surface Mount Technology 2014; 26(1): 1-6.
[13] Judd M, Brindley K. Soldering in Electronics Assembly, Elsevier Ltd., 1999.
[14] Kim DH. Reliability Study of SnPb and SnAg Solder Joints in PBGA Packages, The University of Texas at Austin, 2007.
[15] Kisiel R. Podstawy technologii montażu dla elektroników, Wydawnictwo BTC, 2012.
[16] Lau JH. Solder Joint Reliability, Boston MA: Springer, 1991.
[17] Lau JH. Thermal Stress and Strain in Microelectronic Packaging, US Springer, 1993.
[18] Manson SS, Halford GR. Practical implementation of the double linear damage rule and damage curve approach for treating cumulative fatigue damage, International Journal of Fracture 1981; 17: 169–192.
[19] Matkowski PK, Dudek R, Kreyssig K. Reliability testing of SnAgCu solder joints under combined loading. Micromaterials and Nanomaterials 2011; 13: 86-91.
[20] Miller D, Priest RH, Ellison EG. A review of material response and life prediction techniques, Materials 2015; 8(11): 7757–7769.
[21] Miner MA. Cumulative Damage in Fatigue, Journal of Applied Mechanics 1945; 12 (3): 159-164.
[22] Nelson W. Accelerated Testing, New York: Wiley, 1990.
[23] Nguyen J, et al. Backward compatibility study of lead free area array packages with tin-lead soldering process, IPC - Printed Circuits Expo and the Designers Summit 2006.
[24] Numerical Analysis Guidelines for Microelectronics Packaging Design and Reliability, IPC/JEDEC-9301: 2018.
[25] Qu J. Thermomechanical Reliability of Microelectronic Packaging. Comprehensive Structural Integrity 2007; 8: 219-239.
[26] ReliaSoft, https://www.reliasoft.com/, 2019.
[27] Seraphim DP, Lasky R, Li CY. Principles of Electronic Packaging, New York: McGraw-Hill, 1989.

[28] Stone D, Wilson R, Li CY. The Mechanisms of Damage Accumulation in Solders during Thermal Fatigue. Proccedings of 36th Electronic Components Conference 1986: 630–635.
[29] Suhir E. Interfacial stresses in biomaterial thermostats, Journal of Applied Mechanics 1989; 56: 595-600.
[30] Timoshenko S. Analysis of bi-metal thermostats, Journal of the Optical Society of America 1925: 233-255.
[31] Tummala RR. Fundamentals of Microsystems Packaging, McGraw Hill, 2001.
[32] Tummala RR, Rymaszewski EJ, Klopfenstein AG. Microelectronics Packaging Handbook, Springer Science+Business Media Dordrecht, 1997.
[33] Van Driel WD. Virtual Thermo-Mechanical Prototyping of Microelectronics Devices, The Netherlands: PrintPartners Ipskamp, 2007
[34] Wareing J. Mechanisms of high temperature fatigue and creep-fatigue failure in engineering materials, United Kingdom: Applied Science, 1983.
[35] Wymysłowski A. Numeryczne metody projektowania termomechanicznego w montażu elektronicznym. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 2007.
[36] Wymysłowski A, Dowhań Ł. Application of nanoindentation technique for investigation of elasto-plastic properties of the selected thin film materials, Microelectronics Reliability 2013; 53(3): 443-451.
[37] Zahavi E. Fatigue design - life expectancy of Machine Parts, A Solomon Press Book, 1996.
[38] Zhang GQ. The challenges of virtual prototyping and qualification for future microelectronics, Microelectronics Reliability 2003; 43: 1777-1785.
[39] Zhang GQ, Ernst LJ, de Saint Leger O. Benefiting from thermal and mechanical simulation in microelectronics, Dordrecht: Kluwer Academic Publishers, 2000.
[40] Zhang GQ, van Driel WD, Fan XJ. Mechanics of Microelectronics, Dordrecht: Springer, 2006.
[41] Zienkiewicz OC, Taylor R, Zhu JZ. The Finite Element Method: Its Basis and Fundamentals, Elsevier Butterworth-Heinemann, 2005.

Dr hab. inż. Andrzej Puchalski, prof. UTH[1]
Dr hab. inż. Iwona Komorska, prof. UTH[1]
Dr hab. inż. Marcin Ślęzak, prof. ITS[2]
Prof. dr hab. inż. Andrzej Niewczas[2]

[1] Uniwersytet Technologiczno-Humanistyczny w Radomiu
ul. Malczewskiego 29, 26-600 Radom, Polska
[2] Instytut Transportu Samochodowego
ul. Jagiellońska 80, 03-301 Warszawa, Polska

E-mail: andrzej.puchalski@uthrad.pl
E-mail: iwona.komorska@uthrad.pl
E-mail: marcin.slezak@its.waw.pl
E-mail: andrzej.niewczas@its.waw.pl

Synteza eksploatacyjnych cykli jezdnych samochodów przy wykorzystaniu metody Monte Carlo z zastosowaniem łańcuchów Markowa

Synthesis of maintenance vehicle driving cycles using Monte Carlo method of Markov chains

Słowa kluczowe: *eksploatacyjne cykle jezdne samochodu, synteza cykli jezdnych, modele Markowa, symulacja Monte Carlo*

Streszczenie: Metody symulacyjne powszechnie stosowane w całym procesie projektowania i weryfikacji różnych typów pojazdów mechanicznych wymagają opracowania eksploatacyjnych cykli jezdnych. Optymalizacja parametrów, testowanie i stopniowe zwiększanie stopnia autonomiczności pojazdów nie jest możliwe na bazie standardowych cykli jezdnych. Zapewnienie reprezentatywności syntezowanych szeregów czasowych na podstawie zgromadzonych baz danych wymaga algorytmów wykorzystujących techniki bazujące na modelach stochastycznych i statystycznych. Zaproponowano i zweryfikowano technikę syntezy łączącą metodę Monte Carlo wykorzystującą łańcuch Markowa (MCMC) oraz analizę multifraktalną. Metoda umożliwia proste wyznaczenie profilu prędkości jazdy w porównaniu do klasycznej analizy częstotliwościowej.

Keywords: *vehicle maintenance driving cycles, synthesis of driving cycles, Markov models, Monte Carlo simulation*

Abstract: Simulation methods commonly used throughout the entire design and verification process of various types of motor vehicles require development of naturalistic driving cycles. Optimization of parameters, testing and gradual increase in the degree of autonomy of vehicles is not possible on the basis of standard driving cycles. Ensuring representativeness of synthesized time series based on collected databases requires algorithms using techniques based on stochastic and statistical models. A synthesis technique combining the MCMC method and multifractal analysis has been proposed and verified. The method allows simple determination of the speed profile compared to classic frequency analysis.

1. **Wprowadzenie**

Współczesne metody projektowania i testowania pojazdów mechanicznych bazują na technikach symulacyjnych wymagających stosowania precyzyjnych modeli dynamiki pionowej i poziomej oraz zjawisk o charakterze sekwencji zdarzeń losowych występujących w warunkach ruchu drogowego. Problem ten jest obecny w procesie optymalizacji parametrów pojazdów zarówno z silnikami spalinowymi jak i pojazdów elektrycznych (EV) oraz hybrydowych (HEV). Wybór najbardziej odpowiedniej architektury układu napędowego dla danej klasy pojazdu i cyklu jazdy wymaga optymalizacji wielkości komponentów według funkcji kosztów takich jak najniższa emisja CO_2, najniższa masa, oszczędność paliwa lub dowolna kombinacja tych atrybutów w architekturze [1, 7, 9, 18].

Niezależnie od zastosowanej techniki symulacji, quasi-statycznej wykorzystującej model pojazdu typu „Backward-facing" czy symulacji dynamiczej z modelem „Forward-facing", wymagana jest znajomość reprezentatywnego cyklu jezdnego. W pierwszym przypadku, dla układu otwartego, szereg czasowy opisujący prędkość jest wymuszeniem podawanym na wejście modelu pojazdu w celu obliczenia prędkości obrotowej i momentu obrotowego na kołach. Natomiast w układzie zamkniętym modelu pojazdu, cykl jezdny stanowi wartość zadaną dla bloku kierowcy, którego funkcją jest generacja odpowiedniego momentu obrotowego silnika. Ograniczenia terminowe i kosztowe związane z projektowaniem i testowaniem różnych możliwych architektur pojazdów wymagają metod syntezy cykli jezdnych, które spełnią wymagania inżynierów motoryzacyjnych podczas całego procesu badań i rozwoju w zakresie modelowania i symulacji. Optymalizacja parametrów i stopniowe zwiększanie stopnia autonomiczności pojazdów nie jest możliwe na bazie standardowych cykli jezdnych, nie zapobiega sytuacjom określanym jako „cycle beating". Osiągnięcie reprezentatywności syntezowanych szeregów czasowych na podstawie zgromadzonych baz danych wymaga algorytmów wykorzystujących techniki bazujące na modelach stochastycznych i statystycznych [6, 19]. Proces syntezy, dla zdefiniowania kryteriów równoważności, zamyka sprawdzanie poprawności wyników, tj. każdego wygenerowanego cyklu, za pomocą analizy statystycznej w dziedzinie czasu lub częstotliwości. Często stosuje się kombinację wielu kryteriów [2, 4].

Metody budowy cykli jezdnych wymagają kwantyzacji parametrów ruchu. W zależności od ich przeznaczenia (szacowanie poziomu emisji, szacowanie zużycia paliwa, do celów inżynierii ruchu itd.) syntezę zdefiniowanych stanów można przeprowadzić dla kategorii mikro-podroży (ang. micro-trips), segmentów jezdnych (ang. segments), niejednorodnych klas ruchu (ang. heterogeneous classes) lub cykli modalnych (ang. modal cycles) [17]. Mikro-podróże to modele jezdne pomiędzy postojami obejmujące okresy bezczynności. Sygnalizacje drogowe i przeciążenia reprezentują wzorce jazdy "stop-go" i prowadzą do zwiększonego zużycia paliwa. Micro-podróże dobrze odzwierciedlają zużycie paliwa i emisję. Segmenty jezdne modelują sytuacje dla różnych typów dróg i warunków jazdy klasyfikowanych np. poprzez LOS (ang. Level of Service). Mogą zaczynać się i kończyć różnymi parametrami jazdy, dlatego ich łączenie przy syntezie cyklu wymaga odpowiedniego dopasowania prędkości i przyspieszeń. Cykle jezdne na bazie niejednorodnych klas ruchu, uzyskanych w wyniku statystycznego podziału danych, są konstruowane jako kinematyczne sekwencje przy wykorzystaniu metod probabilistycznych i analizy rozkładów prawdopodobieństwa. Metoda nie jest ukierunkowana na badania emisji i zużycia paliwa. Cykle modalne obejmują zarejestrowane parametry ruchu pojazdów dla określonych przedziałów przyspieszeń, ze stałą prędkością czy na biegu jałowym. W procedurach wykorzystujących teorię procesów stochastycznych przy analizie równań dynamiki ruchu pojazdu, reprezentowanej przez prędkość oraz przyspieszenie, główny kierunek ostatnio prowadzonych badań obejmuje metody bazujące na teorii łańcuchów Markowa [8, 11]. Podjęto także próby z trójwymiarowymi modelami Markowa w procesie syntezy cykli jezdnych, uwzględniających dodatkowo nachylenie drogi [20]. Metody bazujące na wielowymiarowych łańcuchach Markowa umożliwiają realistyczną ocenę zużycia paliwa i emisji CO_2, nawet po dokonaniu kompresji czasowej syntezowanych szeregów czasowych [5]. Charakteryzują się jednak dużym kosztem czasowym symulacji.

W artykule zaproponowano metodę syntezy eksploatacyjnych cykli jezdnych pozwalającą na zastąpienie informacji o chwilowych wartościach przyspieszenia przez poziom multifraktalności oceniany przy wykorzystaniu formalizmu bazującego na analizie liderów falkowych. W ten sposób uzyskano zmniejszenie liczby wymiarów łańcucha Markowa w procesie symulacji. Proces zilustrowano na przykładzie algorytmu Monte Carlo wykorzystującego łańcuch Markowa (ang. MCMC) sygnału prędkości samochodu. Dane wejściowe algorytmu zarejestrowano w czasie serii eksperymentów w warunkach rzeczywistych. Przy selekcji i klasyfikacji modeli ruchu drogowego, równoważnych warunkom rzeczywistym, wykorzystano wskaźniki statystyczne oraz średnią siłę trakcyjną MTF (ang. Mean Tractive Force).

2. Formalizm multifraktalny liderów falkowych w symulacji MCMC

Każdy rzeczywisty cykl jezdny można traktować jak sekwencję losowych przejść między zdefiniowanymi m-stanami pojazdu występującymi w warunkach ruchu drogowego. Częstotliwość występowania określonych stanów jest konsekwencją zarówno parametrów technicznych samochodu jak i intensywności ruchu drogowego oraz zachowania kierowcy. Wyznaczając prawdopodobieństwa pozostania lub przejścia do innego stanu uzyskujemy obraz badanego zjawiska w postaci macierzy prawdopodobieństw przejść TPM (1) (Transition Probability Matrix):

$$P = \begin{bmatrix} P_{11} & \cdots & P_{1m} \\ \vdots & \ddots & \vdots \\ P_{m1} & \cdots & P_{mm} \end{bmatrix} \in R^{mxm} \quad , \tag{1}$$

gdzie wyraz P_{ij} (2) przyjmuje wartość prawdopodobieństwa przejścia ze stanu i do stanu j, gdy $j \neq i$ lub pozostania w stanie i, gdy $j = i$. Prawdopodobieństwo P_{ij} określa następująca zależność:

$$P_{ij} = \frac{N_{ij}}{\sum_j N_{ij}} \quad , \tag{2}$$

gdzie N_{ij} to liczba przejść ze stanu i do stanu j. Suma wartości wyrazów w każdym wierszu równa się jedności. Proces losowy $\{X_n\}_{n \in N}$ nazywamy łańcuchem Markowa, jeżeli dla dowolnego $n \in N$ zachodzi równość $\mathrm{P}\{X_{n+1}|X_n\} = \mathrm{P}\{X_{n+1}|X_0, X_1, \dots, X_n\}$. Przyjęta stacjonarność macierzy TPM implikuje jednorodność łańcuchu Markowa. Zatem dla łańcuchów Markowa rozkład warunkowy zmiennej losowej X_{n+1} zależy tylko od aktualnie znanej wartości X_n. Biorąc pod uwagę aktualny stan jazdy, przyszły stan można zatem wyznaczyć za pomocą symulacji Monte Carlo na bazie macierzy prawdopodobieństwa przejść. Możliwe jest wygenerowanie realizacji testu jezdnego dowolnej długości, co może być wykorzystane w procesie poszukiwania cyklu o wymaganym czasie trwania, przy założonych kryteriach równoważności.

Synteza cyklu jezdnego metodą MCMC, w której oprócz sygnału prędkości bierzemy pod uwagę inne parametry wymaga wielowymiarowego opisu zdefiniowanych stanów pojazdu, co w sposób znaczący komplikuje wyznaczanie macierzy prawdopodobieństwa przejść i wydłuża czas realizacji algorytmu. Jeżeli drugim parametrem jest przyspieszenie, które w większości rzeczywistych cykli jezdnych nie jest mierzone bezpośrednio, to uzyskanie informacji o dynamice ruchu wymaga różniczkowania sygnału prędkości. W takiej sytuacji standardowy, 1-sekundowy, okres próbkowania szeregu czasowego prędkości nie gwarantuje wystarczającej dokładności sygnału przyspieszenia.

Prace, w których analizowano ruch drogowy na podstawie rejestrowanych sygnałów prędkości pojazdów, wskazują na multifraktalny charakter jego dynamiki [3, 16]. Multifraktalność można także zaobserwować zarówno w rzeczywistych jak i w standardowych cyklach jezdnych [~~12~~, 14]. W podjętych badaniach zaproponowano wyeliminowanie sygnału przyspieszenia z wielowymiarowego opisu stanów pojazdu przy wykorzystaniu informacji o dynamice jazdy odwzorowanej za pomocą parametrów multifraktalnych sygnału prędkości. Iterację w tracie symulacji Monte Carlo prowadzono dla określonego czasu, przy wymaganiu dotyczącym dynamiki jazdy. Analiza multifraktalna, która bazuje na oszacowanych wykładnikach skalowania sygnału, jest popularnym narzędziem statystycznym do oceny danych empirycznych. W przypadku szeregów czasowych formalizm matematyczny opierał się początkowo na przyrostach ich wartości, których miarą są punktowe wykładniki Holdera h funkcji czasu $x(t)$ w punkcie t_0 , wyznaczone przez supremum wszystkich wykładników spełniających, dla stałej $C > 0$, warunek: $|x(t) - P_n(t - t_0)| \leq C|t - t_0|^h$, gdzie $P_n(t - t_0)$ jest wielomianem rzędu $n < h$ [13, 15, 16]. Wynikiem algorytmu jest widmo multifraktalne $D(h)$, tj. funkcja opisująca wymiary fraktalne punktów, które mają ten sam wykładnik Holdera.

Wykorzystany w badaniach formalizm multifraktalny w dziedzinie czasowo-częstotliwościowej pozwala na szacowanie parametrów multifraktalnych za pomocą liderów falkowych, które są reprezentantami lokalnych wykładników Holdera sygnału. Algorytm wykazuje niskie koszty obliczeniowe, stabilność numeryczną i dużą uniwersalność w zakresie sygnałów rzeczywistych. Dla współczynników (3) dyskretnej transformaty falkowej (ang. DWT) funkcji $x(t)$ i falki podstawowej o zwartym nośniku $\psi_0(t)$:

$$d_x(j,k) = \int_R \; x(t) 2^{-j} \psi_0\left(2^{-j}t - k\right) dt \, , \tag{3}$$

lidery falkowe (4), dla zbioru największych współczynników $d_x(j', k') \equiv d_{\lambda'}$ w sąsiedztwie 3λ , definiuje w dowolnej skali zależność:

$$L_x(j,k) = sup_{\lambda' \in 3\lambda} |d_{\lambda'}| \, , \tag{4}$$

gdzie j, k są liczbami całkowitymi oraz $3\lambda := 3\lambda_{j,k} = \lambda_{j,k-1} \cup \lambda_{j,k} \cup \lambda_{j,k+1}$ i $\lambda := \lambda_{j,k} = \left[k2^j, (k+1)2^j\right]$.

Można wykazać [10], że wykładniki Holdera są wykładnikami skalowania liderów falkowych: $L_x(j,k) \sim 2^{jh}$. Ponadto funkcję strukturalną (5) zdefiniowaną dla liderów falkowych opisuje zależność potęgowa, której wykładnikiem jest multifraktalny wykładnik skalowania $\zeta(q): R \to R$.

$$Z_L(q,j) = \frac{1}{n_j} \sum_{k=1}^{n_j} L_x(j,k)^q = \mathbb{E} L_x(j,k)^q \sim 2^{j\zeta(q)} \quad , \tag{5}$$

gdzie q jest rzędem funkcji strukturalnej, a n_j liczbą przedziałów analizy wielorozdzielczej.

Funkcja uzyskana poprzez transformację Legendre multifraktalnego wykładnika skalowania $\zeta(q)$, przy łagodnych warunkach regularności sygnału, stanowi górną granicę dla widma multifraktalnego (6) badanego sygnału:

$$D(h) \leq min_{q \neq 0}[1 + qh - \zeta(q)] \qquad (6)$$

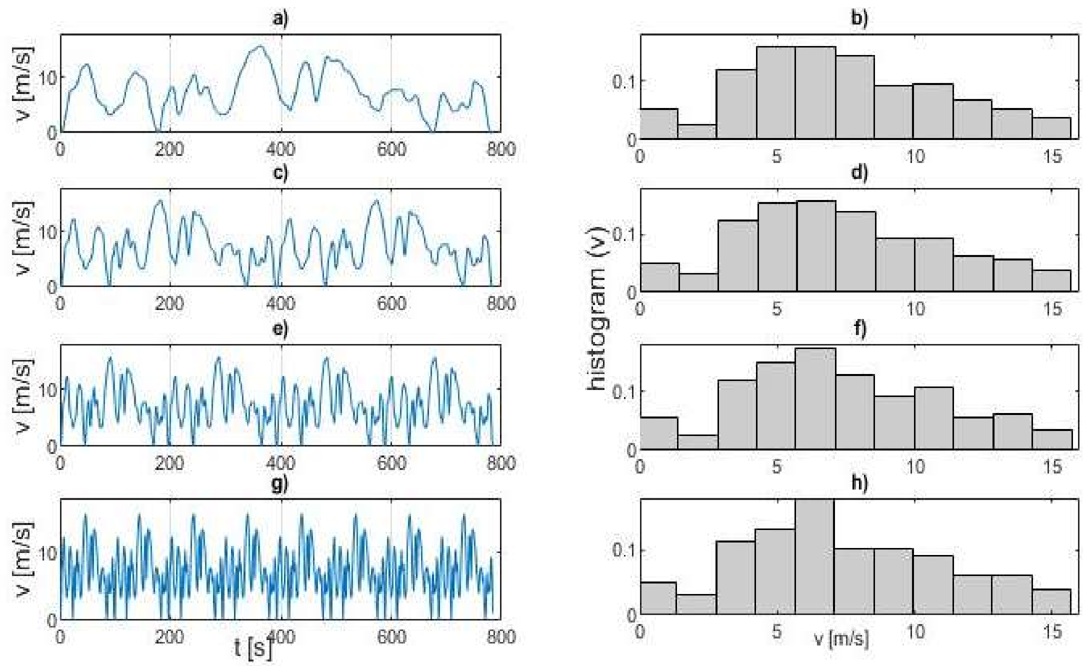
Współczynniki rozwinięcia wykładnika $\zeta(q)$ w szereg Tylora ~~(7)~~ - log kumulanty c_p- stanowią alternatywny opis parametrów widma multifraktalnego analizowanego sygnału.

$$\zeta(q) = \lim_{j \to 0} \frac{\log_2 Z_L(q,j)}{j} = \sum_{p=1}^{\infty} c_p \frac{q^p}{p!} = c_1 q + c_2 \frac{q^2}{2} + c_3 \frac{q^3}{6} + \cdots \qquad (7)$$

W szczególności: współczynnik c_1 opisuje położenie maksimum widma, natomiast współczynniki c_2 oraz c_3 odpowiednio poziom multifraktalności, czyli szerokość spectrum oraz jego asymetrię. Opis własności dynamicznych układów jest z powodzeniem realizowany na podstawie parametrów widm multifraktalnych reprezentatywnych szeregów czasowych [12]. Aproksymacja $\zeta(q)$ (7) , a zatem i widma multifraktalnego D(h) przy wykorzystaniu współczynników c_p znacznie upraszcza algorytmy analizy porównawczej badanych układów.

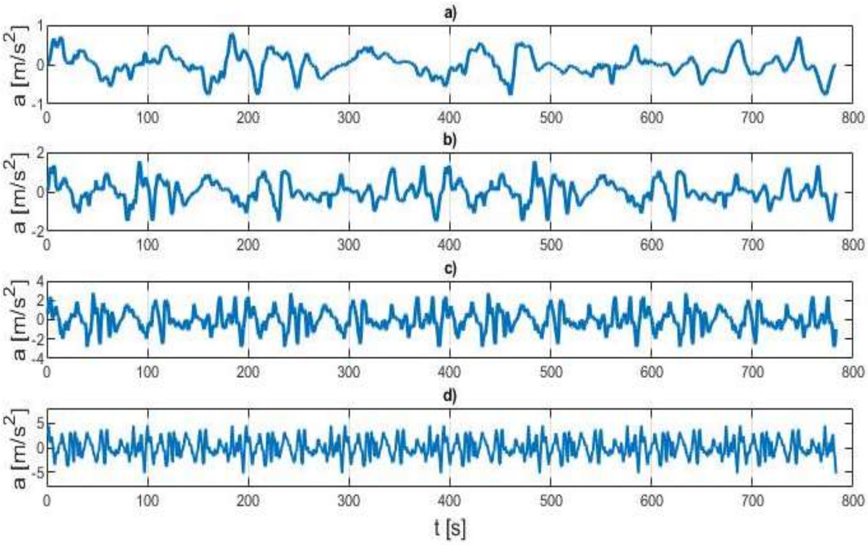
3. Badania symulacyjne zależności dynamiki jazdy i liderów falkowych sygnału prędkości

Związek pomiędzy szeregiem czasowym przyspieszenia i parametrami widma multifraktalnego prędkości zilustrowano na przykładzie syntetycznego sygnału $v(n)$ prędkości pojazdu (Rys.1a). Sygnał ten poddano przepróbkowaniu tak, aby uzyskać sygnały o przyspieszeniu 2x , 4x i 8x większym. Ze względu na operację przepróbkowania histogramy nie są identyczne, lecz porównywalne. Sygnały, które uległy skróceniu powtórzono odpowiednio 2x, 4x i 8x, aby uzyskać sygnały o jednakowej liczbie próbek.

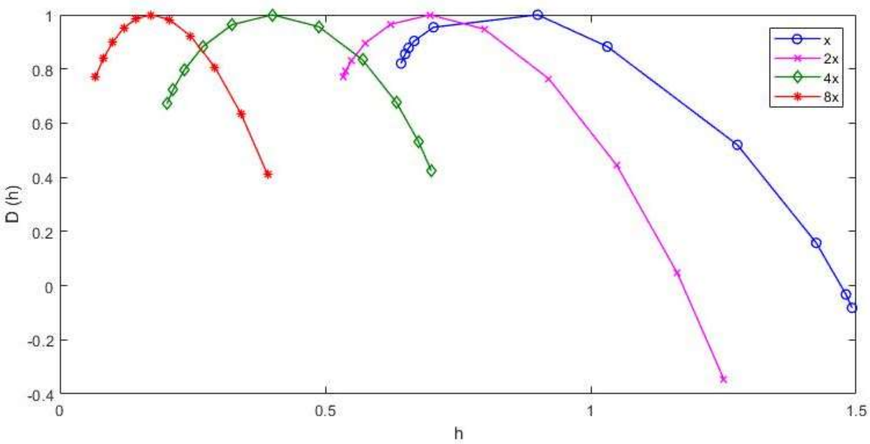


Rys. 1. Syntetyczny sygnał prędkości pojazdu (a) i jego histogramy (b) w zadanym czasie; ten sam przejazd w 2 razy krótszym czasie (c) z porównywalnym rozkładem amplitud (d); ten sam przejazd w 4 razy krótszym czasie (e) z porównywalnym rozkładem amplitud (f); ten sam przejazd w 8 razy krótszym czasie (g) z porównywalnym rozkładem amplitud (h);

Dla sygnałów prędkości z rysunku 1 wyznaczono przyspieszenia (stosując operację różnicowania - Rys. 2) oraz widma multifraktalne (Rys. 3).



Rys. 2. Przebiegi przyspieszenia sygnałów z Rys 1 a), c), e), g)

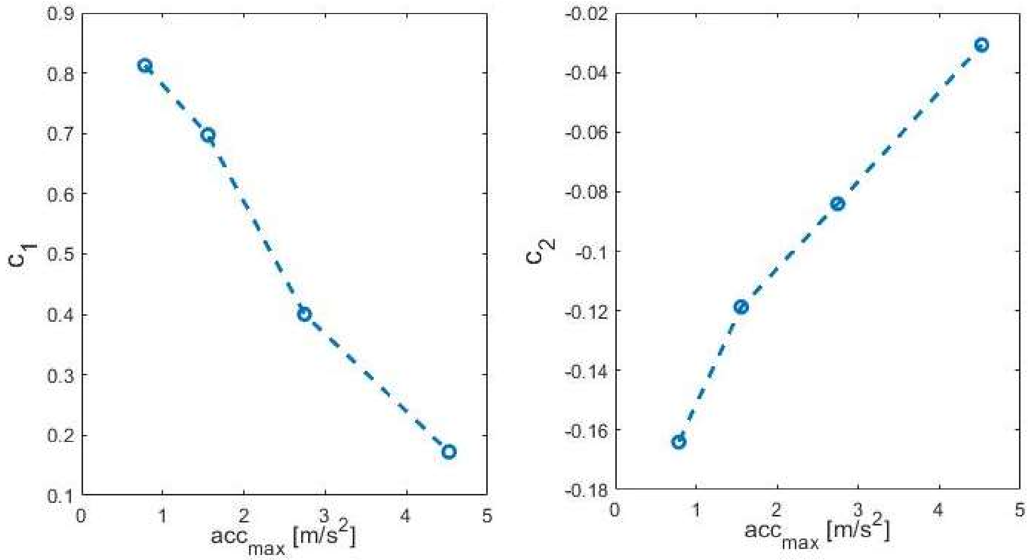


Rys. 3. Widma multifraktalne sygnałów przyspieszenia z Rys. 2 a), b), c), d)

Analiza uzyskanych widm osobliwości wskazuje na zależność położenia ich maksimów i szerokości od przyspieszeń symulowanych sygnałów. Wyznaczono wartości log kumulant syntetycznych sygnałów (Tabela 1) oraz zależności log kumulant i przyspieszenia (Rys. 4). Jako parametry syntetyczne oceny dynamiki jazdy za pomocą widma multifraktalnego zaproponowano pierwszą i drugą log kumulantę, charakteryzujące odpowiednio położenie maksimum widma multifraktalnego oraz jego szerokość.

Log kumulanta	Sygnał			
	x	2x	4x	8x
c_1	0,8127	0,6976	0,4003	0,1723
c_2	-0,1640	-0,1187	-0,0841	-0,0307
c_3	0.1220	0.0282	0.0127	0.0055

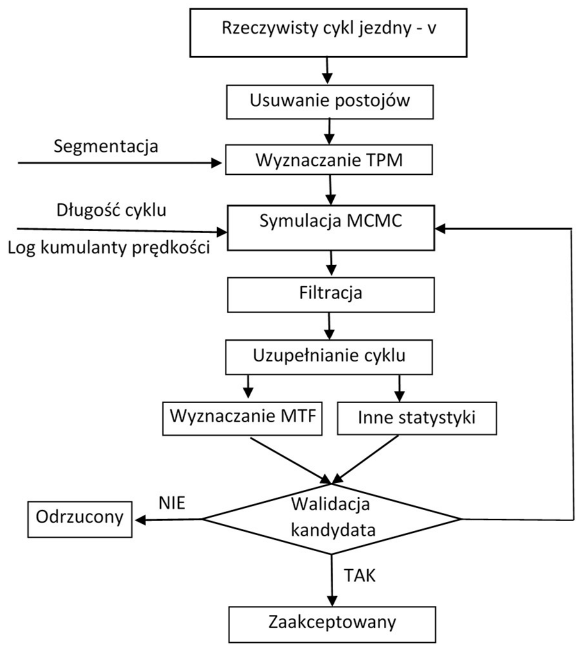
Tabela 1. Wartości log kumulant syntetycznych sygnałów



Rys. 4. Związek pomiędzy log kumulantami i wartością maksymalną przyspieszenia
a) pierwsza log kumulanta
b) druga log kumulanta

4. Realizacja algorytmu syntezy cykli jezdnych i analiza wyników badań (Implementation of the synthesis algorithm of driving cycles of and analysis of research results)

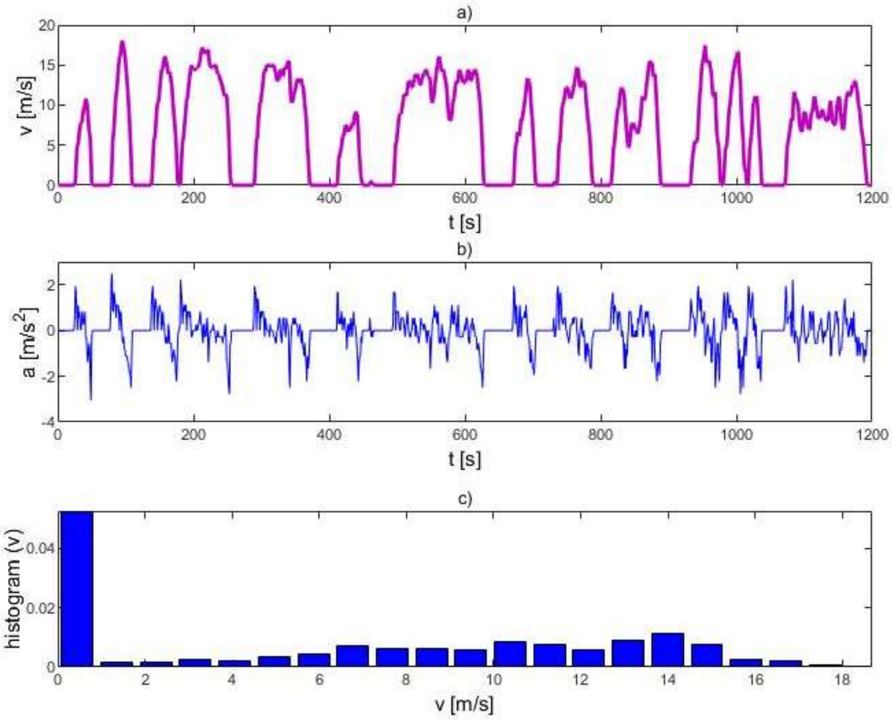
Zaproponowano algorytm generowania eksploatacyjnych cykli jezdnych z wykorzystaniem łańcuchów Markowa pierwszego rzędu oraz formalizmu multifraktalnego bazującego na analizie liderów falkowych (Rys. 5).



Rys. 5. Schemat blokowy algorytmu MCMC syntezy cykli jezdnych

W artykule przedstawiono wyniki badań i analizy ruchu samochodów osobowych w rzeczywistych warunkach drogowych, reprezentowanych przez jazdę miejską w dużej aglomeracji (Rys.6). Analizę przeprowadzono na bazie szeregów czasowych prędkości pojazdu rejestrowanej z okresem próbkowania równym 1s. Opis badań został przedstawiony w pracy [14]. Ze względu na duży, ok. 25% udział w teście "zerowych prędkości" (bieg jałowy) z zarejestrowanych szeregów czasowych usunięto fragmenty odpowiadające postojom (Rys.7), co umożliwiło po segmentacji i wyznaczeniu macierzy prawdopodobieństw przejść TPM, testowanie dynamiki jazdy za pomocą analizy log kumulant. Zarejestrowane prędkości podzielono na 20 równych przedziałów odpowiadających rosnącym prędkościom różnym od zera oraz zakres 21-szy odpowiadający postojom. Rozdzielczość prędkości wynosi ok. 0,9 m/s. Starano się uzyskać dobrą rozdzielczość prędkości samochodu jednocześnie unikając przedziałów o bardzo niskim (lub zerowym) prawdopodobieństwie wystąpienia.

Analizy statystyczne przeprowadzono w środowisku R, natomiast analizę multifraktalną wykonano w programie Matlab.

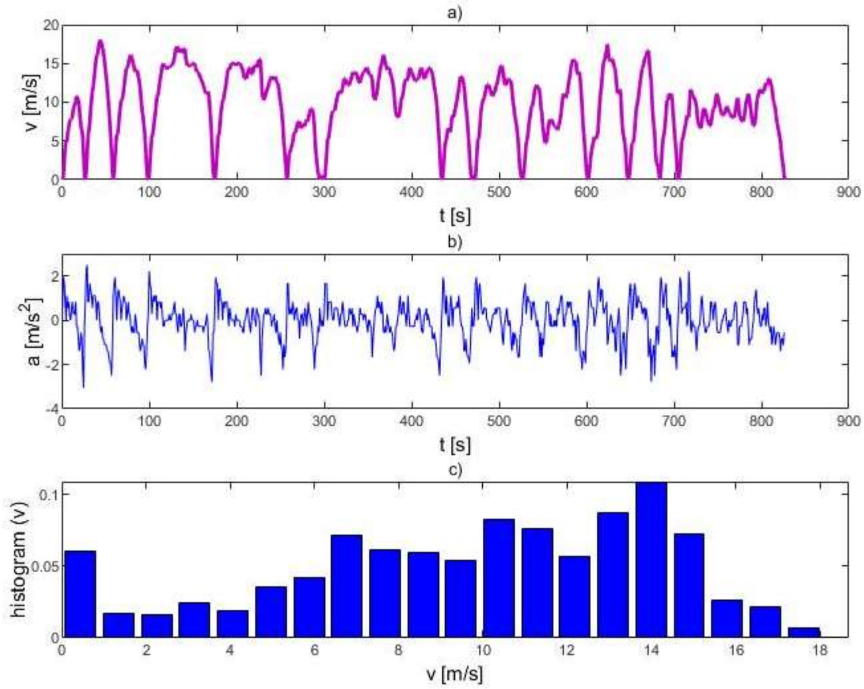


Rys.6. Cykl rzeczywisty reprezentatywny dla prowadzonych badań - fragment 20-minutowy

a) przebieg sygnału prędkości

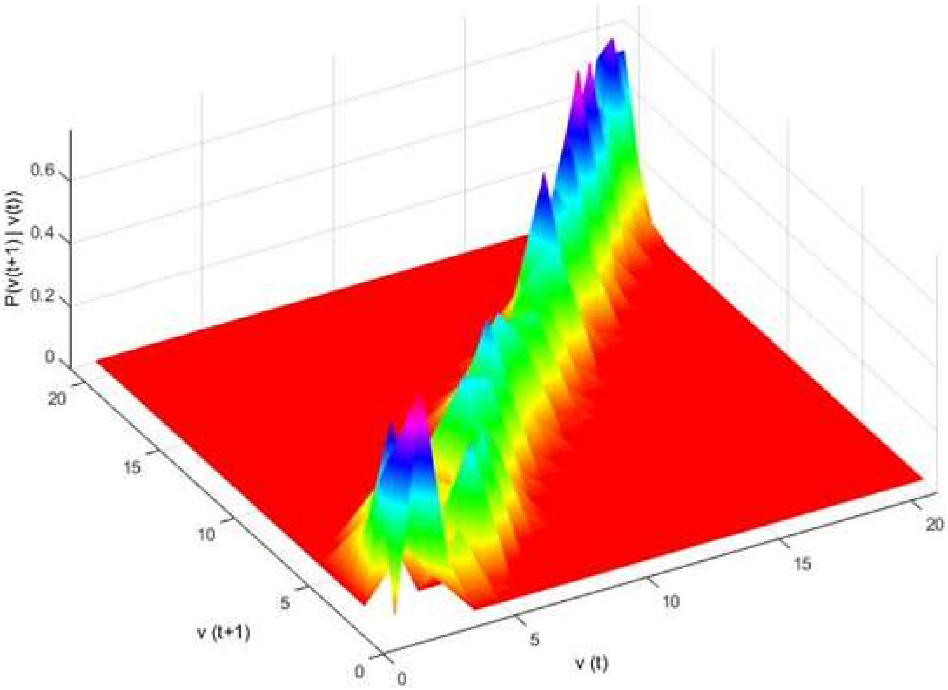
b) przebieg przyspieszenia obliczonego na podstawie prędkości

c) histogram prędkości samochodu



Rys.7. Typowy cykl rzeczywisty po usunięciu postojów
a) przebieg sygnału prędkości (wzorcowy)
b) przebieg przyspieszenia obliczonego na podstawie prędkości
c) histogram prędkości samochodu

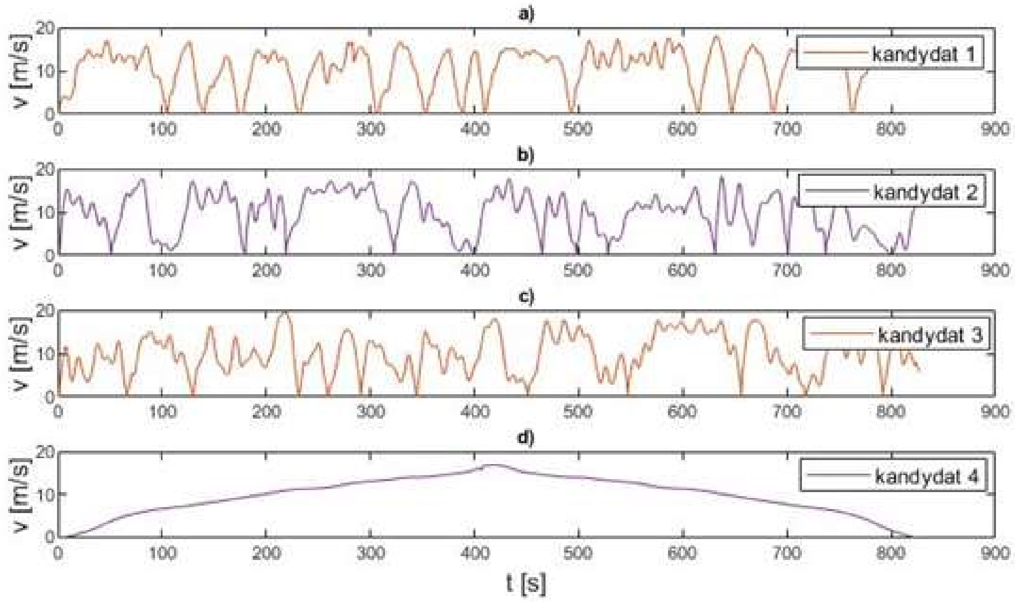
Macierz prawdopodobieństw przejść TPM, obliczona na podstawie sygnału wzorcowego cyklu, uzyskała wymiar 21x21 (Rys.8).



Rys.8. Macierz prawdopodobieństw przejść TPM dla cyklu przedstawionego na Rys. 7a

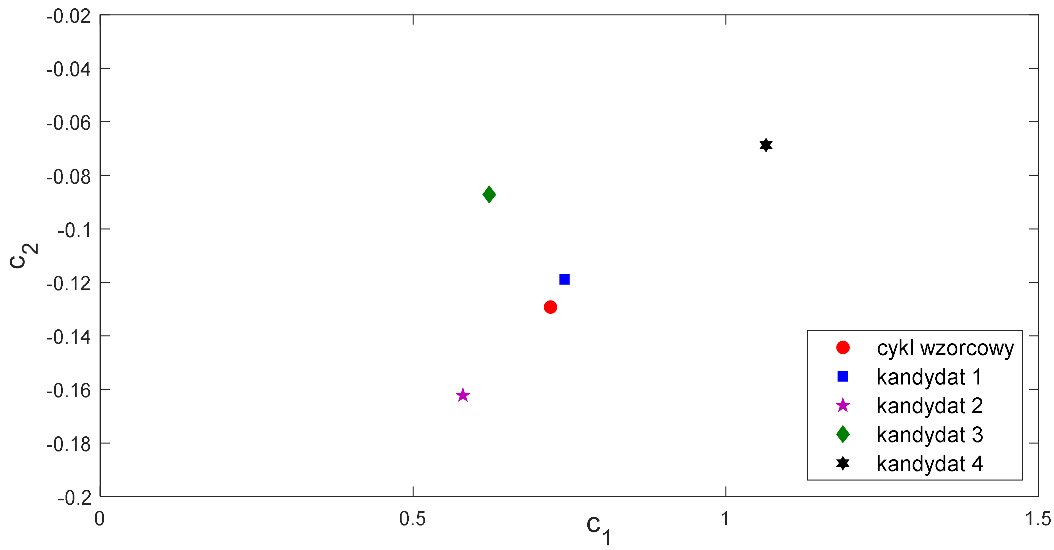
Zgodnie z algorytmem Metropolisa-Hastingsa wykonano symulację 100 cykli. Do zilustrowania wyników algorytmu wybrano trzy przykładowe cykle – kandydatów nr 1, 2, 3-

(Rys.9a-c). Podstawowe statystyki sygnału prędkości (wartość maksymalna, minimalna, średnia, odchylenie standardowe) przykładowych cykli są zbliżone do statystyk przebiegu wzorcowego. Przedstawiono także czwarty cykl – kandydat nr 4 - wygenerowany kontrolnie na podstawie rozkładu amplitud prędkości (Rys.9d).



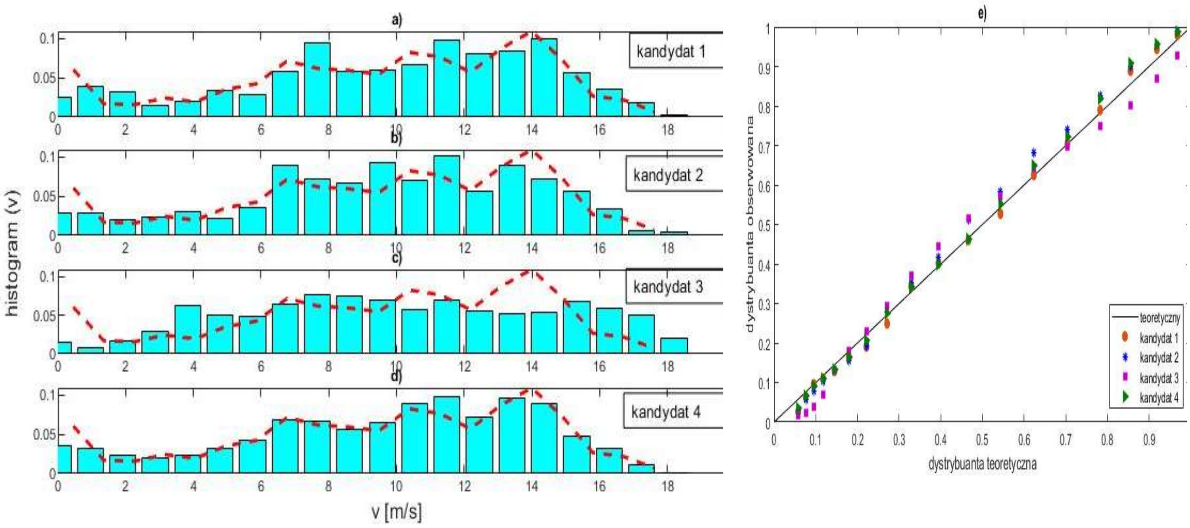
Rys.9. Przykładowe symulowane cykle jezdne

Wyznaczone dwie pierwsze log kumulanty każdego z cykli (Rys.10) wykazują najlepsze dopasowanie dynamiki względem sygnału wzorcowego dla kandydata nr 1.



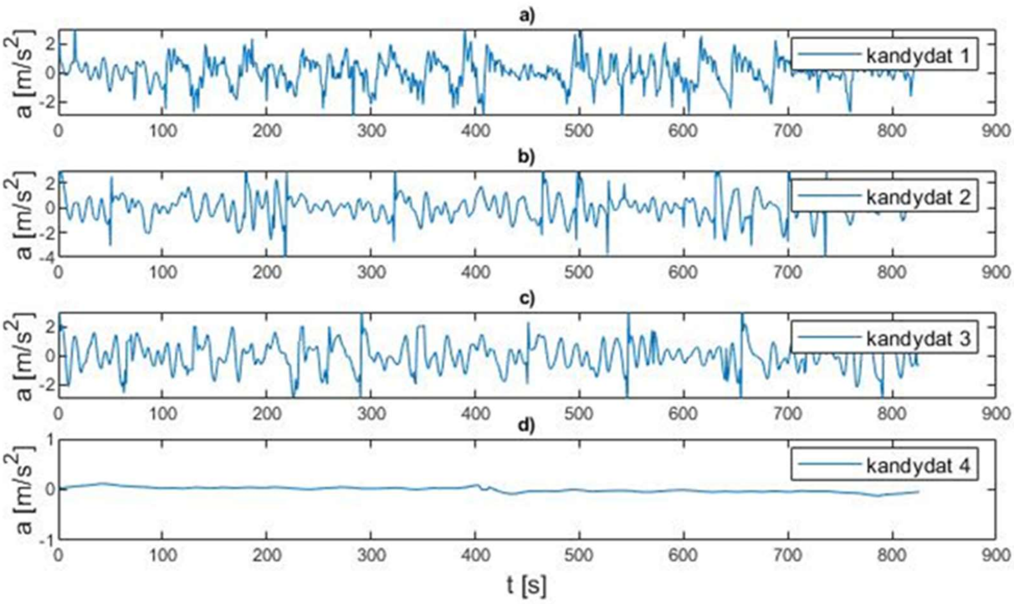
Rys.10. Wykres punktowy log kumulant wyznaczonych dla badanych cykli

Sprawdzono także zgodność rozkładów gęstości prawdopodobieństwa z rozkładem cyklu wzorcowego (Rys. 11a-d). Rozkład cyklu wzorcowego został przybliżony funkcją empiryczną. W celu sprawdzenia dobroci dopasowania danych empirycznych do przybliżonej funkcji rozkładu prawdopodobieństwa można przeprowadzić test Chi-kwadrat lub test Kołmogorova-Smirnova, jednak testy te dla badanej długości cyklu jezdnego odrzucają hipotezę zerową. Aby test potwierdził hipotezę zerową należy znacznie zwiększyć długość cyklu jezdnego, a to nie jest możliwe. W takim przypadku najlepszym sposobem oszacowania jakości dopasowania rozkładu teoretycznego do zaobserwowanego jest porównanie wizualne. Wykorzystano w tym celu wykres prawdopodobieństwo-prawdopodobieństwo (Rys.11e). Poza cyklem kandydata nr 4 najlepsze dopasowanie wykazuje kandydat nr 1.



Rys.11. Histogramy sygnału prędkości dla symulowanych cykli na tle rozkładu wzorcowego (czerwona przerywana linia) a) – d) e) wykres prawdopodobieństwo-prawdopodobieństwo

Przebieg przyspieszenia kandydata nr 4 (Rys.12), który ma rozkład amplitud prędkości idealnie dopasowany do rozkładu wzorcowego, jest zupełnie inny – niemal stały. Przyspieszenia pozostałych kandydatów można jedynie ocenić ze względu na wartości minimalne i maksymalne.



Rys.12. Przebiegi przyspieszeń w cyklach symulowanych obliczone na podstawie prędkości

W przyjętej w pracy metodzie cykli modalnych i segmentacji wg prędkości, przebiegi czasowe uzyskane na podstawie modelu Markowa mają charakter schodkowy, wymagały więc w kolejnym kroku wygładzenia. Spośród różnych metod wygładzania szeregu wybrano metodę lokalnej regresji kwadratowej (ang. Local quadratic regression smoothing). Po zakończeniu procesu iteracji i filtracji szeregi czasowe uzupełniane są okresami postojów i uruchamiany jest proces poszukiwania najbardziej reprezentatywnych cykli spośród wszystkich cykli uzyskanych w wyniku syntezy, dla wybranego wskaźnika równoważności.

W trakcie zrealizowanych badań przeprowadzono analizę wyników algorytmu syntezy równoważnych cykli jezdnych wg wybranych parametrów statystycznych oraz kryterium średniej siły trakcyjnej MTF (8), tj. energii trakcyjnej pojazdu (Tabela 2) przenoszonej przez koła:

$$\overline{F}_{trac} = \frac{1}{x_{total}} \int_{t \in \tau_{trac}} F(t)v(t)dt \quad . \tag{8}$$

gdzie: całkowita siła trakcyjna $F(t)$ jest sumą sił oporu aerodynamicznego F_{air} , oporu toczenia F_{roll} i oporu bezwładności pojazdu F_{iner} , $v(t)$ i $a(t)$ są odpowiednio prędkością i przyspieszeniem jazdy w trakcie cyklu o długości x_{total} oraz τ_{trac} oznacza przedziały czasowe, w których $F(t) > 0$.

F_{air}[N]	Opór aerodynamiczny		$0{,}4v^2(t)$
F_{roll}[N]	Opór toczenia		383
F_{iner}[N]	Opór bezwładności		$1300a(t)$

Tabela 2. Parametry pojazdu

Przy obliczaniu wskaźnika MTF, najbardziej znaczącą jest siła oporu bezwładności, proporcjonalna do przyspieszenia. Najlepsze dopasowanie wg kryterium MTF do cyklu rzeczywistego występuje dla kandydata nr 1. Podstawowe parametry brane pod uwagę przy weryfikacji cykli zestawiono w Tabeli 3. Pominięto wartości minimalne i maksymalne prędkości i przyspieszenia, które są weryfikowane już na początku.

Cykl jezdny	Dopasowanie rozkładu	Wartość średnia	Odchylenie standardowe	Log Kumulanta 1	Log Kumulanta 2	MTF
	[-]	[m/s]	[m/s]	[-]	[-]	[N]
Cykl odniesienia		9,7	4,4	0,72	-0,13	689
Kandydat 1	+	10,0	4,6	0,74	-0,12	743
Kandydat 2	+/-	9,7	4,4	0,58	-0,16	820
Kandydat 3	+/-	10,4	4,5	0,62	-0,10	796
Kandydat 4	+	9,7	4,6	1,06	-0,07	450

Tabela 3. Zestawienie wybranych wartości charakterystycznych dla badanych cykli

Wszystkie syntetyzowane cykle jezdne wykazują prawidłową wartość średnią i odchylenie standardowe. Selekcji nie można także wykonać na bazie rozkładów gęstości prawdopodobieństwa. Kandydaci nr 1 i 4 charakteryzują się najlepszym dopasowaniem rozkładu prawdopodobieństwa amplitud prędkości.

Jeżeli przyjmiemy ponad 10% rozbieżność wskaźnika MTF wygenerowanego cyklu względem cyklu odniesienia za nieakceptowalny poziom równoważności rzeczywistym warunkom jazdy, to log kumulanty testowane w fazie syntezy cykli kandydatów oraz wskaźnik

MTF wykorzystany do weryfikacji ich równoważności, wskazują na kandydata nr 1.

5. Podsumowanie

Przedstawione wyniki badań wnoszą nowe spojrzenie na statystyczno- losowe metody syntezy rzeczywistych cykli jezdnych pojazdów. Wskazano na możliwość odwzorowywania dynamiki jazdy, reprezentowanej przez przyspieszenie, za pomocą parametrów multifraktalnych sygnału prędkości. Zastosowanie liderów falkowych do badania dynamiki jazdy umożliwiło przeprowadzenie syntezy cykli, uwzględniającej parametry prędkości i przyspieszenia, za pomocą symulacji Monte Carlo z jednowymiarowym łańcuchem Markowa. Algorytm syntezy równoważnych cykli jezdnych zweryfikowano wg kryterium średniej siły trakcyjnej MTF.

Wykorzystana dotąd baza danych pochodziła z badań pojazdów z silnikami spalinowymi. Następne badania autorów obejmą analizę procesu prognozowania cyklu jazdy i modelowania ruchu drogowego dla potrzeb sterowania układem napędowym i zarządzaniem energią elektryczną pojazdów elektrycznych. Oczekiwane rezultaty będą przydatne do projektowania infrastruktury stacji ładowania samochodów napędzanych elektrycznie.

Literatura

1. Brady J, O'Mahony M. Development of a driving cycle to evaluate the energy economy of electric vehicles in urban areas. Applied Energy 2016; 177: 165–178.
2. Chłopek Z. The synthesis of driving tests according to the criterions of similarity of frequency characteristics. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2016; 18(4): 572-577.
3. Dai M, Zhang Ch, Zhang D. Multifractal and singularity analysis of highway volume data. Physica A 2014; 407: 332-340.
4. Esser A, Zeller M, Foulard S, Rinderknecht S. Synthesis of Representative and Multidimensional Driving Cycles. SAE International 2018; 2018-01-0095.
5. Hereijgers K, Silvas E, Hofman T, Steinbuch M. Effect of using Synthetized Driving Cycles on Vehicle Fuel Consumption. IFAC papers 2017; 50-1: 7505-7510
6. Lee T-K, Filipi Z.S. Synthesis of real-world driving cycles using stochastic process and statistical methodology. International Journal of Vehicle Design 2011; 57(1): 17-36.
7. Liu Z, Ivanco A, Filipi Z. Naturalistic drive cycle synthesis for pickup trucks. Journal of Safety Research 2015; 54: 109–115.
8. Man Z, Shuming S, Nan L, Bingjian . High-efficiency Driving Cycle Generation Using a Markov Chain Evolution Algorithm. IEEE Transactions on Vehicular Technology December 2018; DOI: 10.1109/TVT.2018.2887063.
9. Mayakuntlaa SK, Vermab A. A novel methodology for construction of driving cycles for Indian cities. Transportation Research Part D 2018; 65: 725–735.
10. Mukli P, Nagy Z, Eke A. Multifractal formalism by enforcing the universal behavior of scaling functions, Physica A 2015; 417: 150–167.
11. Nyberg P, Frisk E, Nielsen L. Generation of Equivalent Driving Cycles Using Markov Chains and Mean Tractive Force Components. Proc. of IFAC, 2014; 8787–8792.
12. Puchalski A, Komorska I. Binomial multifractal features of worldwide harmonized light duty vehicles test cycle. JVE Vibroengineering Procedia 2017; 13:175–179.
13. Puchalski A, Komorska I. Multifractal Nature of Diesel Engine Rattle Noise in Vehicle. Archives of Acoustics 2017; 42(3): 469-474.

14. Puchalski A, Ślęzak M, Komorska I, Wiśniowski P. Multifractal analysis vehicle's in-use speed profile for application in driving cycles. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2018; 20 (2): 177–181.
15. Puchalski A., Komorska I. A generalised entropy in multifractal time signals analysis of mechanical vibration. JVE Journal of Vibroegineering 2018; 20 (4):1667 – 1675.
16. Shang P, Lu Y, Kamae S. Detecting long-range correlation of traffic time series with multifractal detrended fluctuation analysis. Chaos, Solitions and Fractals 2008; 36: 82-90.
17. Silvas E, Hereijgers K, Peng H, Hofman T, Steinbuch M. Synthesis of Realistic Driving Cycles with High Accuracy and Computational Speed, Including Slope Information. IEEE Transaction Vehicular Technology 2016; 65(6): 4118–4128.
18. Tutuianu M, Marotta A, et al. Development of a World-wide Worldwide harmonized Light duty driving Test Cycle (WLTC) and a possible pathway for its introduction in the European legislation. Transportation Research Part D Transport & Environment 2015; 40(4), 61-75.
19. Yi Y, Pengjian S. Multiscale multifractal detrended cross-correlation analysis of traffic flow. Nonlinear Dynamics 2015; 81:1329–1347.
20. Zhao B, Lv Ch, Hofman T. Driving-Cycle-Aware Energy Management of Hybrid Electric Vehicles Using a Three-Dimensional Markov Chain Model. Automotive Innovation 2019; 2:146–156.

Dr hab. Inż. Maciej Tabaszewski[1]

Poznan University of Technology
Institute of Applied Mechanics
ul. Jana Pawła II 24, 60-965 Poznań, Poland
e-mail: maciej.tabaszewski@put.poznan.pl

Dr hab. inż. Grzegorz. M. Szymański

Poznan University of Technology
Institute of Transport
ul. Piotrowo 3, 60-965 Poznań, Poland
e-mail: grzegorz.m.szymanski@put.poznan.pl

Diagnostyka luzu zaworów silnika spalinowego z wykorzystaniem sygnału drganiowego i metod uczenia maszynowego

Słowa kluczowe: *silnik spalinowy, diagnostyka, drgania, uczenie maszynowe*

Streszczenie: Dynamiczny rozwój konstrukcji silników spalinowych generuje potrzebę wprowadzenia strategii eksploatacji jednostek napędowych, opartej na znajomości ich stanu technicznego. W artykule poddano analizie zagadnienia, związane z drganiową diagnostyką luzu zaworów tłokowego silnika spalinowego, istotnego ze względu na efektywność pracy silnika i jego trwałość. Zaproponowano wykorzystanie metod klasyfikacji do oceny poprawności luzu zaworowego. Przeprowadzono i opisano eksperymenty, które miały na celu dostarczenie informacji koniecznych do zbudowania i zweryfikowania zaproponowanych metod. W przeprowadzonych badaniach pozyskano sygnały drganiowe z trójosiowego czujnika przyspieszeń drgań zlokalizowanego na głowicy silnika. Dokonano parametryzacji uzyskanych przebiegów czasowych sygnału drganiowego dla silnika pracującego pod różnym obciążeniem, z różnymi prędkościami obrotowymi oraz z różnymi luzami zaworowymi. Parametryzacja dotyczyła zarówno cech sygnału przyspieszeń drgań, pochodnej przyspieszeń drgań względem czasu jak i obwiedni tej pochodnej. W pierwszym podejściu zbudowano klasyfikator w postaci zbioru drzew binarnych, który przy okazji pozwolił na wyodrębnienie istotnych, ze względu na przyjęte klasy, cech. Dla porównania zbudowano także klasyfikatory w postaci sieci neuronowej jak i algorytmu k – najbliższych sąsiadów z metryką euklidesową. Na podstawie przeprowadzonych badań i analiz zaproponowano metodę oceny luzu zaworowego.

1. Wprowadzenie

Tłokowy silnik spalinowy jest powszechnie stosowany do napędu pojazdów oraz urządzeń stacjonarnych. Zamienia on energię zawartą w paliwie na pracę mechaniczną obracającego się wału korbowego i tak, jak wszystkie urządzenia mechaniczne podlega zużywaniu i starzeniu. Trwałość silnika określona jest właściwościami konstrukcyjnymi i w dużym stopniu zależy od warunków eksploatacji oraz charakteru obciążeń. Wraz z postępującym zaawansowaniem procesów degradacji konstrukcji (oddziaływanie zmiennych temperatur, procesy trybologiczne, kawitacja, korozja chemiczna i elektrochemiczna, starzenie itp.) pogarszają się parametry niezawodności i sprawności. W konsekwencji dochodzi do zużycia i uszkodzenia obiektu lub

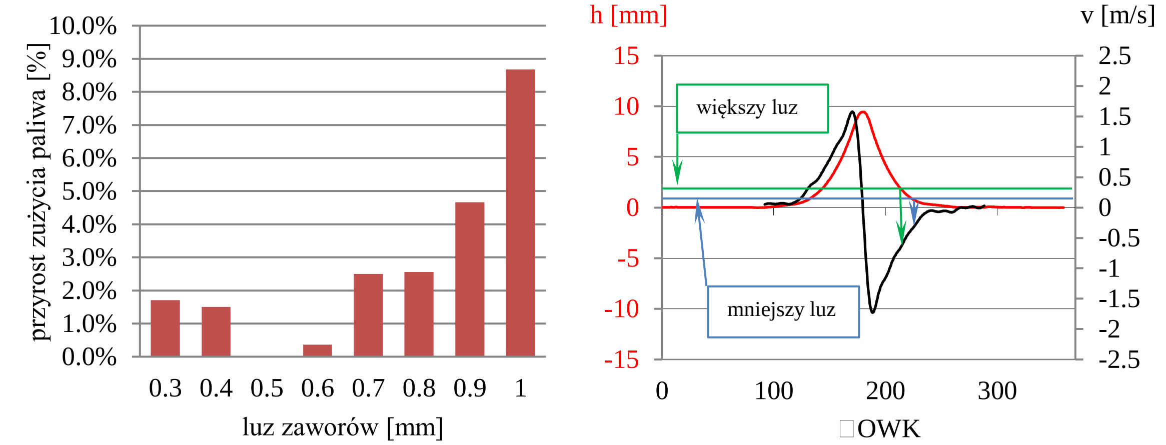
[1] Corresponding author . tel.: +48 61 665 2390; fax: 48 61 665 2307
Email address: maciej.tabaszewski@put.poznan.pl

jego wyłączenia z eksploatacji ze względów ekonomicznych lub innych (np. ekologicznych). Już od początków historii silników spalinowych zauważono, że jednym z kluczowych problemów mających wpływ na jego prawidłową pracę jest poprawne ustawienie parametrów regulacyjnych.

Degradacja struktury silnika oraz nieprawidłowo ustawione parametry regulacyjne mogą być przyczyną następujących zjawisk w silniku spalinowym: pogorszenia efektywności pracy silnika spalinowego, zmniejszenia mocy, związanego ze zmniejszeniem sprawności mechanicznej, sprawności cieplnej oraz współczynnika napełnienia, zwiększenie emisji związków toksycznych w spalinach, możliwość uszkodzenia elementów silnika spalinowego.

Na rysunku 1a. przedstawiono różnice w zużyciu paliwa przy zmianie luzu zaworów rozrządu silnika spalinowego, natomiast zmiany prędkości zderzenia przy uwzględnieniu wzniosu krzywki h) zaworu z gniazdem zaworowym na rys. 1.b. Z danych przedstawionych na rys. 1a wynika, że zmiany luzu zaworowego mogą powodować wzrost zużycia paliwa przez badany silnik spalinowy o ok. 9%, natomiast analiza rys. 1b. pozwala stwierdzić, ze wraz ze wzrostem luzu zaworowego (linie: niebieska i zielona rys.1.b.) zwiększa się prędkość uderzenia zaworu o gniazdo, powodując dodatkowe niepożądane obciążenia dynamiczne elementów głowicy cylindrów silnika spalinowego.

a) b)



Rys. 1. a) Przyrost zużycia paliwa w funkcji luzu zaworów, b) zarys krzywki (linia czerwona), prędkość punktu na krzywce (linia czarna) oraz zmiany prędkości zderzenia zaworu z gniazdem w silniku spalinowym spowodowane zmianami luzów zaworów (pionowe strzałki: zielona i niebieska), OWK – kąt obrotu wału korbowego, *h* – wznios krzywki, v – prędkość puntu na krzywce

Opracowano wiele metod diagnozowania stanu technicznego silników spalinowych. Można je podzielić na metody wykorzystujące procesy robocze (indykowanie, zmiany momentu obrotowego w funkcji obrotu wału korbowego, pomiar ciśnienia i temperatury spalin, ciśnienia w przestrzeni nad i pod tłokiem, parametrów zasilania, zadymienia spalin, itp.) oraz procesy resztkowe (drgania, hałas, procesy termiczne, elektryczne i inne). Na podstawie badań procesów roboczych można wnioskować o ogólnym stanie silnika spalinowego, natomiast procesy resztkowe niosą informacje o stanie poszczególnych podzespołów i par kinematycznych. Dlatego procesy resztkowe wykorzystuje się, jako autonomiczne lub wspomagające metody diagnostyczne. Wszystkie metody oparte na analizie drgań i hałasu dla określenia stanu technicznego obiektu noszą nazwę diagnostyki wibroakustycznej.

Mechanizm rozrządu zaworowego jako jeden z podstawowych mechanizmów wchodzących w skład tłokowych silników spalinowych już z samej zasady swego działania jest źródłem sygnału wibroakustycznego, ponieważ pracy rozrządu zaworowego obejmującego: otwieranie i zamykanie zaworów, współpracę krzywki z popychaczem, kasowanie luzu w łożyskach dźwigni zaworowych i innym towarzyszą zderzenia współpracujących elementów, a te z kolei wywołują drgania. Wykorzystanie sygnałów drgań do diagnozowania układu rozrządu silnika spalinowegoprzedstawiono w następujących pracach:

W artykule [12] omówiono opcję zastosowania rozkładu pakietu falek podczas filtrowania sygnału akustycznego silnika spalinowego w celu zdiagnozowania nadmiernego luzu zaworowego. Autorzy przygotowali algorytm umożliwiający wybór wybranych szczegółów i zbliżenie analizy falkowej do komponentów o niskiej częstotliwości, które stanowią hałas, a także do komponentów o wysokiej częstotliwości, zawierających informacje o możliwym powiększeniu luzu zaworowego silnika. Następnie, w oparciu o wybrane komponenty sygnału akustycznego wysokiej częstotliwości, opracowano metodę automatycznego wykrywania powiększonych luzów zaworów, zakładając, że udział energetyczny emitowanego sygnału akustycznego powinien być ustalany podczas otwierania i zamykania poszczególnych zaworów. Autorzy artykułu [13] opisali badania dotyczące diagnostyki stanu zaworów wydechowych w dużych morskich silnikach wysokoprężnych. Badania wykonano na czterocylindrowym 2-suwowym morskim silniku wysokoprężnym o średnicy tłoka 500 mm w Centrum Badawczym MAN B&W Diesel w Kopenhadze w Danii. Eksperymenty obejmowały trzy różne stany zaworu, przy czym dwa dotyczyły sztucznie wywołanych sytuacji przepalenia zaworu. Podstawowymi pomiarami monitorowania były drgania i fale naprężeń strukturalnych, znane również jako emisja akustyczna (AE). Wyniki wykazały, że sygnały AE mają znaczną przewagę nad innymi zaangażowanymi czujnikami, co wskazuje na wrażliwość zarówno na działanie mechaniczne, jak i mechaniczno-płynne spalanie. Zarejestrowane dane zostały wstępnie przetworzone, a funkcje wyodrębnione przy użyciu analizy głównych składników (PCA). Na podstawie szeregu zastosowanych heurystyk i statystyk przeprowadzono poszukiwanie optymalnej podprzestrzeni głównych komponentów. Wybrana przestrzeń cech została wykorzystana do klasyfikacji stanu zaworu wydechowego poprzez zastosowanie zarówno znormalizowanych klasyfikatorów neuronowych ze sprzężeniem zwrotnym, jak i dyskryminatorów liniowych. Natomiast w artykule [28] obliczono rozkłady jądra w kształcie stożka (CKD) sygnałów przyspieszenia drgań pozyskanych z głowicy cylindra w ośmiu różnych stanach mechanizmu rozrządu i przedstawiono je na obrazach szarych. Do dekompozycji danych wielowymiarowych wykorzystano nieujemną faktoryzację macierzy (NMF), a zespół sieci neuronowej (NNE), który ma lepszą zdolność uogólnienia do klasyfikacji niż pojedyncza sieć neuronowa. NNE został wykorzystany do przeprowadzenia inteligentnej diagnozy na podstawie rozkładów czasowo-częstotliwościowych. Na podstawie wyników eksperymentalnych wykazano, że uszkodzenia mechanizmu rozrządu w silniku diesla można dokładnie sklasyfikować za pomocą proponowanej metody. W pracy [31] poddano analizie zagadnienia, związane z drganiową diagnostyką automatycznych kompensatorów luzu zaworów tłokowych silników spalinowych. Opisano i przeprowadzono eksperymenty naukowe, które miały na celu dostarczenie informacji koniecznych do zbudowania i zweryfikowania modeli diagnostycznych umożliwiających ocenę stanu technicznego tych elementów silnika spalinowego, istotnych ze względu na efektywność pracy silnika i jego trwałość. Na podstawie wyznaczonego modelu diagnostycznego opracowano algorytm oceny stanu technicznego automatycznych kompensatorów luzu zaworowego. Natomiast w artykule [32] obliczono rozkłady Wignera-Ville'a (WVD) sygnałów przyspieszenia drgań, które zostały zarejestrowane na głowicy cylindrów w ośmiu różnych stanach rozrządu zaworowego i wyświetlone na obrazach szarych. Do klasyfikacji obrazów czasowo-częstotliwościowych po ich znormalizowaniu zostały bezpośrednio wykorzystane probabilistyczne sieci neuronowe (PNN). W ten sposób diagnostyka uszkodzeń mechanizmu rozrządu została przeniesiona do klasyfikacji obrazów czasowo-częstotliwościowych. P Wyniki eksperymentów pokazują, że usterki w zespołach zaworów diesla można dokładnie sklasyfikować za pomocą proponowanych metod.

Zastosowanie różnych technik analizy sygnałów wibroakustycznych do oceny stanu technicznego silników spalinowych przedstawili autorzy prac:

Autorzy pracy [1] wykazali, że sygnały akustyczne pochodzące z silnika spalinowego bogate są w informacje dotyczące parametrów pracy silnika i jego stanu. Niestety, te informacje są złożone i charakteryzują się dużym szumem tła. Na podstawie analizy zmierzonych parametrów wykazali, że można diagnozować niedomagania silnika spowodowane spadkiem stopnia sprężania, zmianami ciśnienia wtrysku, zmianami w układzie wylotowym, zmianą luzu zaworu ssącego i wylotowego. Oceny stanu technicznego silnika przeprowadzili na podstawie analizy zmian wartości skutecznej oraz kurtozy sygnałów akustycznych każdego z cylindrów silnika. Autorzy artykułu [2] stwierdzili, że sygnały akustyczne wywołane przez mechanizmy często muszą być opisywane przez nieliniowe modele w dziedzinie czasu. Natomiast w dziedzinie częstotliwości model liniowy jest w wielu przypadkach wystarczający do opisu kanałów propagacji dźwięku. W pracy dokonali porównania metod obliczeniowych pod względem dokładności, czasu obliczania i możliwości do wykonywania badań w czasie eksploatacji obiektu. W artykule [3] przedstawiono szybką i automatyczną metodę diagnostyki silnika opartą na jednym parametrze emisji akustycznej (AE). Metoda oparta jest na porównaniu drgań i energii AE z wartościami odniesienia w celu ustalenia czy stan silnika jest wadliwy. Metoda została zastosowana w silniku testowym i okazała się zadowalająca. Praca [9] dotyczy monitorowania dużych silników Diesla poprzez analizę zmian prędkości kątowej wału korbowego. Skoncentrowano się na 20-cylindrowym silniku wysokoprężnym z naturalnymi częstotliwościami wału korbowego w zakresie prędkości roboczych. Zmiany prędkości kątowej modelowano na wolnym końcu wału korbowego. Modelowanie obejmowało zarówno zachowania dynamiczne wału korbowego, jak i momenty wzbudzenia. Ponieważ silnik jest bardzo duży, pierwsze tryby skręcania wału korbowego są w zakresie niskich częstotliwości. Wymagany jest model z założeniem elastycznego wału korbowego. Momenty wzbudzenia zależą od krzywej ciśnienia w cylindrze. Ten ostatni modelowano za pomocą modelu fenomenologicznego. Parametry mechaniczne i spalania modelu optymalizowano za pomocą rzeczywistych danych. Zaproponowano zautomatyzowaną diagnozę opartą na systemie wykorzystującym sztuczną inteligencję. Sieci neuronowe wykorzystywano do rozpoznawania wzorców przebiegów prędkości kątowej w normalnych i wadliwych warunkach. Wzory odniesienia wymagane na etapie szkolenia obliczano z modelu, skalibrowanym przy użyciu niewielkiej liczby rzeczywistych pomiarów. Uzyskano obiecujące wyniki. Podczas badań weryfikacyjnych pomyślnie zdiagnozowano uszkodzenie polegające na wycieku paliwa. W pracy [10] wykonano sprzężoną symulację dynamiki tłoka i trybologii silnika (trybodynamiki) przy użyciu quasi-statycznych i przejściowych kodów numerycznych do modelowania uderzeń tłoka o ściankę cylindra. Potwierdzenie przydatności proponowanych metod określono na podstawie pomiarów eksperymentalnych wykonanych na jednocylindrowym silniku benzynowym w warunkach laboratoryjnych Poprzez pomiar przyspieszeń drgań powierzchni bloku silnika. Autorzy artykułu [34] zaproponowano system diagnostyki uszkodzeń silników spalinowych wykorzystujących techniki transformacji pakietów falkowych (WPT) i sztucznej sieci neuronowej (ANN). W artykule [5] omówiono zastosowanie czujnika prądu jonowego do wykrywania rezonansu spalania w silniku wysokoprężnym z bezpośrednim wtryskiem. Zmodyfikowana świeca żarowa służą do pomiaru prądu jonowego oprócz jego głównej funkcji polegającej na podgrzewaniu komory spalania. Dokonano porównania rezonansu spalania określonego na podstawie sygnałów czujnika prądu jonowego, przetwornika ciśnienia w cylindrze i czujnika drgań silnika. Stwierdzono, że sygnał prądu jonowego można wykorzystać do określenia synchronizacji, amplitudy, częstotliwości i czasu trwania rezonansu. Wyjście czujnika może zostać wykorzystane jako sygnał zwrotny do ECU (elektronicznej jednostki sterującej) w celu zminimalizowania drgań i hałasu silnika. Artykuł [8] dotyczy najnowocześniejszych strategii i technik diagnostycznych opartych na sygnałach wibroakustycznych, które mogą służyć do monitorowania i diagnozowania silników spalinowych (ICE) zarówno na stanowisku badawczym, jak i w warunkach pracy. W tym artykule po raz pierwszy przedstawiono krótkie podsumowanie generowania dźwięku i drgań w ICE w kontekście dalszej dyskusji na temat diagnostyki wibroakustycznej. Przedstawiono również przegląd technik monitorowania i diagnostyki opisanych w literaturze, przy użyciu sygnałów drganiowych i akustycznych.

Na podstawie analizy dotychczasowych dokonań w dziedzinie diagnostyki wibroakustycznej układów silników spalinowych stwierdzono, że: prowadzono badania dotyczące zastosowania parametrów sygnałów drgań do oceny stanu technicznego zespołów silnika spalinowego lub zachodzących w nim procesów, badania dotyczyły zagadnień metodycznych (np. ustalenia warunków pracy silnika w trakcie pomiarów drgań, wyboru punktów pomiarowych) oraz zagadnień modelowania (budowania modeli diagnostycznych i ich weryfikacji) w badaniach wykorzystywano zarówno proste metody opisu sygnałów (np. miary punktowe) jak i bardzo zaawansowane techniki przetwarzania sygnałów (np. sztuczne sieci neuronowe, analiza czasowo-widmowa). W analizowanych pracach dotyczących drganiowej diagnostyki silników spalinowych dostrzeżono metody, w których modele diagnostyczne bazowały na odpowiedziach struktury obiektu na wymuszenia impulsowe. Analizy dotyczące odpowiedzi stanowiły w przeprowadzonych badaniach podpowiedź co do optymalnego wyboru punktu pomiarowego jak i ewentualnie zakresu częstotliwości, które powinna objąć analiza.

W niniejszym artykule przedstawiono nowe podejście do identyfikacji luzu zaworowego silnika pracującego z różnym obciążeniem i prędkością obrotową na podstawie odpowiednio przetworzonego sygnału drganiowego. Zaproponowano metodykę identyfikacji klasy luzu zaworowego w oparciu o przyspieszenia drgań bezwzględnych mierzonych na głowicy silnika i systemy uczone pod nadzorem – klasyfikatory. Takie rozwiązanie pozwala oceniać automatycznie prawidłowość luzu zaworowego na pracującym silniku bez poszukiwania matematycznego modelu opisującego związek pomiędzy sygnałem drganiowym a luzem.

Klasyfikatory jako systemy uczone pod nadzorem, mają zastosowanie w bardzo wielu dziedzinach przy przetwarzaniu bardzo dużych zasobów danych i automatyzacji procesu wnioskowania. Nie sposób dokonać pełnego przeglądu zastosowań wspomnianych algorytmów, a nawet dziedzin ich zastosowania. Przykładowo można tylko wspomnieć tak różne obszary jak: przewidywania wyników egzaminów uczniów [6], monitorowanie zmian urbanistycznych [11], klasyfikację chropowatości dróg [15], segmentację defektów jabłek [19], inteligentny system detekcji uszkodzeń maszyn wirnikowych [20], klasyfikację ruchu sieciowego [24], czy przetwanie i analizę obrazu [35]. Znanych jest wiele prób wykorzystania uczenia maszynowego także w diagnostyce maszyn i podzespołów w celu określenia stanu technicznego jak i cech stanu. Przykładowo w pracy [16] rozpatrywane jest zastosowanie sieci konwolucyjnej z pewnymi modyfikacjami do klasyfikacji stanu maszyn. Autorzy zprezentowali obiecujące wyniki działania metody na przykładzie sygnałów z łożysk tocznych. W pracy [18] autorzy z powodzeniem zastosowali klasteryzację metodą k- means oraz klasyfikację metodą SVM (Support Vector Machine) w celu oceny stanu zużycia ostrzy maszyny pakującej. W pracy [23] zastosowano statystyczne miary odległości do rozróżnienia stanów uszkodzenia łożysk tocznych. Autorzy pracy [27] wykazali skuteczność metod klasyfikacji w odniesieniu do monitorowania stanu sprężarek tłokowych zaistalowanych w urządzeniach chłodniczych. Wykorzytsano tu między innymi sieci neronowe ale także maszny do uczenia ekstremalnego (ELM). W pracy dotyczącej diagnostyki turbin wiatrowych [29] stosowano reprezentacje sygnałów czasowych w postaci obrazów, wykorzytywano różne cechy tekstur i dokonywano klasyfikacji za pomocą tych cech. Omówienie wielu metod uczenia maszynowego w kontekście ogólnych zastosowań diagnostycznych można znaleźć także w pozycji [17]. Metody uczenia maszynowego wykorzystywano także w diagnostyce silników spalinowych tłokowych. W pracy [4] prównano różne metody klasyfikacji wykorzytsane do identyfikacji zjawiska wypadania zapłonu. Autorzy pracy [7] dokonywali diagnozy stanu wstryskiwaczy silnika za pomocą sygnału drganiowego, dyskretnej transformaty falkowej i sieci neuronowej. Z kolei w pracy [21] analizowano tymi samymi metodami problem wyłączenia cylindrów silnika. Wiele niesprawności silnika spalinowego identyfikowano także za pomocą metod probablistycznych [33]. Nie rozważano tutaj jednak problemu nieoptymalnych luzów zaworowych. W pracy [14] zastosowano klasyfikacje do identyfikacji wielu uszkodzeń – między innymi zbyt małego lub dużego luzu zaworowego. Badania zrealizowano z zastosowaniem zaawansowanych metod ekstremalnego uczenia maszynowego. Za anlizy pracy wynika, że trafność klasfikacji uszkodzeń uzyskanych przez badaczy nie przekracza 96%. Wydaje się, że pożądane jest znalezienie prostszej metody klasyfikacji luzu ze względu na możliwość jej łatwej implementacji praktycznej w diagnosyce pokładowej, stąd zdaniem autorów tej pracy konieczne są dalsze poszukiwania i badania w tym kierunku. Istotne jest również to, by mimo prostosty metody zachować wysoką pewność diagnozy.

Proces klasyfikacji można technicznie przeprowadzić wieloma metodami o specyficznych właściwościach i możliwościach. Z pośród szergu możliwości można wymienić tutaj drzewa klasyfikacyjne, sieci neuronowe, klasyfikatory odległościowe, aproksymacyjne, rozmyte itp. W pracy zastosowano trzy metody, które zdaniem autorów są najprostsze do praktycznej implementacji w pokładowej diagnostyce silnika: drzewo klasyfikacyjne, sztuczne sieci neuronowe MLP i klasyfikator k – najbliższych sąsiadów.

Zaletą struktury drzewa jest przystępny dla człowieka sposób reprezentacji wiedzy, którą można uzyskać po procesie uczenia. Inną zaletą jest brak wymagań dotyczących założeń na temat związku zmiennej zależnej ze zmiennymi objaśniającymi. Ponadto drzewo pozwala wyłowić automatycznie te miary diagnostyczne, które mają istotne znaczenie w procesie klasyfikacji stanu. Te które nie zostaną wykorzystane przez algorytm nie mają znaczenia w budowie drzewa, a więc w klasyfikacji stanu czy identyfikacji uszkodzeń przeprowadzonych tą metodą. Dodatkowo metoda może być stosowana w zbiorach danych o licznych brakach danych co może mieć znaczenie w przypadku dostępnych baz pochodzących z kilku źródeł.

Innymi metodami klasyfikacji stosowanymi także w diagnostyce siników, są sieci neuronowe Sztuczne sieci neuronowe, wykorzystywane tutaj do klasyfikacji, pozwalają na przetwarzanie informacji w sposób równoległy. W przypadku sztucznej sieci neuronowej podstawowe znaczenie ma optymalizacja struktury sieci (liczba warstw ukrytych czy liczba neuronów w warstwach) co stanowi żmudny proces przeprowadzany przeważnie metodą prób i błędów. Ważny jest także dobór atrybutów wejściowych, co w przypadku drzew klasyfikacyjnych wbudowane jest bezpośrednio w algorytm jego budowy.

Innym klasyfikatorem stosowanym w tej pracy jest klasyfikator k – najbliższych sąsiadów. Jego zaletą jest niewątpliwie prostota realizacji i łatwość ustalenia w jaki sposób doszło do predykcji danej klasy. Wadami natomiast są konieczność przechowywania dużej ilości danych w pamięci, konieczność optymalizacji doboru parametru k, miary odległości pomiędzy danymi jak i zbioru atrybutów.

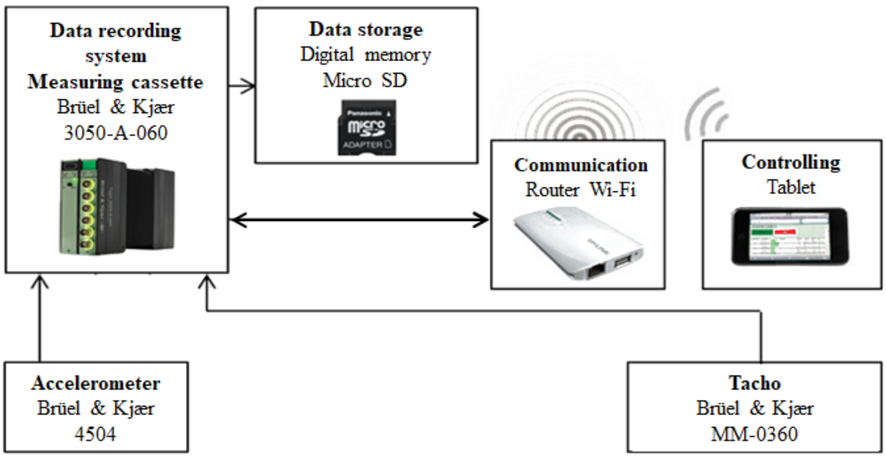
2. Metodyka badań

Obiektem badań był jednocylindrowy silnik badawczy SB 3.1 zbudowany na bazie silnika SW 680. Silnik SB 3.1 przeznaczony jest do prac badawczych z zakresu przebiegu spalania oraz oceny innych parametrów pracy silników SW 680 budowanych na licencji firmy Leyland w WSK Mielec. W silniku badawczym zastosowano następujące zespoły z silnika SW 680: korbowód, tłok z pierścieniami, tuleję cylindrową, zawory rozrządu, sterowanie rozrządu, wtryskiwacz, głowicę (wykonaną poprzez przeróbkę głowicy silnika SW 680). Konstrukcja silnika pozwala na: pomiar ciśnienia w cylindrze, regulację stopnia sprężania w zakresie ε =

14-20, możliwość zmiany początku tłoczenia paliwa w sposób ciągły, możliwość zainstalowania przekaźnika drogi tłoka, możliwość zmiany faz rozrządu w sposób ciągły, zmiany wyrównoważenia sił masowych I rzędu.

Badania przeprowadzono dwuetapowo. W pierwszym etapie przeprowadzono testy impulsowe mające na celu identyfikację częstotliwości rezonansowych co zdeterminowało pasmo pomiarowe oraz pozwoliło na wyznaczenie punktów akwizycji drgań na głowicy silnika spalinowego. W etapie drugim badano drgania głowic silnika dla różnych konfiguracji luzu zaworów i nastaw punktu pracy silnika. Na podstawie wyników badań etapu drugiego, opracowano algorytm oceny luzu zaworów w badanym silniku.

Ogólny schemat układu pomiarowego stosowanego do rejestracji sygnałów drganiowych podczas badań uruchomionego silnika przedstawiono na rys. 2.



Rys. 2. Schemat układu pomiarowego stosowanego do pomiarów drgań głowicy w czasie badań uruchomionego silnika

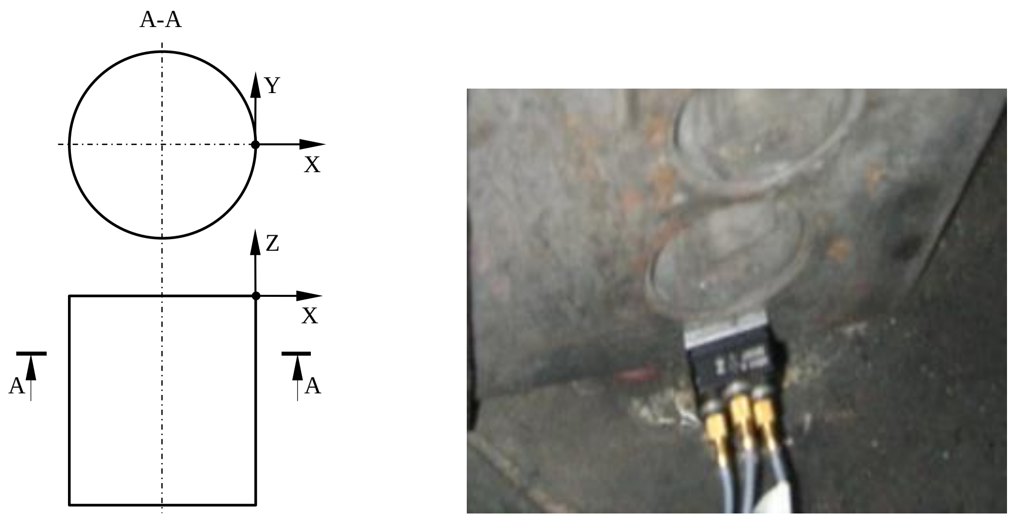
Metodyka badań została opracowana w oparciu o założenia eksperymentu czynnego [22]. Podczas wykonywania eksperymentu zmieniano: luz zaworów, obciążenie silnika spalinowego oraz prędkość obrotową wału korbowego, równocześnie rejestrowano przyspieszenia drgań głowicy silnika.

Pomiary wykonano zgodnie z zasadą trzech uruchomień tzn. każdą serię pomiarów wykonano trzykrotnie, pomiędzy każdą serią pomiarów dokonano wyłączenia silnika z ruchu. Wykorzystano opisaną metodę prowadzenia badań, chcąc uniknąć przypadkowych wartości parametrów charakterystyk sygnałów drganiowych.

Przetworniki drgań typu 4504 firmy Brüel&Kjær wybrano na podstawie wskazówek zawartych w pracy [26, 30] liniowe pasmo przenoszenia wybranych przetworników wynosiło do 18 kHz. Podczas badań rejestrowano sygnały w paśmie 0,1 Hz–25 kHz. Akcelerometry zamocowano na głowicy silnika spalinowego za pomocą kleju. Przy wyborze miejsc pomiarowych do testów impulsowych przyjęto zasadę, że przetwornik powinien znajdować się jak najbliżej miejsca generacji sygnału drgań związanego z pracą zaworów oraz w miejscu dostępnym [25]. Kierunki pomiaru drgań przyjęto następująco: kierunek X równoległy do średnicy cylindra, kierunek Z równoległy do osi cylindra, kierunek Y prostopadły do dwóch pozostałych (rys. 3a). Częstotliwość próbkowania ustawiono na 65536 Hz. Do rejestracji sygnałów drgań zastosowano multianalizator PULSE firmy Brüel&Kjær, który umożliwia rejestrację przebiegów szybkozmiennych równolegle na 6 kanałach z dynamiką do 160 dB.

Orientację przestrzenną przetworników do pomiaru drgań przedstawiono na rys. 3a natomiast miejsce ich mocowania na badanym obiekcie na rys. 3b.

a) b)



Rys. 3. a). Orientacja kierunków pomiaru drgań na głowicy cylindrów silnika, b) widok przetwornika drgań zamocowanego na głowicy cylindrów silnika

Wybór punktu pomiaru drgań został poprzedzony analizą konstrukcji głowicy, badaniami opisanymi w pracy [26] dotyczącymi określenia wpływu luzu zaworów silnika spalinowego o ZS na wybrane parametry drgań oraz wspomnianymi wcześniej testami impulsowymi, polegającymi na uderzaniu zaworów o gniazda zaworowe. Uderzenia realizowano przez usuwanie, umieszczonej pomiędzy trzonkiem zaworu a dźwigienką, wzorcowej płytki pomiarowej. Powtarzano to kilkakrotnie dla każdego zaworu w celu wyeliminowania błędów przypadkowych oraz wykonania procesu uśredniania. Ważne było ustalenie takiego punktu pomiaru sygnałów drganiowych, który umożliwiałby ocenę uderzenia każdego zaworu. Po wykonaniu analiz wyników testów impulsowych do badań eksploatacyjnych wybrano jeden punkt. Biorąc pod uwagę dynamikę sygnałów zarejestrowanych podczas badań impulsowych zauważono, że kierunek X może nieść najwięcej informacji dotyczącej luzu zaworowego.

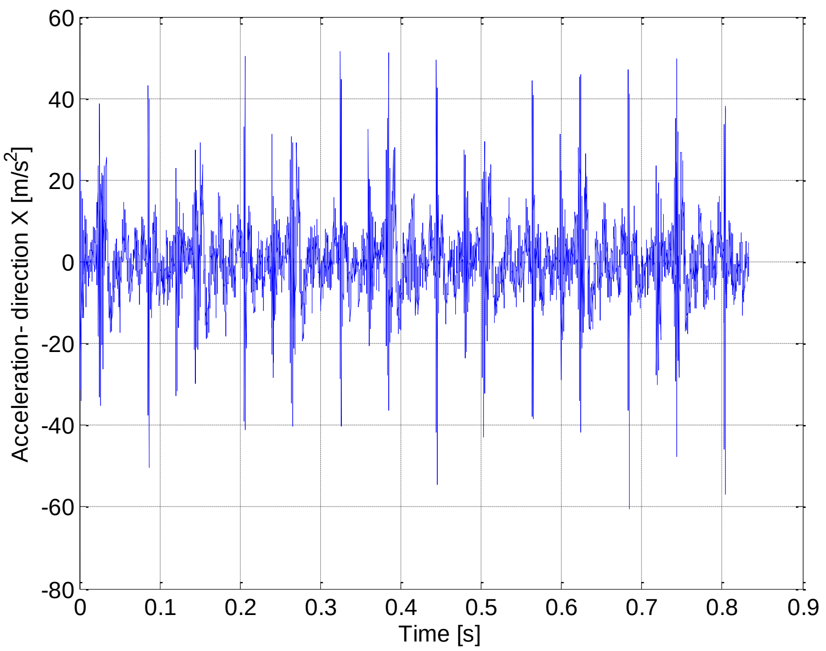
Badania pracującego silnika przeprowadzono przy: prędkości obrotowej wału korbowego 700 obr/min, 1000 obr/min, 1200 obr/min, 1500 obr/min, 1700 obr/min, obciążeniu: brak obciążenia zewnętrznego, 22.5 Nm, 45 Nm, 67.5 Nm, 90 Nm., temperaturze cieczy chłodzącej utrzymywano na poziomie 75°C. Dla wymienionych wyżej warunków dokonano rejestracji sygnału przyspieszeń drgań w trzech kierunkach.

Przykładowy fragment rejestracji w kierunku X sygnału drganiowego przedstawiano na rysunku 4.

W przebiegu widocznych jest szereg cykli pracy silnika. Dominującym zjawiskiem są chwilowe zdarzenia związane z samym zapłonem, natomiast zjawiska drganiowe związane z zamknięciem zaworu są trudniej dostrzegalne zwłaszcza dla niewielkich wartość luzu zaworowego.

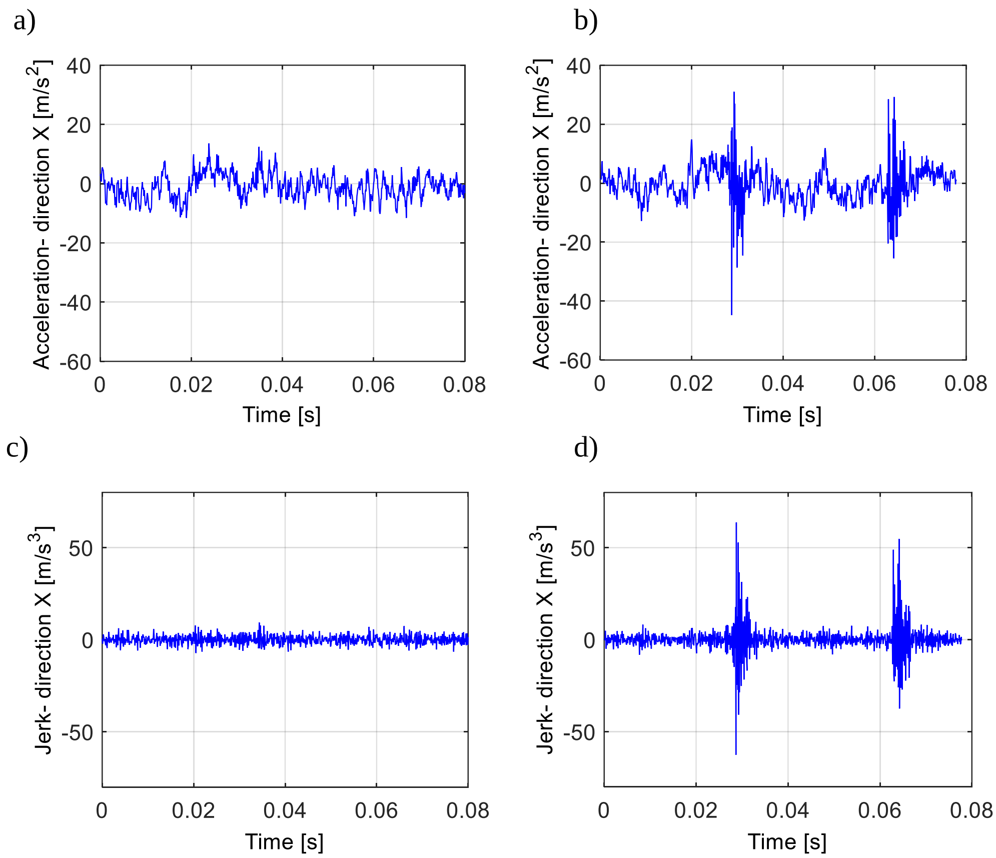
Przed analizą sygnały drganiowe poddano selekcji kątowej. Oznacza to, że analizowano tylko te fragmenty sygnałów, które były zsynchronizowane czasowo lub kątowo z procesem zamykania zaworów.

W praktyce można tego dokonać w łatwy sposób dysponując odpowiednimi znacznikami na kole zamachowym, a nawet w oparciu sam sygnał przyspieszeń drgań biorąc pod uwagę próg wartości szczytowej osiąganej podczas zapłonu silnika i odpowiednią bramkę czasową. W ten sposób uzyskano ponad 32 tys. wycinków sygnału czasowego dla różnych obciążeń i prędkości obrotowej silnika.



Rys.4. Przykładowy przebieg przyspieszeń drgań rejestrowanych w kierunku X – fragment rejestracji

Rysunek 5 przedstawia przykładowe wycinki rejestracji dla skrajnych wartości luzu zaworowego 0,3 mm oraz 1,0 mm.



Rys.5. Fragmenty przebiegów czasowych przyspieszeń (rysunki a,b) i pochodnej przyspieszenia drgań (rysunki c,d) związanych z zamykaniem zaworów; rysunek a, c – luz = 0,3 mm; rysunek b, d – luz = 1,0 mm

W przypadku niewielkiego luzu moment zamknięcia zaworu jest praktycznie niewidoczny (rys. 5a), natomiast w przypadku skrajnie dużego luzu łatwo go dostrzec (rys. 5b). Widoczna jest wyraźna odpowiedź układu spowodowana uderzeniem zaworów w gniazdo zaworowe. Najtrudniejsze do analizy są przykłady pośrednie. Należy także zauważyć, że poszczególne fragmenty nawet dla tego samego luzu, różnią się od siebie co widoczne jest już na rysunku 4, a uzyskane zbiory obarczone sporym rozrzutem. Ilustruje to tabela 1 gdzie przedstawiono wartości skuteczne przyspieszeń drgań dla wyciętych fragmentów sygnałów oraz ich odchylenia standardowe.

Konstrukcyjna wartość luzu zaworowego wynosi dla badanego sinika 0,5 mm. Jak wynika z analizy tabeli 1, wobec znacznej dyspersji wyników prawidłowe rozpoznanie luzu zaworowego na podstawie tylko wartości skutecznej jest niemożliwe. Można założyć, że identyfikacja luzu w oparciu same miary amplitudowe może być obarczona znaczą niepewnością. Wobec powyższego wydaje się zasadna ocena sygnałów w świetle wielu miar, a następnie ich selekcja ze względu na największy wkład informacji w detekcji klasy uwzględniającej dany luz zaworowy.

Tab.1. Zestawienie średniej wartości skutecznej i jej odchylenia standardowego mierzonej w pasmie do 6kHz dla różnych kierunków pomiarowych i różnych wartości luzu zaworowego silnika

Luz [mm]	Kierunek X		Kierunek Y		Kierunek Z	
	Średnie rms m/s^2	Odchylenie standardowe m/s^2	Średnie rms m/s^2	Odchylenie standardowe m/s^2	Średnie rms m/s^2	Odchylenie standardowe m/s^2
0,3	9.40	4.32	12.47	4.29	10.43	3.50
0,4	9.10	4.49	12.74	4.01	10.48	3.52
0,5	10.11	4.64	12.99	4.29	12.15	4.45
0,6	8.25	2.59	13.12	3.97	11.01	3.80
0,7	10.43	4.42	13.71	4.29	12.48	4.47
0,8	9.35	2.85	13.59	4.07	12.96	4.48
0,9	11.78	5.51	13.41	4.02	14.05	4.73
1,0	12.33	5.08	14.29	4.40	16.53	5.50

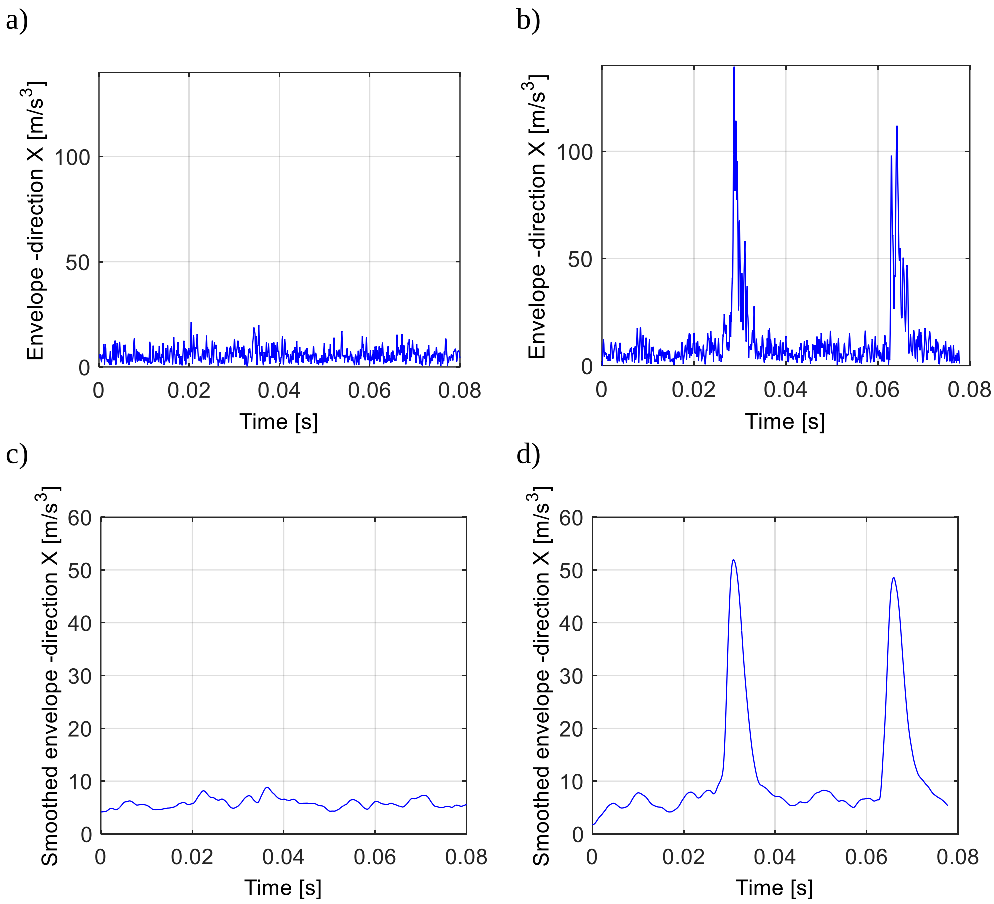
W celu uwypuklenia zjawisk, a także zmniejszenia wpływu niskich częstotliwości wyznaczono także zmianę przyspieszeń w czasie (jerk). Na rysunku 5c i 5d przestawiono przebiegi tej wielkości odpowiadające zaprezentowanym przebiegom przyspieszeń drgań.

W celu wydobycia najbardziej istotnej informacji z wyizolowanych przebiegów wyznaczono obwiednię sygnałów stosując transformatę Hilberta i wyznaczając magnitudę sygnału analitycznego. Dodatkowo dokonano wygładzenia obwiedni modelem wyrównywania wykładniczego Browna:

$$\begin{cases} \hat{S}_1 = S_1 \\ \hat{S}_t = \alpha S_t + (1-\alpha)\hat{S}_{t-1} \quad \text{dla} \quad t > 1 \end{cases} \tag{1}$$

gdzie: S_t – wartości pierwotnych obserwacji, $\hat{S}_t$ - wartości obserwacji po wygładzeniu, α - współczynnik wygładzania (tutaj przyjęty arbitralnie jako α=0,1) oraz t – numer obserwacji.

Na rysunkach 6a i 6b przedstawiono uzyskaną obwiednię przebiegów z rysunku 5c i 5d oraz wygładzoną modelem wyrównywania wykładniczego obwiednię (rysunki 6c i 6d).



Rys.6. Obwiednia sygnału pochodnej przyspieszenia dla dwóch skrajnych przypadków luzu zaworowego (rysunki a i b), oraz wygładzona obwiednia (rysunki c i d); rysunek a, c – luz = 0,3 mm; rysunek b, d – luz = 1,0 mm

W celu wytrenowania systemu klasyfikacji w rozróżnianiu klas luzu zaworowego dokonano parametryzacji otrzymanych wycinków przebiegów przyspieszeń drgań, jego pochodnej oraz wygładzonej obwiedni. Uzyskaną obwiednię podzielono na dwa fragmenty odpowiadające okresom czasu, w których następuje zamknięcie pierwszego i drugiego zaworu.

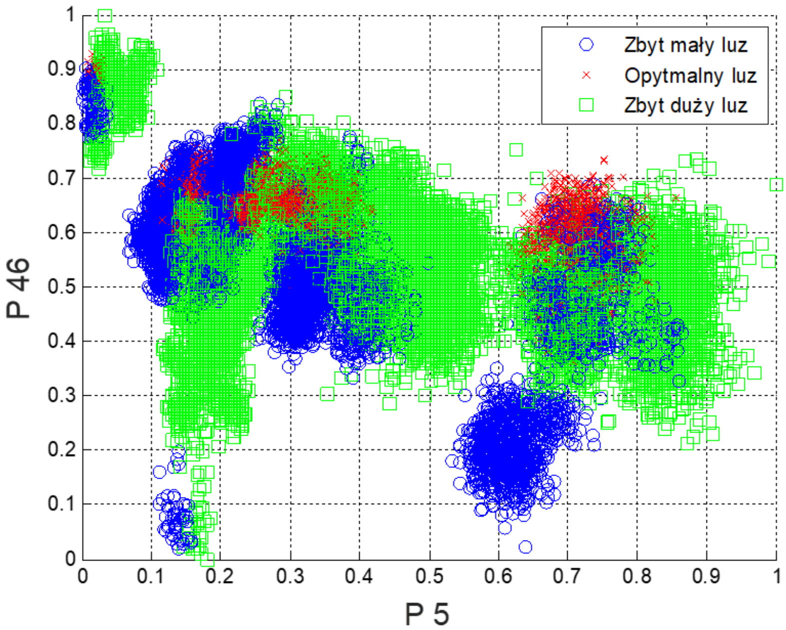
Dla pierwotnego sygnału przyspieszeń, jego pochodnej względem czasu oraz sygnału odfiltrowanego w paśmie powyżej 2000Hz zastosowano miary sygnału takie jak: odcięta środka ciężkości kwadratu sygnału, momenty zwykłe i centralne rzędu pierwszego i drugiego, momenty unormowane tych samych rzędów, wartość skuteczna sygnału, wartość szczytowa, między szczytowa, wartość średnia, pole powierzchni pod krzywą, współczynnik szczytu, kształtu, impulsowości, kurtoza sygnału. Dla obwiedni zastosowano zliczenie próbek powyżej zadanych poziomów (9 przyjętych poziomów), przy czym obwiednię podzielono na dwa fragmenty związane z zamknięciem pierwszego jak i drugiego zaworu. Analizy takie powtórzono dla trzech rejestrowanych kierunków. Ostatecznie wygenerowano ponad 300 parametrów. Dodatkowo jako atrybuty rozpatrzono informację o obciążeniu silnika oraz jego prędkości obrotowej). Dostępne próbki sygnału pozwoliły na wygenerowanie 32054 wektorów uczących. W porównaniu do liczby parametrów opisujących wektory uczące jest to mało liczny zbiór przykładów, stąd jak opisano dalej, dokonano selekcji cech diagnostycznych redukując wymiar wektorów uczących do 15 parametrów.

Każdy wektor uczący opatrzono etykietą związana z luzem, przy czym dokonano podziału danych na trzy klasy: luz zbyt mały (mniejszy od 0,5 mm), luz optymalny (ok. 0,5 mm), luz zbyt duży (ponad 0,5 mm). W tabeli 2 zestawiono liczności dostępnych przykładów dotyczących odpowiednich klas. Jak wynika z tabeli liczby przykładów dla poszczególnych klas znacznie się różnią co determinuje sposób oceny wyników błędów klasyfikacji.

Tab. 2. Liczność dostępnych przykładów reprezentujących klasy

Klasa	Liczność przykładów
Zbyt mały luz	8160
Luz optymalny	4016
Zbyt duży luz	19878

Na rysunku 7 przedstawiono przykładową przestrzeń cech utworzoną przez dwa znormalizowane (normalizacja typu min-max) atrybuty oznaczone jako P5 i P46.



Rys. 7. Dostępne przykłady uczące na płaszczyźnie wyznaczonej przez dwa przykładowo wybrane atrybuty

3. Analiza danych

W pracy zaproponowano system identyfikacji luzu zaworowego pracującego silnika bazujący na metodach uczenia maszynowego. W tym celu porównano trzy metody: zbiór trzech drzew binarnych typu CART (Classification and Regression Tree) z zastosowaniem strategii OvA – one versus all, klasyfikator *k*- najbliższych sąsiadów (*k*-Nearest Neighbors), oraz sieć neuronową jednokierunkową MLP (multilayer perceptoron) z trzema wyjściami związanymi z każdą z klas. Klasyfikator bazujący na zbiorze trzech drzew uczonych niezależnie rozpoznawać poszczególne klasy dotyczące luzu zaworowego. O przynależności do danej klasy decydowała pozytywna odpowiedź jednego z trzech wytrenowanych klasyfikatorów. Przy takim podejściu możliwe są sytuacje, w których żaden klasyfikator nie daje na wyjściu informacji o przynależności do danej klasy, jak i takie, w których pozytywnie odpowiada więcej niż jedno drzewo. Przyjęto rozwiązanie, w którym brak jednoznacznej klasyfikacji wymusza kolejną. W praktyce, gdy wystąpi taka sytuacja, system rozpoznawania dokona klasyfikacji kolejnych odcinków czasowych pomijając przypadki niepewne. Można również wielokrotnie powtarzać klasyfikację praktycznie w czasie rzeczywistym i podjąć decyzję na podstawie istotnej większości rozpoznań danej klasy. Wobec generacji dużej liczby danych przez pracujący silnik, w relatywnie krótkim okresie czasu, zignorowanie niepewnych klasyfikacji nie rodzi żadnych problemów w praktyce. Podobny problem występuje także w zaproponowanym rozwiązaniu opartym o sieć neuronową. Tutaj można zastosować podejście, w którym o przynależności do danej klasy decyduje neuron w warstwie wyjściowej, który daje największą wartość wyjścia. Jest mało prawdopodobne, aby na więcej niż jednym wyjściu pojawiła się dokładnie taka sama wartość. Jednak może się zdarzyć, że odpowiedzi na wszystkich wyjściach sieci będą relatywnie na małym poziomie, co może zostać zinterpretowane jako przypadek nierozpoznany. Przyjęto, metodą prób i błędów, że wartość na wyjściu sieci musi przekroczyć próg 0,7, aby można było uznać przypisane danego przypadku do określonej klasy. W przypadku nieprzekroczenia tej wartości na żadnym wyjściu sieci uznaje się, że wystąpił stan nierozpoznany i sposób postępowania jest dokładnie taki sam jak w przypadku zbioru drzew klasyfikacyjnych.

W celu redukcji wymiaru wektorów cech określono, które z cech są najbardziej przydatne do klasyfikacji luzu zaworowego. W tym celu dokonano w pierwszej kolejności klasyfikacji za pomocą omówionego wcześniej klasyfikatora w postaci zbioru drzew klasyfikacyjnych. Jest wiadome, że algorytmy budowy drzewa stosują określona miarę dobroci podziału i oceniają poszczególne cechy na podstawie tej miary. W zastosowanym algorytmie jako miarę jakości podziału wykorzystano indeks Giniego. Pod uwagę wzięto tylko te miary spośród wszystkich dostępnych, które posłużyły do budowy drzew. Następnie, na podstawie kolejnego kroku analizy, stwierdzono, że niektóre z tych miar są istotnie skorelowane liniowo. Po usunięciu niektórych z nich z zestawu danych i testach nie zauważono istotnego wzrostu błędu klasyfikacji za pomocą zbioru drzew. Finalnie otrzymano zbiór piętnastu cech, którego dalsze zmniejszanie objawiało się mniejszym lub większym wzrostem błędu testowania. Ostateczny zbiór cech zawierał między innymi takie miary jak: wartość skuteczna i współczynnik kształtu sygnału przyspieszeń, momenty zwykłe i centralne rzędu pierwszego i drugiego z sygnału przyspieszeń, wartość średnia i maksymalna, kurtoza z pochodnej sygnału przyspieszeń, liczby zliczeń próbek powyżej określonych poziomów (o małych wartościach) dla obwiedni pochodnej sygnału przyspieszeń i sygnału odfiltrowanego oraz prędkość obrotowa. Zdecydowana większość wybranych przez algorytm budowy drzewa miar dotyczyły kierunków pomiarowych X i Y. Informacja o obciążeniu nie okazała się istotna.

W celach porównawczych, te same cechy zastosowano w przypadku pozostałych klasyfikatorów. Oczywiście można zastosować tutaj także inne metody selekcji cech. Należy się liczyć z tym, że być może wyłoniłby one inne atrybuty, lepsze dla innych klasyfikatorów. Porównanie jednak wielu możliwości nie było tutaj celem badań.

W klasyfikatorze k- najbliższych sąsiadów zmieniano wartość parametru k w zakresie od 1 do 11 oraz miarę odległości obserwując błąd testowania. Ostatecznie parametr ten przyjęto jako k=6 i metrykę euklidesową.

Podobnie dokonano optymalizacji sieci MLP dokonując porównania błędów na zbiorze testowym. Najlepsze wyniki udało się uzyskać dla sieci z dwoma warstwami ukrytymi z 7 i 6 neuronami sigmoidalnymi w poszczególnych warstwach. Zwiększanie jak i zmniejszanie liczby neuronów skutkowało zwiększaniem błędu uczenia. Do uczenia zastosowano algorytm gradientowy Levenberga – Marquardta.

Za każdym razem ocena błędów klasyfikatorów odbyła się za pomocą powtarzanej metody Hold-Out, przy czym jako zbiór testowy przyjęto arbitralnie 25% przykładów oraz przyjęto 100 powtórzeń testu. Należy zaznaczyć, że metoda ta ma tendencję do przeszacowywania błędu rozpoznawania [26]. Uzyskana wartość odchylenia standardowego podczas powtarzania testu pozwala na określenie jak odporny jest dany algorytm klasyfikacyjny na zmiany danych uczących. Takie rozwiązanie ma także swoje wady gdyż jest mało prawdopodobne, że uda się podczas testów wyczerpać cały zestaw danych.

Ze względu na znaczną różnicę w liczności przedstawicieli różnych klas, w celu porównania metod obliczono błąd ważony klasyfikacji wyrażony wzorem:

$$\varepsilon = \frac{1}{K}\sum_{i=1}^{K}\frac{\sum_{j=1,\, j\neq i}^{K} a_{ij}}{K_i} \tag{2}$$

gdzie: K – liczba klas, K_i – liczba elementów w i – tej klasie, a_{ij} elementy macierzy rozkładu klas (macierzy pomyłek) z poza przekątnej.

Taka definicja błędu pozwala uwzględnić zróżnicowanie liczności przykładów z różnych klas.

W tabeli 3 przedstawiono uśrednione wyniki uzyskane podczas testów porównywanych metod, natomiast poniżej macierze rozkładu klas. W przypadku sieci neuronowych jak i klasyfikatora k-NN zaprezentowano wyniki dla najlepszych klasyfikatorów (o najlepiej dobranej strukturze sieci w przypadku klasyfikatora *k*-NN o optymalnym parametrze *k* jak i mierze odległości). Ze względu na zróżnicowaną liczność reprezentantów danej klasy liczbę pomyłek odniesiono do liczby przedstawicieli danej klasy.

Tab.3. Wyniki testowana klasyfikatorów

Metoda	Średni błąd ważony klasyfikacji [%]	Odchylenie standardowe błędu - miara odporności klasyfikatora na zmianę zbioru uczącego [%]	Udział klasyfikacji niepewnych [%]
Zbiór drzew	0,93	0,13	3,14
k-NN (k=6)	1,98	0,30	0
MLP [7,6]	1,73	0,70	3,27

W tabelach 4-6 zaprezentowano macierze rozkładu klas dla różnych metod identyfikacji klasy.

Tab.4. Macierz rozkładu klas dla zbioru drzew klasyfikacyjnych

		Klasy rzeczywiste		
		Zbyt mały luz [%]	Optymalny luz [%]	Zbyt duży luz [%]
Klasy rozpoznane	Zbyt mały luz	99,4	0,4	0,2
	Optymalny luz	0,1	98,1	0,1
	Zbyt duży luz	0,5	1,5	99,7

Tab. 5. . Macierz rozkładu klas dla klasyfikatora k-NN (k=6)

		Klasy rzeczywiste		
		Zbyt mały luz [%]	Optymalny luz [%]	Zbyt duży luz [%]
Klasy rozpoznane	Zbyt mały luz	99,2	1,1	0,4
	Optymalny luz	0,4	96,0	0,8
	Zbyt duży luz	0,4	2,9	98,8

Tab.6. Macierz rozkładu klas dla klasyfikatora w postaci sieci neuronowej [7,6] (7 i 6 neuronów w warstwach ukrytych)

		Klasy rzeczywiste		
		Zbyt mały luz [%]	Optymalny luz [%]	Zbyt duży luz [%]
Klasy rozpoznane	Zbyt mały luz	98,8	0,6	0,7
	Optymalny luz	0,2	97,1	0,4
	Zbyt duży luz	1,0	2,3	98,9

Z porównania klasyfikatorów wynika, że najkorzystniejszym spośród rozpatrywanych rozwiązaniem identyfikacji klasy związanej z luzem zaworowym dla badanego silnika jest zbiór klasyfikatorów binarnych. Klasyfikator *k*-NN w zastosowanym algorytmie nie identyfikuje przypadków niepewnych. Tłumaczy to większy błąd klasyfikacji niż w przypadku pozostałych klasyfikatorów. „Trudne" przypadki zostają sklasyfikowane, a nieuznawane za niepewne. Istnieje oczywiście możliwość wprowadzenia algorytmu (k,l) najbliższych sąsiadów, w którym w przypadku zbyt małej liczby sąsiadów głosujących za daną klasą (mniej niż *l*), klasyfikację uznajemy za niepewną. Jednak wprowadza to kolejny parametr *l* do optymalizacji klasyfikatora. W przypadku zbioru drzew jak i sieci, jak wspomniano poprzednio, klasyfikacje niepewne są odrzucane, a decyzja co do przynależności do klasy odbywać się może na podstawie kolejnych próbek sygnału. Po odrzuceniu niepewnych klasyfikacji zbiór drzew pozwala na klasyfikację luzu z dokładnością rzędu 99% pomimo wpływu różnego obciążenia i prędkości obrotowej. W przypadku pozostałych klasyfikatorów wyniki są słabsze (ok. 98%) jednak akceptowalne, zwłaszcza, że dla postawionego problemu można, w krótkim czasie wielokrotnie powtórzyć diagnozę.

Klasyfikator oparty o drzewa klasyfikacyjne cechował się również największą odpornością na zmianę zbioru uczącego mierzoną za pomocą odchylenia standardowego błędu. Wszystkie klasyfikatory popełniają najwięcej błędów w prawidłowym rozpoznaniu luzu optymalnego i najczęściej jest on mylony ze zbyt dużym luzem (odpowiednio 1,5%, 2,9 % i 2,3 % przypadków luzu optymalnego zostało sklasyfikowanych jako luz zbyt duży). Najlepsze wyniki pod tym względem uzyskuje zbiór drzew klasyfikacyjnych. W przypadku tego klasyfikatora najlepiej, z pośród porównywanych metod, identyfikowany był także nadmierny luz zaworowy.

4. Podsumowanie

W wyniku przeprowadzonych analiz można zaproponować metodę klasyfikacji luzu zaworowego silnika w oparciu o sygnały drganiowe mierzone na głowicy. Dla rozpatrywanego silnika udaje się to zrobić z trafnością ok. 99%. Biorąc pod uwagę możliwość wielokrotnego powtórzenia procesu klasyfikacji w czasie prawie rzeczywistym i wybierając najczęściej występującą klasę można podjąć decyzję o klasie luzu zaworowego z dużą dozą pewności. Do określenia tego stanu w przypadku badanego silnika wystarcza 15 parametrów liczbowych nietrudnych do uzyskania. Oczywiście w przypadku innych typów silnika będzie konieczne przeprowadzenie na nowo całego procesu budowy klasyfikatora.

W pracy zaproponowano zastosowanie do rozwiązania przedstawionego problemu zbiór trzech drzew binarnych wyspecjalizowanych w identyfikacji każdej klasy luzu zaworowego. Takie podejście pozwala osiągnąć relatywnie małe błędy klasyfikacji (dla rozpatrywanego przypadku) jak i automatyzuje dobór składowych wektora uczącego. Dodatkowo drzewo pozwala w prosty sposób wygenerować czytelne dla człowieka reguły, łatwe także do zaimplementowania w systemie, który miałby działać automatycznie.

Źródło finansowania

Prezentowane wyniki badań, zostały współfinansowane z subwencji przyznanej przez Ministerstwo Nauki i Szkolnictwa Wyższego 05/52/SBAD/0295 and 02/21/SBAD/3558

Bibliografia

1 Albarbar A, Ball A, Starr A. On acoustic measurement-based condition monitoring of internal combustion engines. Insight: Non-Destructive Testing and Condition Monitoring 2008; 50(1): 30–34.

2 Albertson F, Bodén H, Gilbert J. Comparison of different methods to couple nonlinear source descriptions in the time domain to linear system descriptions in the frequency domain - Application to a simple valveless one-cylinder cold engine. Journal of Sound and Vibration 2006; 291(3–5): 963–985.

3 Arroyo J, Muñoz M, Moreno F, Bernal N, et al. Diagnostic method based on the analysis of the vibration and acoustic emission energy for emergency diesel generators in nuclear plants. Applied Acoustics 2013; 74(4): 502–508.

4 Babu A K, Raj A A, Kumersan G. Misfire detection a multi- cylinder diesel engine: a machune learning approach. Journal of Engineering Science and Technology 2016; 11(2): 278–295.

5 Badawy T, Shrestha A, Henein N. Detection of Combustion Resonance Using an Ion Current Sensor in Diesel Engines. Journal of Engineering for Gas Turbines and Power 2012; 134(5): 052-802.

6 Ahmed A B, Elaraby I S. A prediction for Student's Performance Using Decision Tree ID3 Method. India - World Journal of Computer Application and Technology 2014; 2(2): 43–47.

7 Czech P, Bąkowski H. Diagnosing of car engine fuel injectors damage using dwt analysis and PNN neural networks. Transport Problems 2013; 8(3): 85–91.

8 Delvecchio S, Bonfiglio P, Pompoli F. Vibro-acoustic condition monitoring of Internal Combustion Engines: A critical review of existing techniques. Mechanical Systems and Signal Processing 2018; 99: 661–683.

9 Desbazeille M, Randall R B, Guillet F, El Badaoui M, et al. Model-based diagnosis of large diesel engines based on angular speed variations of the crankshaft. Mechanical Systems and Signal Processing 2010; 24(5): 1529–1541.

10 Dolatabadi N., Theodossiades S., Rothberg S.J., On the identification of piston slap events in internal combustion engines using tribodynamic analysis. Mechanical Systems and Signal Processing 2015; 58: 308–324.

11 Erbek F S, Özkan C, Taberner M. Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. International Journal of Remote Sensing 2004; 25(9): 1733–1748

12 Figlus T, Liščák Š, Wilk A, Łazarz B. Condition monitoring of engine timing system by using wavelet packet decomposition of a acoustic signal. Journal of Mechanical Science and Technology 2014; 28(5): 1663–1671.

13 Fog T.L., Hansen L.K., Larsen J., Hansen H.S., et al., On condition monitoring of exhaust valves in marine diesel engines. Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop 1999; 554–563.

14 Gao F, Lv J. Fault Diagnosis for Engine Based on Single-Stage Extreme Learning Machine. Mathematical Problems in Engineering 2016; 2016: 1–10.

15 Gorges C, Öztürk K, Liebich R. Impact detection using a machine learning approach and experimental road roughness classification. Mechanical Systems and Signal Processing 2019; 117: 738–756.

16 Jia F, Lei Y, Lu N, Xing S. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. Mechanical Systems and Signal Processing 2018; 110: 349–367.

17 Korbicz J, Kościelny J M, Kowalczuk Z, Cholewa W (Eds.). Models, Artificial Intelligence, Applications. Springer-Verlag Berlin Heidelberg, 2004.
18 Lahrache A, Cocconcelli M, Rubini R. Anomaly detection in a cutting tool by k-means clustering and support vector machines. Diagnostyka 2017; 18(3): 21–29.
19 Leemans V, Magein H, Destain M F. Defect segmentation on 'Jonagold' apples using colour vision and a Bayesian classification method. Computers and Electronics in Agriculture 1999; 23: 43–53.
20 Lei Y, He Z, Zi Y. Application of an intelligent classification method to mechanical fault diagnosis. Expert Systems with Applications 2009; 36(6): 9941–9948.
21 Madej H, Czech P. Discrete wavelet transform and probabilistic neural network in IC engine fault diagnosis. Eksplotacja i Niezawodnosc - Maintenance and reliability 2010; 4(48): 47–54.
22 Matzke W. Four-stroke engine timing, Warszawa: WKiŁ, 1967.
23 Mechefske C K, Mathew J. Fault detection and diagnosis in low speed rolling element bearings Part II: The use of nearest neighbour classification. Mechanical Systems and Signal Processing 1992; 6(4): 309–316.
24 Nguyen T T, Armitage G. A survey of techniques for internet traffic classification using machine learning. IEEE Communications Surveys Tutorials 2008, 10(4): 56–76.
25 Niziński S, Michalski R. Diagnostics of technical objects. Radom: Department of Vehicle and Machine Operation, Faculty of Technical Sciences, University of Warmia and Mazury in Olsztyn, Institute for Sustainable Technologies in Radom, 2002.
26 Osiecki J, Ziemba S. Basics of mechanical vibration measurements, Warszawa: PWN, 1968.
27 Potočnik P, Govekar E. Semi-supervised vibration-based classification and condition monitoring of compressors. Mechanical Systems and Signal Processing 2017; 93: 51–65.
28 Qinghua W, Youyun Z, Lei C, Yongsheng Z. Fault diagnosis for diesel valve trains based on non-negative matrix factorization and neural network ensemble. Mechanical Systems and Signal Processing 2009; 23(5): 1683–1695.
29 Ruiz M, Mujica L E, Alférez S, Acho L, et al. Wind turbine fault detection and classification by means of image texture analysis. Mechanical Systems and Signal Processing 2018; 107: 149–167.
30 Serridge M, Licht T R. Piezoelectric accelerometers and vibration preamplifiers, Nearum: Brüel & Kjær, 1987.
31 Szymański G.M., Tomaszewski F., Diagnostics of automatic compensators of valve clearance in combustion engine with the use of vibration signal. Mechanical Systems and Signal Processing 2016; 68–69: 479–490.
32 Wang C, Zhang Y, Zhong Z. Fault diagnosis for diesel valve trains based on time-frequency images. Mechanical Systems and Signal Processing 2008; 22(8), 1981–1993.
33 Wong P K, Zhong J, Yang Z, Vong C M. Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis. Neurocomputing 2016; 174: 331–343.
34 Wu J, Liu C. An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. Expert Systems with Applications 2009, 36: 4278–4286.
35 Zhang H, Li J, Huang Y, Zhang L. A Nonlocal Weighted Joint Sparse Representation Classification Method for Hyperspectral Imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2014; 7(6): 2056–2065

dr inż. Jakub Korta

E-mail: jakub.korta@ieee.org

dr hab. inż. Paweł Skruch

Akademia Górniczo - Hutnicza im. Stanisława Staszica w Krakowie
Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej
Katedra Automatyki i Robotyki
al. A. Mickiewicza 30, 30-059, Kraków, Poland
E-mail: pawel.skruch@agh.edu.pl

Koncepcja integracji komory parowej z obudową układów elektronicznych w celu zwiększenia ich niezawodności

Streszczenie: Systematycznie wzrastająca moc obliczeniowa oraz postępująca miniaturyzacja urządzeń elektronicznych stosowanych w pojazdach samochodowych powodują trudności w utrzymaniu temperatury pracy elementów półprzewodnikowych w dozwolonym zakresie, przyczyniając się do ich przedwczesnego zużycia, a w skrajnych przypadkach, uniemożliwiając nawet ich normalną pracę. Wydajne i trwałe układy chłodzące stają się więc nieodzownym komponentem współczesnych podzespołów samochodowych, o krytycznym znaczeniu dla ich niezawodności. Urządzeniami mogącymi w niedalekiej przyszłości wspomagać działanie układów chłodzenia systemów elektronicznych wykorzystywanych w motoryzacji są komory parowe (płaskie rurki cieplne), w których transport energii termicznej zachodzi poprzez przemianę fazową i samoistne przemieszczanie się czynnika roboczego. Współcześnie, tego rodzaju urządzenia nie są komercyjnie stosowane w układach chłodzenia sterowników samochodowych, pozostając przedmiotem prac badawczo-rozwojowych związanych z ich wpływem na szeroko pojętą niezawodność termomechaniczną urządzeń elektronicznych. W niniejszym artykule opisano koncepcję zintegrowania komory parowej z aluminiową obudową kontrolera elektronicznego pracującego w warunkach podwyższonej temperatury otoczenia, odpowiadającej warunkom użytkowania komponentów samochodowych. Ponadto, ocenie poddano wpływ zastosowania tego urządzenia na temperaturę pracy chłodzonego elementu półprzewodnikowego i jego niezawodność, wyrażoną jako przewidywany czas jego bezawaryjnego funkcjonowania.

Słowa kluczowe: *chłodzenie, elektronika, analiza termiczna, rurka cieplna, komora parowa, przemiana fazowa, motoryzacja, niezawodność*

1. Potrzeba chłodzenia urządzeń elektronicznych

W ostatnich latach, w branży motoryzacyjnej obserwuje się coraz większe zainteresowanie systemami elektronicznymi wysokiej mocy, pozwalającymi na podniesienie komfortu i bezpieczeństwa podróży. Istotnym obszarem wykorzystania wysokowydajnej elektroniki samochodowej są systemy multimedialne (infotainment) oraz różnego typu zaawansowane systemy wspomagania kierowcy, nazywane zbiorczo systemami ADAS (*ang.* Advanced Driver Assistance Systems). Działanie sterowników ADAS oparte jest na akwizycji różnego rodzaju sygnałów – pochodzących np. z modułów komunikacji bezprzewodowej, kamer, radarów, czujników ultradźwiękowych, itd. – i wysokowydajnym ich przetwarzaniu, co ze względów bezpieczeństwa, często wykonywane jest w czasie bliskim rzeczywistemu.

Rejestrowanie i analizowanie tak dużych strumieni informacji wymaga wykorzystania istotnych mocy obliczeniowych, czego częstym, niepożądanym efektem jest wytwarzanie nadmiernych ilości ciepła. W konsekwencji często prowadzi to do znaczącego podniesienia temperatury elementów elektronicznych, wydatnie skracając okres ich bezawaryjnego funkcjonowania, a w skrajnych przypadkach, skutkować może nawet zatrzymaniem ich pracy. Istotnym czynnikiem jest również temperatura otoczenia na działanie której narażone są komponenty samochodowe, której wartość nierzadko przekracza 50℃, utrudniając utrzymanie termicznych parametrów pracy elementów półprzewodnikowych poniżej dopuszczalnych wartości granicznych. Według wyników badań przytoczonych w [20], wysoka temperatura pracy systemów elektronicznych była przyczyną 49% przypadków wycofania wadliwych produktów ze sprzedaży, odnotowanych w branży samochodowej w latach 2005-2015. Jak podano w [6,16], obok drgań mechanicznych, wysoka temperatura pracy oraz jej cykliczne zmiany są najczęstszymi przyczynami zużycia urządzeń elektronicznych.

Przewiduje się, że pomimo ciągłej optymalizacji algorytmów obliczeniowych oraz zmian technologii produkcji układów scalonych, w niedalekiej przyszłości kompletne samochodowe systemy elektroniczne, w szczególności kontrolery jazdy autonomicznej, wytwarzać będą coraz większe strumienie ciepła, o wartościach przekraczających 1kW [8]. Aby umożliwić funkcjonowanie takich systemów, dąży się do opracowania wydajnych, niezawodnych i zarazem tanich metod chłodzenia, umożliwiających dyssypację strumieni ciepła dużej gęstości, rzędu kilkudziesięciu watów wytwarzanych przez elementy półprzewodnikowe o powierzchniach kilku cm^2. Przykładem takiego urządzenia może być układ Nvidia Xavier, dedykowany do zadań jazdy autonomicznej, wytwarzający do 30W TDP (*ang*. Thermal Design Power), na półprzewodniku o powierzchni $3.5cm^2$.

Niepożądany wpływ działania podwyższonej temperatury pracy komponentów elektronicznych na ich niezawodność jest szeroko opisywany w literaturze naukowo-technicznej. W [10] autorzy zaprezentowali przegląd mechanizmów zniszczenia wywołanych działaniem wysokiej temperatury obserwowanych na poziomie układów scalonych oraz ich wpływ na niezawodność tych urządzeń. Wynik oddziaływania cykli temperaturowych na niezawodność tranzystorów wysokiej mocy wykorzystywanych w motoryzacji opisano w [18]. Autorzy [11] przedstawili próbę eksperymentalnego oszacowania wytrzymałości zmęczeniowej elementów sterownika samochodowego narażonego na cykliczne oddziaływanie wysokich temperatur. Nowatorska metoda prowadzenia przyspieszonych badań starzeniowych tego typu została opisana w [19]. W [22] przedstawiono metodę predykcji niezawodności układów elektronicznych opartą na tzw. modelu fizyki uszkodzeń – PoF (*ang*. Physics of Failure), uwzględniającym zmiany temperatury pracy układu scalonego na jego wytrzymałość termomechaniczną. W [4] opisano analizę niekorzystnego wpływu termicznych warunków pracy na niezawodność urządzeń i systemów elektronicznych wykorzystywanych we współczesnych systemach ADAS oraz procedurę pozwalającą na diagnozowanie tego rodzaju problemów już na wstępnym etapie ich projektowania. Szczegółowy przegląd współcześnie stosowanych metod analitycznych i eksperymentalnych stosowanych w szacowaniu niezawodności komponentów elektronicznych opisano w [5]. Wykorzystanie techniki chłodzenia układów elektronicznych opartej na zjawisku zmiany fazy (odparowania i kondensacji) czynnika chłodzącego oraz jego wysoką wydajność opisano w [1] oraz w [3]. W obu przytoczonych przykładach zastosowano aktywne układy chłodzenia, wykorzystujące konwekcję wymuszoną do rozproszenia energii termicznej. Systemy takie wymagają szeregu dodatkowych urządzeń, jak wymiennik ciepła, pompa, zawory, przewody, etc., dlatego pomimo wysokiej wydajności, aktywne systemy chłodzenia elektroniki w motoryzacji stosowane są w ograniczonym zakresie. Wykorzystanie zjawiska przemiany fazowej zachodzącej w pasywnym urządzeniu chłodzącym – komorze parowej (płaskiej rurce cieplnej) – walidację modelu jego modelu numerycznego oraz porównanie z działaniem płyty miedzianej przedstawiono w [13]. Autorzy [17] przedstawili szczegółowe porównanie działania komór parowych i płytek miedzianych w różnych konfiguracjach, stosowanych jako rozpraszacze strumienia ciepła zintegrowane z elementami półprzewodnikowymi. W obu tych pracach wskazano na większą wydajność urządzeń wykorzystujących przemianę fazową.

W przestrzeni publicznej znaleźć można niewiele publikacji dotyczących zastosowania komór parowych lub rurek cieplnych do chłodzenia systemów elektronicznych wykorzystywanych w pojazdach transportu drogowego. W [12] opisowo przedstawiono przegląd potencjalnych możliwości wykorzystania dwufazowych urządzeń chłodzących w konstrukcjach samochodowych. W artykule tym wyszczególniono następujące potencjalne obszary zastosowania tego rodzaju urządzeń: kontrola temperatury reflektorów wykorzystujących diody LED, układy chłodzenia baterii pojazdów elektrycznych oraz elementów sterowników elektronicznych. Wykorzystanie rurek cieplnych w układzie chłodzenia motocyklowej lampy przedniej wyposażonej w diody LED opisano w [15]. W odniesieniu do klasycznego rozwiązania, autorzy zauważyli nieznaczną poprawę niezawodności tak zaprojektowanego urządzenia. W [14] zaproponowano zastosowanie komory parowej do dyssypacji energii termicznej generowanej przez moduł tranzystorów wysokiej mocy, odnotowując obniżenie temperatury poszczególnych komponentów o około 9℃. Ponadto, wykazano zmniejszenie gradientu temperatury na powierzchni modułu, co w wydatny sposób poprawia niezawodność mechaniczną całego ustroju. W [9] zaproponowano zastosowanie rurki cieplnej jako urządzenia wspomagającego chłodzenie komponentów radia samochodowego, działającego w temperaturze pokojowej, bez dodatkowych elementów wymuszających obieg powietrza. Opisany w artykule rezultat to redukcja temperatury głównego źródła ciepła o około 3℃.

Niniejszy artykuł przedstawia koncepcję zintegrowania niewielkiej komory parowej z obudową sterownika elektronicznego, pracującego w temperaturze otoczenia odpowiadającej warunkom użytkowania komponentów samochodowych. Celem przeprowadzonej i opisanej w artykule analizy jest zrozumienie wpływu zastosowania płaskiej rurki cieplnej na temperaturę złącza (t_j)

chłodzonego komponentu elektronicznego oraz oszacowanie zmiany długości oczekiwanego czasu bezawaryjnego funkcjonowania jego elementu półprzewodnikowego. Wobec ograniczonej dostępności prac poświęconych pasywnym systemom chłodzenia sterowników samochodowych opartych na zjawisku przemiany fazowej czynnika roboczego, artykuł ten stanowić może istotne źródło informacji dla inżynierów i naukowców zainteresowanych termomechaniczną niezawodnością urządzeń elektronicznych wykorzystywanych w pojazdach drogowych.

Rozdział 2 niniejszego artykułu przedstawia podstawowy podział metod chłodzenia urządzeń elektronicznych oraz opisuje fizyczne mechanizmy transportu ciepła, które należy w tym kontekście rozważać. Rozdział 3 prezentuje zasady działania pasywnych urządzeń wspomagających transport energii termicznej poprzez wykorzystanie przemiany fazowej czynnika roboczego. W rozdziale 4 zaprezentowano studium koncepcji integracji komory parowej z aluminiową obudową urządzenia elektronicznego. W rozdziale 5 przedstawiono szacowany wpływ tej modyfikacji na przewidywaną długość czasu użytkowania chłodzonego komponentu elektronicznego. Rozdział 6 zawiera podsumowanie artykułu i wnioski płynące z przedstawionych w nim rozważań.

2. Chłodzenie urządzeń elektronicznych

Systemy chłodzenia elektroniki podzielić można na dwie grupy: systemy pasywne i aktywne. Pierwsza grupa opiera swoje działanie wyłącznie o naturalne procesy wymiany ciepła, tj. przewodzenie, konwekcję i promieniowanie. Systemy aktywne wspomagają te zjawiska poprzez zastosowanie urządzeń wymagających zewnętrznych źródeł energii, takich jak wentylatory lub pompy cieczy chłodzących. Inny podział spotykany w literaturze przedmiotu zakłada podział na metody chłodzenia wykorzystujące czynnik roboczy niezmieniający stanu skupienia (np. chłodzenie powietrzem) oraz takie, w których czynnik roboczy zmienia fazę (np. odparowanie i kondensacja wody). Wybór odpowiedniej metody i urządzenia chłodzącego uzależniony jest od ilości ciepła jakie należy usunąć z chłodzonych komponentów. W rozwiązaniach spotykanych w branży samochodowej, preferowane są pasywne metody chłodzenia, ponieważ, ze względu na brak elementów ruchomych, ich zastosowanie wiąże się z mniejszym ryzykiem awarii. Ponadto, brak konieczności zasilania, pozytywnie wpływa na całościowy bilans energetyczny systemu.

Wspomniane mechanizmy wymiany ciepła, tj. przewodzenie, konwekcję oraz promieniowanie, opisać można za pomocą wzorów (1) – (3) [2]. W przypadku przewodzenia, ilość ciepła jaka przepływa przez element w jednostce czasu wyrażona jest poprzez wzór (1):

$$\dot{Q} = -k\frac{\Delta T A}{\Delta x} \tag{1}$$

gdzie: $\dot{Q}$ [W] – strumień ciepła mierzony w jednostce czasu, ΔT [K] – różnica temperatur w kierunku przepływu ciepła, Δx [m] – grubość elementu, A [m^2] – powierzchnia przekroju porzecznego elementu, $k[\frac{\mathrm{W}}{\mathrm{m \cdot K}}]$ – przewodność cieplna przewodzącego materiału.

Drugim wspomnianym mechanizmem wymiany ciepła jest konwekcja, która rozumiana jest jako makroskopowy ruch gazu lub cieczy. Jeśli wywołana jedynie poprzez niejednorodną gęstość płynu, konwekcja nazywana jest swobodną (naturalną). Wywołana obecnością czynnika zewnętrznego, np. działaniem urządzenia wentylacyjnego, nazywana jest konwekcją wymuszoną. Przepływ ciepła wywołany konwekcją opisać można równaniem (2):

$$\dot{Q} = Ah(T_s - T_\infty) \qquad (2)$$

gdzie: T_s [K] oraz T_∞[K] – kolejno temperatura źródła ciepła oraz otoczenia, mierzona w dostatecznie dużej odległości. Współczynnik h $[\frac{\mathrm{W}}{\mathrm{m^2K}}]$ to współczynnik wnikania ciepła, zależny od kształtu i stanu powierzchni, rodzaju konwekcji i innych czynników. Mechanizm konwekcji jest często wykorzystywany zarówno w pasywnych (radiatory) jak i aktywnych układach chłodzenia (wentylatory).

Ostatnim wspomnianym powyżej zjawiskiem fizycznym powodującym wymianę ciepła jest promieniowanie termiczne. Jest to mechanizm wywołany powstawaniem fal elektromagnetycznych, wzbudzanych przez poruszające się w materii cząsteczki posiadające ładunek elektryczny. Podobnie jak w przypadku powyższych zjawisk, również promieniowanie dotyczy każdej substancji, której temperatura jest wyższa niż zero bezwzględne. Ilość wypromieniowanego ciepła w jednostce czasu opisać można za pomocą wzoru (3):

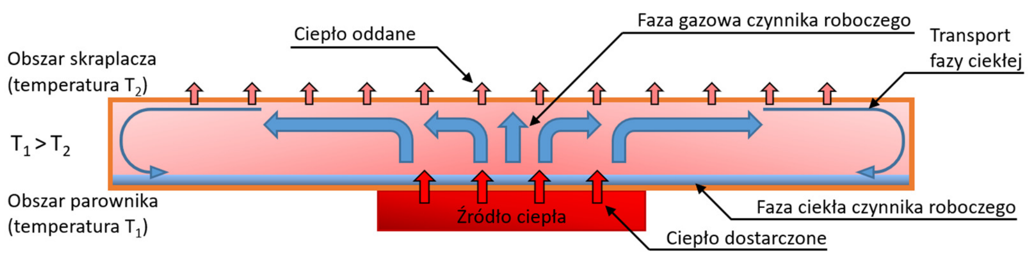
$$\dot{Q} = \varepsilon\sigma A(T_S^4 - T_A^4) \qquad (3)$$

gdzie: ε – bezwymiarowy współczynnik emisyjności zależny od rodzaju powierzchni źródła ciepła, $\sigma = 5.67x10^{-8}$ $[\frac{\mathrm{W}}{\mathrm{K^4m^2}}]$ – stała Stefana – Boltzmanna, T_A[K] – temperatura otoczenia.

W praktyce, wszystkie opisane zjawiska jednocześnie przyczyniają się do dyssypacji energii termicznej chłodzonych urządzeń, umożliwiając ich pracę w dopuszczalnym zakresie temperatur. Wpływ każdego z nich z osobna jest zależny od zamysłu konstruktorskiego oraz warunków otoczenia, w jakim pracuje urządzenie.

3. Dwufazowe metody chłodzenia

Jednym z możliwych rozwiązań stosowanych w pasywnych systemach odprowadzania ciepła są metody oparte o zjawisko przemiany fazowej czynnika chłodzącego. Przemiana ta następuje najczęściej pomiędzy stanem ciekłym, a gazowym. W jej trakcie występuje odebranie lub oddanie energii w postaci ciepła, z lub do otoczenia, co związane jest ze zmianą stanu skupienia substancji. Na potrzeby chłodzenia elektroniki, zjawisko to wykorzystywane jest najczęściej w rurkach cieplnych lub komorach parowych (tzw. płaskich rurkach cieplnych). Jak pokazano schematycznie na przykładzie budowy komory parowej na Rys. 1, w konstrukcji tych urządzeń wyróżnić można strefy pełniące odrębne funkcje. W obszarze parownika, ze względu na podwyższoną temperaturę, czynnik roboczy ulega odparowaniu, zmieniając stan skupienia z ciekłego na gazowy. W trakcie tego procesu pobierana jest ze ścianek urządzenia niezbędna energia, w ilości odpowiadającej ciepłu parowania wykorzystanej substancji. Następnie, w postaci gazowej, czynnik przemieszcza się wewnątrz komory do skraplacza. Ze względu na niższą temperaturę tego obszaru, następuje w nim odwrotna do poprzedniej przemiana fazowa, powodując powrót czynnika roboczego do stanu ciekłego. W trakcie tego procesu następuje wyzwolenie do otoczenia energii termicznej, w ilości równej ciepłu parowania. Następnie, skroplony czynnik roboczy przemieszcza się do parownika, gdzie opisany cykl rozpoczyna się na nowo. Transport skroplonego czynnika możliwy jest dzięki wykorzystaniu specjalnie dobranych struktur (porowatych lub kształtowych), wzdłuż których ciecz przemieszcza się w efekcie działania sił kapilarnych.



Rys. 1. Schemat działania dwufazowego systemu chłodzenia na podstawie komory parowej.

W urządzeniach elektroniki użytkowej, najczęściej stosowanym czynnikiem roboczym jest woda demineralizowana. Rzadziej stosowane są inne substancje, jak metanol lub aceton. Do specyficznych zastosowań, stosuje się również inne czynniki robocze, których wybór podyktowany jest m.in. możliwą temperaturą pracy. Ciepło przemiany fazowej popularnych czynników roboczych mierzone w temperaturze ich wrzenia $h_e\ [\frac{\mathrm{kJ}}{\mathrm{kg}}]$ zestawiono w Tabeli 1.

Tab. 1. Ciepło parowania popularnych czynników roboczych dwufazowych systemach chłodzenia [2].

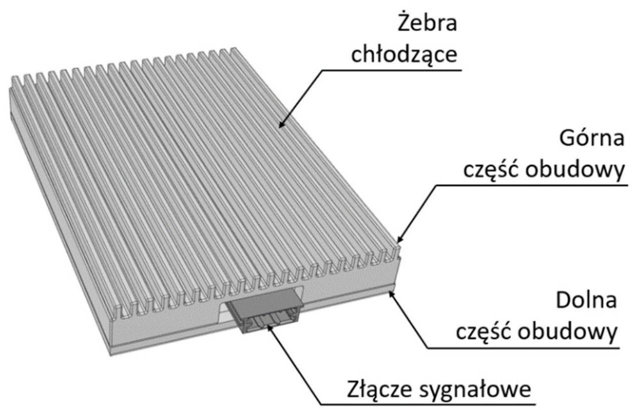
Czynnik roboczy	Ciepło parowania $h_e\ [\frac{\mathrm{kJ}}{\mathrm{kg}}]$
Woda	2257
Metanol	1100
Aceton	518

Rezultatem zastosowania opisanych urządzeń jest transport energii termicznej z obszaru parownika, do obszaru skraplacza, a więc od źródła ciepła do wybranego rejonu, gdzie ulega ona rozproszeniu do otoczenia.

W dalszej części artykułu opisano koncepcję zastosowania komory parowej na potrzeby obniżenia wartości temperatury pracy urządzenia elektronicznego, zamkniętego w obudowie aluminiowej. Porównano wynik działania tak zaprojektowanego pasywnego systemu chłodzenia z działaniem klasycznego rozwiązania, opartego o przewodzenie ciepła, od jego źródła, bezpośrednio do obudowy wyposażonej w użebrowanie chłodzące. Na potrzeby oszacowania wydajności obu wariantów konstrukcyjnych przeprowadzono analizy numeryczne wykorzystując metodę obliczeniowej mechaniki płynów - CFD (*ang*. Computational Fluid Dynamics).

4. Koncepcja zastosowania komory parowej

Rysunek 2 przedstawia analizowane urządzenie zamknięte w dwuczęściowej obudowie wykonanej ze stopu aluminium EN AC-44300 ($k_{Al} = 130\frac{W}{m \cdot K}$). W celu zwiększenia powierzchni wymiany ciepła z otoczeniem, górna część obudowy pokryta została żebrami chłodzącymi. Ponadto, na rysunku widoczne jest złącze sygnałowe, za pośrednictwem którego wewnętrzny układ elektroniczny jest zasilany oraz komunikuje się z urządzeniami peryferyjnymi.

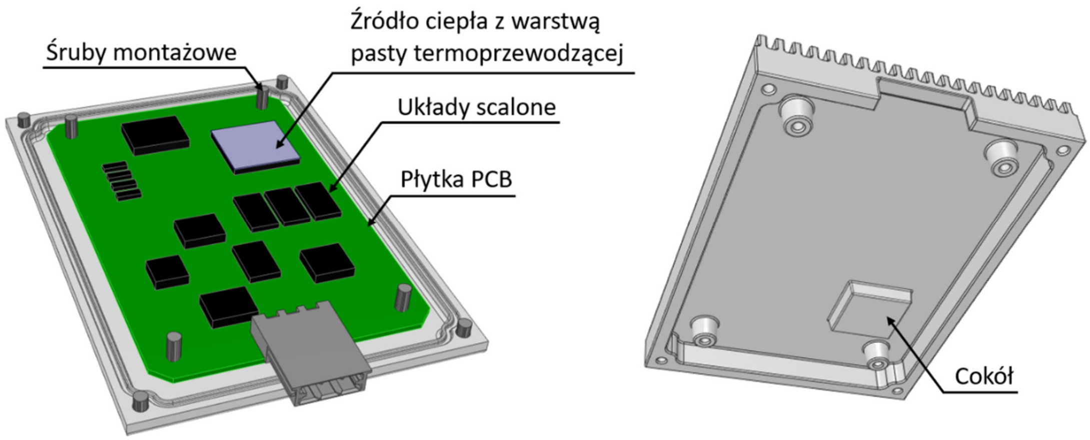


Rys. 2. Analizowane urządzenie elektroniczne zamknięte w aluminiowej obudowie.

Rysunek 3 przedstawia wnętrze urządzenia, na które składają się: obwód drukowany PCB wraz z komponentami elektronicznymi oraz śruby montażowe. Jeden z zamontowanych wewnątrz urządzenia układów scalonych (Komponent I) wydziela podczas pracy moc $P = 18$W, podczas gdy całkowita moc generowana przez urządzenie to 20.84W (Rys. 4). Aby umożliwić przepływ energii termicznej od tego układu scalonego na zewnątrz urządzenia, w pierwszej iteracji projektowej połączono jego górną powierzchnię z obudową. Dla zmniejszenia powierzchniowej rezystancji termicznej tego połączenia, pomiędzy układem scalonym, a cokołem obudowy zastosowano warstwę pasty termoprzewodzącej. Tak przygotowany model został w dalszej kolejności poddany analizie numerycznej, na potrzeby której założono, że temperatura otoczenia w jakiej pracować będzie opisywany sterownik wynosić będzie $t_{amb} = 55°C$. Największa dopuszczalna temperatura złącza dla Komponentu I wynosi $t_j^{max} = 125°C$, do której określenia wymagana jest znajomość wartości temperatury na górnej, zewnętrznej powierzchni jego obudowy t_{pow}. Wykorzystując tę informację, rzeczywistą temperaturę złącza t_j oszacować można stosując wzór (4) [2]:

$$t_j = t_{pow} + \left(\mathrm{R}_{\theta jc} \cdot P\right) \tag{4}$$

gdzie poprzez $R_{\theta jc}$ oznaczono współczynnik rezystancji termicznej pomiędzy obszarem złącza wewnątrz układu scalonego, a górną powierzchnią jego obudowy. Wielkość $R_{\theta jc}$ najczęściej podana jest przez producenta, a dla analizowanego komponentu wynosi $R_{\theta jc} = 1.22\frac{°C}{W}$.

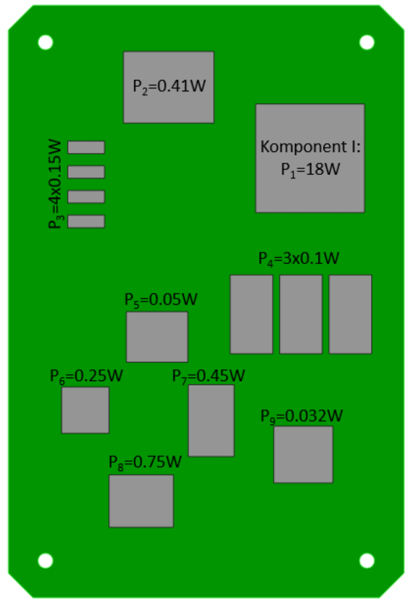


Rys. 3. Wnętrze analizowanego urządzenia, wyposażonego w pojedynczy cokół na obudowie.

4.1. Wyniki analiz numerycznych

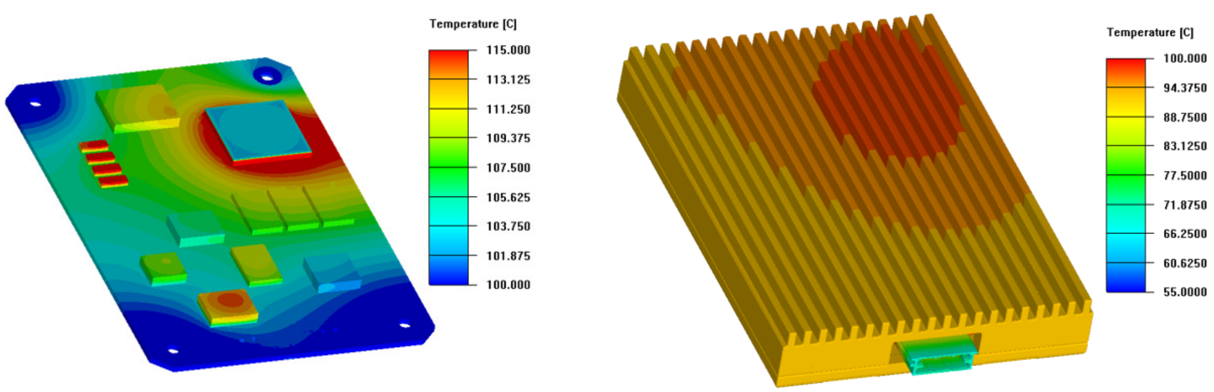
Opisane w niniejszym artykule symulacje numeryczne prowadzone były z wykorzystaniem metody CFD, pozwalającej na analizę zmienności przebiegu parametrów obliczeniowych w trójwymiarowej przestrzeni modelu. W celu opisania ruchu płynu (powietrza) w obszarze dziedziny obliczeniowej (wewnątrz i na zewnątrz urządzenia), zastosowana metoda dostarczyła rozwiązania równań Naviera – Stokesa. W obliczeniach uwzględniono również przestrzenny transport energii termicznej, realizowany za pośrednictwem jednocześnie występujących zjawisk promieniowania, konwekcji i przewodzenia. Przyjęto, że wymiana ciepła z otoczeniem zachodzi w procesie konwekcji naturalnej, którą odwzorowano stosując model Boussinesq. W odniesieniu do przepływu powietrza przyjęto, że w całej dziedzinie obliczeniowej będzie on laminarny. Proces promieniowania cieplnego uwzględniono w symulacjach stosując model DO (*ang*. Discrete Ordinates), zakładający dyskretną liczbę kierunków promieniowania z każdego źródła – zastosowano pięciokrotny podział każdego z oktantów zarówno w kierunkach azymutalnym jak i zenitalnym. Problem obliczeniowy zdefiniowany został jako zagadnienie rozwiązywane w stanie ustalonym. Modele przygotowane na potrzeby opisanych analiz składały się z około 4 milionów elementów. Warunkiem zatrzymania każdej z opisanych symulacji było osiągnięcie odpowiednio małych wartości residuów dla rozwiązywanych równań zachowania masy, prędkości oraz transportu energii (kolejno 10^{-4}, 10^{-4} i 10^{-8}) oraz przy pomijalnie małych zmianach wartości temperatury obserwowanej na powierzchni źródła ciepła w kolejnych iteracjach obliczeń.

Rysunek 5 prezentuje rezultat symulacji numerycznej, przedstawiający pole temperatury na powierzchni obwodu drukowanego oraz na pozostałych komponentach elektronicznych. Jak odczytano z wyników analizy, wartość temperatury górnej powierzchni Komponentu I sięga około $103.7°C$.



Rys. 4. Moc komponentów analizowanego układu elektronicznego.

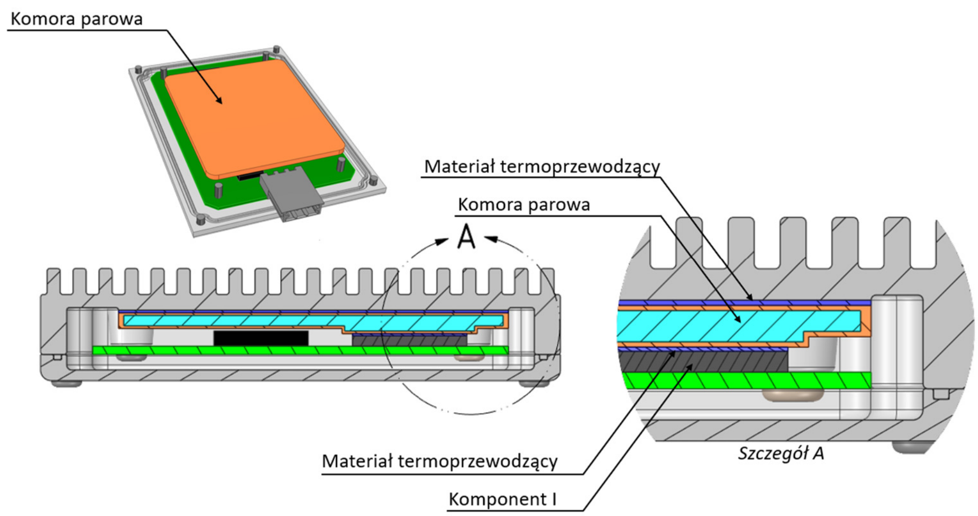
Na podstawie wzoru (4), temperatura złącza analizowanego układu scalonego wynosi w tym wypadku: $t_j = 103.7 + (1.22 \cdot 18) = 125.66°C$, nieznacznie wykraczając poza dopuszczalny limit $125°C$. Niemniej jednak, możliwe zmiany warunków pracy urządzenia (np. zabrudzenie powierzchni obudowy lub zwiększonego zewnętrznego promieniowania termicznego pochodzącego od innych urządzeń znajdujących się w pobliżu), powodować mogą dalsze przekroczenie tej wartości, powodując potencjalne problemy związane z eksploatacją rozważanego komponentu oraz wydatnie skrócić jego żywotność.



(A) *(B)*

Rys. 5. Wyniki analizy termicznej sterownika bez komory parowej: rozkład pola temperatury dla obwodu drukowanego i układów scalonych (A) oraz na obudowie urządzenia (B).

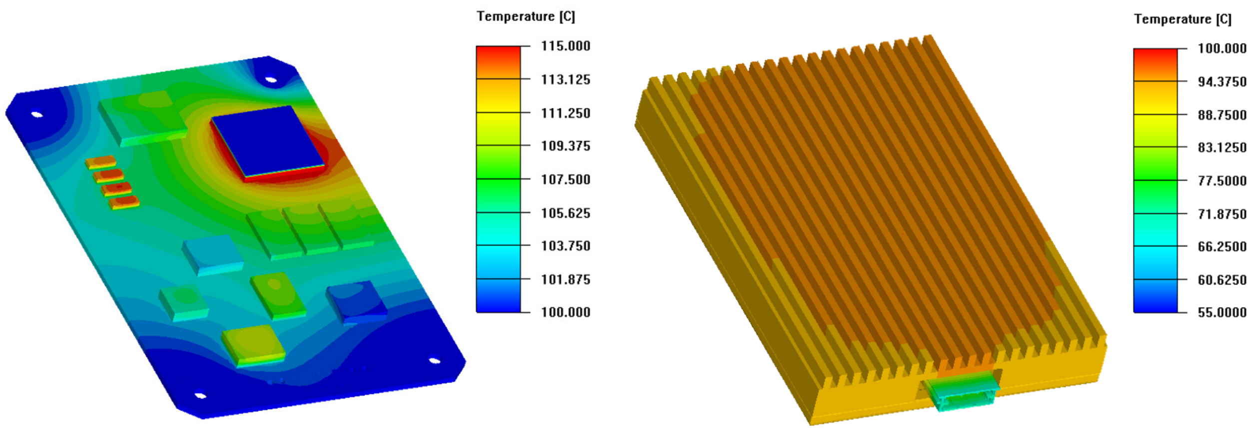
W celu poprawy warunków użytkowania rozważanego urządzenia, w kolejnej iteracji projektowej podjęto próbę zintegrowania go z komorą parową, której obecność miała na celu zwiększenie efektywności odprowadzania ciepła od Komponentu I, do otoczenia. Założono, że czynnikiem roboczym w komorze będzie woda demineralizowana, jej obudowa zaś wykonana będzie z miedzi charakteryzującej się współczynnikiem przewodności cieplnej $k_{Cu} = 385 \frac{\mathrm{W}}{\mathrm{m \cdot K}}$. Wymiary zewnętrzne tego komponentu to 75x90x4 mm. Jak pokazano na Rys. 6, komora została umieszczona w obudowie sterownika w taki sposób, aby pozostawać w kontakcie jednocześnie z górną powierzchnią analizowanego układu scalonego oraz obudową sterownika. Aby zmniejszyć rezystancję termiczną występującą na styku tych komponentów, z obu stron komory parowej zastosowano warstwę materiału termoprzewodzącego ($k = 3.5 \frac{W}{m \cdot K}$) o grubości warstwy równej 0.5mm.



Rys. 6. Komora parowa zintegrowana z obudową analizowanego sterownika.

W celu efektywnego wykorzystania działania komory parowej zmodyfikowano kształt cokołu wewnętrznego obudowy sterownika w taki sposób, aby przylegał on do komory parowej na całej jej górnej powierzchni. Tak przygotowany model geometryczny poddany został analizie numerycznej, na potrzeby której założono, że urządzenia to pracować będzie w warunkach procesu ustalonego. Ponadto, założono, że opisać je można stosując izotropowy model materiałowy. Takie podejście powszechnie stosowane jest w celu uproszczenia opisu zjawisk występujących wewnątrz komory (tj. parowania, transportu fazy gazowej czynnika roboczego, skraplania oraz przemieszczania się fazy ciekłej wywołanego odziaływaniem kapilarnym ze strukturą porowatą). Autorzy prac [7] oraz [13] przedstawili wyniki badań związanych z uproszczeniem modelowania zachowania komór parowych, dochodząc do zbieżnych wniosków mówiących o możliwości opisu tych urządzeń jako homogenicznych bloków materiału o wysokiej przewodności termicznej, której wartość sięga $8300 - 10000 \frac{W}{m \cdot K}$. W pracy [13] opisano zależność pomiędzy zastępczą przewodnością termiczną uproszczonego modelu komory parowej, a jej temperaturą pracy. Zgodnie z przytoczonymi przez autorów obserwacjami, wraz ze wzrostem temperatury źródła ciepła, wzrasta również wartość szacowanej przewodności zastępczej urządzenia. Jest to podyktowane intensywnością odparowania czynnika roboczego, zmianą rozkładu i wartości pola ciśnienia panującego w komorze parowej oraz, w konsekwencji, zmianą sposobu rozprzestrzeniania się pary czynnika roboczego we jej wnętrzu. Z przedstawionych w [13] wyników eksperymentalnych wynika, że dla komory parowej pracującej w temperaturze o $50°C$ wyższej niż temperatura otoczenia, przewodność zastępcza urządzenia osiąga wartość około $9750 \frac{W}{m \cdot K}$.

Na potrzeby opisanej poniżej symulacji przyjęto, że modelowana komora posiadać będzie izotropową charakterystykę przewodności termicznej, opisaną wartością zastępczą równą $k_{kp} = 9000 \frac{W}{m \cdot K}$. Rysunek 7 prezentuje wyniki symulacji termicznej tak przygotowanego modelu.



(A) (B)

Rys. 7. Wyniki analizy termicznej sterownika z komora parową: (A) rozkład pola temperatury dla obwodu drukowanego i układów scalonych oraz (B) na obudowie urządzenia.

Wyniki zaprezentowane na Rys. 7, pokazują, że temperatura komponentów elektronicznych, obwodu drukowanego oraz obudowy sterownika uległa zmniejszeniu. Szczególnie widoczna jest zmiana temperatury powierzchni Komponentu I, który jest głównym źródłem ciepła w analizowanym sterowniku, a który jak opisano powyżej, jako jedyny pozostawał w kontakcie z komorą. Korzystając ze wzoru (4) oszacować można wpływ zastosowania komory parowej na temperaturę złącza t_j^{kp} rozważanego komponentu. Najwyższa obliczona temperatura na górnej powierzchni Komponentu I wynosi w analizowanym przypadku $t_{max}^{kp} = 99.95°C$, co przekłada się na $t_j^{kp} = 121.91°C$, a więc $\Delta t_j = 125.66°C - 121.91°C = 3.75°C$. Uzyskane drogą symulacyjną wyniki zestawiono w Tab. 2.

Rysunek 7 przedstawia również rozkład temperatury na zewnętrznej powierzchni obudowy urządzenia. Jak zauważyć można, zastosowanie komory parowej w opisany sposób prowadzi do rozproszenia energii termicznej na większym obszarze, a w konsekwencji do zmniejszenia gęstości strumienia ciepła przepływającego przez aluminiową obudowę.

Tab. 2. Zestawienie wyników analiz termicznych omawianego urządzenia.

	Temperatura na powierzchni Komponentu I	***Temperatura złącza Komponentu I***
Wariant I: Obudowa aluminiowa wyposażona w cokół	$103.7°C$	$125.66°C$
Wariant II: Obudowa aluminiowa wyposażona w komorę parową	$99.95°C$	$121.91°C$

5. Żywotność Komponentu I

Elementy elektroniczne zbudowane na bazie krzemu, pracując w warunkach podwyższonej temperatury, podlegają szeregowi zjawisk fizyko-chemicznych, powodujących ich przyspieszone starzenie. Do najważniejszych czynników degradacyjnych zaliczyć należy efekt elektromigracji występujący w materiale półprzewodnikowym, polegającym na stopniowym przemieszczaniu się atomów degradującego materiału w warunkach przepływu prądu. Intensywność tego niepożądanego efektu zwiększa się wraz ze wzrostem temperatury materiału półprzewodnikowego.

Często stosowanym sposobem szacowania tempa zwiększonego zużycia komponentu elektronicznego pracującego w warunkach podwyższonej temperatury jest zastosowanie wzoru Arrheniusa, pozwalającego wyznaczyć szybkość przebiegu procesu degradacyjnego. Jednak, jak opisano w [21], w sytuacji gdy krzemowy element półprzewodnikowy pracuje w temperaturze powyżej $105°C$, podejście to nie zapewnia dokładnego opisu przebiegu procesu starzenia. W zamian, wykorzystać można zaproponowaną w [21] procedurę analizy procesu degradacji układów scalonych, opartą o dane empiryczne. Procedura ta pozwala na szacowanie wartości współczynnika przyspieszonego zużycia AF, na podstawie którego, stosując wzór (5), przewidywać można czas poprawnego działania analizowanego komponentu:

$$L_U = L_D \cdot AF \tag{5}$$

gdzie: L_D – nominalna długość czasu pracy komponentu półprzewodnikowego szacowana dla $t_j = 105°C$, L_U – przewidywana długość czasu pracy w zadanych warunkach.

Według przytoczonej procedury, w zakresie temperatur pracy $110°C - 125°C$, zależność pomiędzy wartością współczynnika AF i temperaturą t, opisać można wzorem (6):

$$AF = -0.02t + 2.7 \tag{6}$$

gdzie wartość temperatury t podana jest w stopniach Celsjusza.

Dla omawianego w niniejszym artykule przypadku, współczynnik AF dla Komponentu I, przyjmuje wartości: $AF = 0.19$ oraz $AF^{kp} = 0.26$, kolejno dla wariantu obudowy wyposażonej w cokół oraz komorę parową. Przy założeniu profilu obciążenia Komponentu I jako ciągłej jego pracy w obliczonych temperaturach $t_j = 125.66°C$ oraz $t_j^{kp} = 121.91°C$, na podstawie zależności (5) obliczyć można, że zastosowanie komory parowej pozwoli na dłuższe jego użytkowanie, przy czym uzysk w tym wypadku wynosić będzie $0.07L_D$, a więc przewidywany czas pracy urządzenia będzie o 7% dłuższy.

6. Podsumowanie

W niniejszym artykule zaprezentowano koncepcję integracji obudowy przykładowego urządzenia elektronicznego pracującego w warunkach podwyższonej temperatury otoczenia z komorą parową. Efekt zastosowania takiej modyfikacji układu odprowadzania ciepła porównano z działaniem klasycznej obudowy wykonanej ze stopu aluminium. Otrzymane na drodze analiz numerycznych wyniki dowodzą, że w porównaniu z klasycznym rozwiązaniem, zastosowanie układu chłodzącego wykorzystującego komorę parową pozwala na obniżenie temperatury złącza krytycznego podzespołu o $\Delta t_j = 3.75°C$. W założonych warunkach otoczenia, zmiana ta skutkuje obniżeniem temperatury pracy materiału półprzewodnikowego poniżej krytycznej wartości $t_j^{max} = 125°C$, wydłużając tym samym przewidywany czas jego użytkowania o $0.07L_D$.

Literatura

1. Aranzabal I, de Alegría I M, Delmonte N, Cova P, Kortabarria I. Comparison of the Heat Transfer Capabilities of Conventional Single-and Two-Phase Cooling Systems for an Electric Vehicle IGBT Power Module. IEEE Transactions on Power Electronics 2018; 34(5): 4185-4194.
2. Cengel Y A. Heat transfer: a practical approach, 2nd Edition. McGraw-Hill Higher Education, 2002.
3. Chainer T J, Schultz M D, Parida P R, Gaynes M A. Improving data center energy efficiency with advanced thermal management. IEEE Transactions on Components, Packaging and Manufacturing Technology 2017; 7(8): 1228-1239.
4. Chang N, Pan S, Srinivasan K, Feng Z, Xia W, Pawlak T, Geb D. Emerging ADAS Thermal Reliability Needs and Solutions. IEEE Micro 2018; 38(1): 66-81.
5. Gadalla B, Schaltz E, Blaabjerg F. A survey on the reliability of power electronics in electro-mobility applications. Proceedings of: 2015 Intl Aegean Conference on Electrical Machines & Power Electronics (ACEMP), 2015 Intl Conference on Optimization of Electrical & Electronic Equipment (OPTIM) & 2015 Intl Symposium on Advanced Electromechanical Motion Systems (ELECTROMOTION) 2015; 304-310.
6. Gu J, Pecht M. Predicting the reliability of electronic products. Proceedings of: 8th International Conference on Electronic Packaging Technology 2007; 1-8.

7. Gurevich A, Steiner I, Huang E. Design of thermal systems based on combination of Thermoelectric and Vapor Chamber technologies. Proceedings of: 33rd Thermal Measurement, Modeling & Management Symposium (SEMI-THERM) 2017; 1-5.
8. Hager M, Gromala P, Wunderle B, Rzepka S. Affordable and Safe High Performance Vehicle Computers with Ultra-Fast On-Board Ethernet for Automated Driving. Advanced Microsystems for Automotive Applications 2018: Smart Systems for Clean, Safe and Shared Vehicles. Springer, 2019.
9. Hammoud J, Dudley S, Apte N. Use of Heat Pipe Technology for Multi Media Thermal Management. Proceedings of: SAE 2006 World Congress & Exhibition 2006; 1-3.
10. Khazaka R, Mendizabal L, Henry D, Hanna R. Survey of high-temperature reliability of power electronics packaging components. IEEE Transactions on power Electronics 2014; 30(5): 2456-2464.
11. Ma X, Guo Y, Wang L, Ji W. Exploration of the reliability of automotive electronic power steering system using device junction electrothermal profile cycle. IEEE Access 2016; 4: 7054-7062.
12. Mochizuki M. Latest development and application of heat pipes for electronics and automotive. Proceedings of: IEEE CPMT Symposium Japan 2017; 87-90.
13. Niittymaki L, Biber C, Carbone M. Flexible CFD simulation model of a thin vapor chamber for mobile applications. Proceedings of: 22nd International Workshop on Thermal Investigations of ICs and Systems (THERMINIC) 2016; 152-157.
14. Qi F, Wang Y, Bob-Manuel C, Li H, Jones S, Li B, Chen Y, Yan Y. Advanced Cooling Solutions of High Power Automotive Module. Proceedings of: PCIM Asia 2017: International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management 2017; 1-3.
15. Ram Sundar R, Sarweshwara Sarma N S V. Performance study of motorcycle driving-beam LED headlight with different heat sink models and LED pitch. Proceedings of: 16th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm) 2017; 784-792.
16. Riches S, Johnston C. Electronics design, assembly and reliability for high temperature applications. Proceedings of: IEEE International Symposium on Circuits and Systems (ISCAS) 2015; 1158-1161.
17. Robinson A J, Colenbrander J, Kempers R, Chen R. Solid and Vapor Chamber Integrated Heat Spreaders: Which to Choose and Why. IEEE Transactions on Components, Packaging and Manufacturing Technology 2018; 8(9): 1581-1592.
18. Simon D, Boianceanu C, De Mey G, Ţopa V. Experimental reliability improvement of power devices operated under fast thermal cycling. IEEE Electron Device Letters 2015; 36(7): 696-698.
19. Sun L, Gu X H, Song P, Di Y. A generalized equivalent temperature model in a time-varying environment. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2017; 19(3): 432–440.

20. Tan C M, Narula U, Kapoor D. Reliability paradox for worldwide automotive electronics. Proceedings of: IEEE Annual Reliability and Maintainability Symposium (RAMS) 2017; 1-7.
21. Webber A. Calculating Useful Lifetimes of Embedded Processors: Application Report. Texas Instruments Incorporated, 2017.
22. Zeng S, Sun B, Tong C. A modified model of electronic device reliability prediction. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2009; 4(44): 4-9.