Vol. 16. No 3, 2014
ISSN 1507-2711
Cena: 25 zł (w tym 5% VAT)

EKSPLOATACJA I NIEZAWODNOŚĆ
MAINTENANCE AND RELIABILITY

Polskie Naukowo Techniczne Towarzystwo Eksploatacyjne
Warszawa

Polish Maintenance Society
Warsaw

Scientific Board

Professor Andrzej Niewczas, PhD, DSc (Eng)

Chair of Scientific Board
President of the Board of the Polish Maintenance Society

Professor Holm Altenbach, PhD, DSc (Eng)
Martin Luther Universität, Halle-Wittenberg, Germany

Professor Dr Gintautas Bureika
Vilnius Gediminas Technical University, Vilnius, Lithuania

Professor Zdzisław Chłopek, PhD, DSc (Eng)
Warsaw University of Technology, Warsaw

Professor Jan Dąbrowski, PhD, DSc (Eng)
Białystok Technical University, Białystok

Professor Sławczo Denczew, PhD, DSc (Eng)
The Main School of Fire Service, Warsaw

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

Professor Olgierd Hryniewicz, PhD, DSc (Eng)
Systems Research Institute of the Polish Academy of Science, Warsaw

Professor Hong-Zhong Huang, PhD, DSc
University of Electronic Science and Technology of China, Chengdu, Sichuan, China

Professor Krzysztof Kołowrocki, PhD, DSc
Gdynia Maritime University

Professor Štefan Liščak
Žilinská univerzita, Žilina, Slovak Republic

Professor Vaclav Legat
Czech University of Agriculture, Prague, Czech Republic

Professor Jerzy Merkisz, PhD, DSc (Eng)
Poznań University of Technology, Poznań

Professor Gilbert De Mey
University of Ghent, Belgium

Professor Tomasz Nowakowski, PhD, DSc (Eng)
Wrocław University of Technology, Wrocław

Professor Marek Orkisz, PhD, DSc (Eng)
Rzeszów University of Technology, Rzeszów

Professor Stanisław Piasecki, PhD, DSc (Eng)
Systems Research Institute of the Polish Academy of Sciences, Warsaw

Professor Leszek Piaseczny, PhD, DSc (Eng)
Polish Naval Academy, Gdynia

Professor Stanisław Radkowski, PhD, DSc (Eng)
Warsaw University of Technology, Warsaw

Professor Andrzej Seweryn, PhD, DSc (Eng)
Białystok Technical University, Białystok

Professor Zbigniew Smalko, PhD, DSc (Eng)
Air Force Institute of Technology, Warsaw

Professor Marian Szczerek, PhD, DSc (Eng)
Institute for Sustainable Technologies, Radom

Professor Jan Szybka, PhD, DSc (Eng)
AGH University of Science and Technology, Cracow

Professor Katsumi Tanaka
Kyoto University, Kyoto, Japan

Professor David Vališ, PhD, DSc (Eng)
University of Defence, Brno, Czech Republic

Professor Irina Yatskiv
Riga Transport and Telecommunication Institute, Latvia

Co-financed by the Minister of Science and Higher Education

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.

All the scientific articles have received two positive reviews from independent reviewers.

Our IF is 0.293

Editorial staff:	Dariusz Mazurkiewicz, PhD, DSc (Eng), Associate Professor (Editor-in-Chief, Secretary of the Scientific Board) Tomasz Klepka, PhD, DSc (Eng) (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 20-618 Lublin, ul. Nadbystrzycka 36, Poland e-mail: office@ein.org.pl http://www.ein.org.pl/
Circulation:	550 copies

Science and Technology

IN MEMORIAM

COMMODORE PROFESSOR LESZEK PIASECZNY

Cdre Professor Leszek Piaseczny, PhD, D.Sc. (Eng). was born on 1 January 1947, in Puławy. In 1968, he graduated from the Higher School of Naval Forces in Gdynia with the title of mechanical engineer and a specialization in "Machinery and engine rooms" and the military rank of acting sub-lieutenant. In 1976, he was awarded the degree of doctor of technical sciences, conferred by a resolution of the Council of the Ship Research Institute of the Gdańsk University of Technology.

In 1995, the Council of the Faculty of Machines and Motor Vehicles of the Poznań University of Technology conferred upon Dr Leszek Piaseczny the degree of doctor of technical science (habilitatus).

In 1996, Dr Leszek Piaseczny, D.Sc. Eng. was appointed to the position of Associate Professor of the Naval Academy. In the years 1993 – 2003, Professor Leszek Piaseczny held the position of Head of the Scientific Department of the Naval Academy, and then in the period 2003 – 2006 he was the Commander-Dean of the Faculty of Mechanical and Electrical Engineering of the Naval Academy. In 2003, Polish President conferred on Leszek Piaseczny the title of Professor of technical sciences.

Professor Leszek Piaseczny's professional interests involved the operation and design of technical facilities with particular emphasis on ships and internal combustion engines. Professor Leszek Piaseczny concentrated his research on the durability and reliability of the technical systems of vessels and the optimization of their maintenance systems, including the technology of the repair of internal combustion engines of ships and the use of polymer composites in the processes of repairing marine equipment. Of particular interest is the pioneering work of Professor Piaseczny, both theoretical and practical, in the field of internal combustion engines. Professor Piaseczny's research on the emissions of pollutants from internal combustion engines of vessels has been applied in practice.

Overall, Professor Piaseczny's scientific contribution includes the authorship of three monographs, two patents, and more than 200 scientific articles.

Professor Leszek Piaseczny promoted four doctors and was involved as a reviewer in numerous doctoral and habilitation proceedings as well as applications for the title of professor.

Professor Leszek Piaseczny was a founding member of the Polish Scientific Society of Combustion Engines and member of the Polish Maintenance Society. Professor Piaseczny was member of the Scientific Board of the quarterly Eksploatacja i Niezawodność – Maintenance and Reliability.

He educated many generations of engineers and masters of science and was a respected lecturer and an author of original lectures and graduate seminars.

He combined his undisputed professional position with a rich personality of a man of letters and an erudite in many areas of culture. On account of his activity, knowledge and scientific expertise, as well as extremely positive features of his personality, Professor Leszek Piaseczny was a widely respected and liked colleague.

He died on April 4, 2014 after a long illness

Prof. Zdzisław Chłopek

RYBAK P. **Operating loads of impulse nature acting on the special equipment of the combat vehicles**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 347–353.

Providing the combat vehicles with high operation effectiveness, safety and reliability during execution of complex tasks makes a priority. Therefore the armament and the military equipment have to meet very high requirements in that aspect when used in various conditions. This paper presents basic sources of dynamic loads affecting the combat vehicles. Attention is paid to the loads of impact nature as they mostly affect the effectiveness and reliability of a vehicle, electronic equipment and psychophysical condition of the combat vehicle crew. These loads result from off-road drives, firing the gun, the influence of the land mines or IED, hitting by enemy's missile. As a result, some fragments of the experimental and model tests on combat vehicles are presented. Results of these tests can be helpful when designing internal vehicle equipment including special equipment. Particularly in the aspect of normative requirements for that class of vehicles and their special equipment.

YANG Y-J, HUANG H-Z, 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.

The electrohydraulic servo valve (EHSV) is widely used in many engineering fields. Its reliability is of great importance to the reliability and safety of entire servo control systems. With the aim of analyzing and evaluating reliability of EHSV, this paper firstly presents the physical structure and functional principle of EHSV. It is followed by the Failure Mode, Effects and Criticality Analysis (FMECA). From the analysis, the common cause failures (CCF) in the studied EHSV are identified. Lastly, a method that can quantitatively analyze reliability and failure rate of EHSV with considering the common cause failures is proposed. It is observed from the study that the failure rate of the EHSV with CCF is lower than the failure rate without considering CCF.

PROCHOWSKI L, ŻUCHOWSKI A. **Analysis of the influence of passenger position in a car on a risk of injuries during a car accident**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 360–366.

Experimental tests have been prepared and carried out in order to deepen the analysis of the influence of a position of a rear seat passenger in a passenger car on a risk of injuries during a road accident. This risk is considered and calculated in the aspect of a car passenger position that slightly deviates from the one planned by manufacturers of individual protection devices. Attention was focused on the analysis of measurable effects of the position modification in the area of torso and head movement and neck deformation. The experimental tests included a physical simulation of a frontal collision of a car and a rigid obstacle. They showed that relatively small changes to the initial dummy position could be a reason of significant differences in the movement trajectory and dummy position at the final stage of the crash test. These differences are described in details. The analysis also included the criteria index values that refer to the head, neck and chest injuries. Index calculation results confirm the influence of a small change to the leg and torso position on the index values, thus on the probability of injuries of the rear seat passengers in a passenger car.

ZBROWSKI A, SAMBORSKI T. **Study on electromechanical drives used in ventilation and smoke extraction systems**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 367–376.

The paper presents experimental tests for electromechanical drives for use in ventilation and smoke extraction systems of buildings. The determined characteristics (force, shift, current and voltage as a function of time) enable the assessment of applicability of these drives as the drives of flaps and windows in fire protection systems. The study was conducted using a developed test stand facilitating the determination of characteristics of work of linear and rotary drives in changeable temperatures.

MAZURKIEWICZ D. **Computer-aided maintenance and reliability management systems for conveyor belts**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 377–382.

Operational reliability of conveyor transport systems is a problem that relates to the provision of an adequate level of availability of a continuous transport system, which in the case of belt conveyors depends not only on their usability, as determined by their design and manufacture, but also on the appropriate level of their use, understood, among others, as the degree of sophistication and effectiveness of the methods and tools applied in industrial diagnosis of those devices. The present article describes the most popular diagnostic systems used in the maintenance of internal transport conveyor systems. Also a new method of computer-aided maintenance of such systems is presented.

RYBAK P. **Obciążenia eksploatacyjne o charakterze udarowym działające na wyposażenie specjalne wozów bojowych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 347–353.

Zapewnienie wozom bojowym wysokiej skuteczności działania, bezpieczeństwa oraz niezawodności podczas realizacji złożonych zadań jest traktowane priorytetowo. A zatem uzbrojenie i sprzęt wojskowy musi spełniać bardzo wysokie wymagania w tym aspekcie podczas eksploatacji w różnych warunkach. W pracy przedstawiono podstawowe źródła obciążeń dynamicznych działające na wozy bojowe. Uwagę skupiono na obciążeniach mających charakter udarowy, gdyż one głównie wpływają na sprawność i niezawodność pojazdu, urządzeń wewnętrznych i stan psychofizyczny załogi. Obciążenia te wynikają z jazd terenowych, strzelania z armaty, oddziaływania miny lub IED, trafienia pociskiem przeciwnika. W rezultacie przedstawiono niektóre fragmenty z badań eksperymentalnych i modelowych wozów bojowych. Wyniki tych badań mogą być pomocne przy projektowaniu urządzeń wewnętrznych pojazdu w tym urządzeń specjalnych. Szczególnie w aspekcie wymagań normatywnych dla tej klasy pojazdów oraz ich urządzeń specjalnych.

YANG Y-J, HUANG H-Z, LIU Y, ZHU S-P, PENG W. **Analiza niezawodności serwozaworu elektrohydraulicznego narażonego na uszkodzenia spowodowane wspólną przyczyną**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 354–359.

Serwozawory elektrohydrauliczne (EHSV) mają szerokie zastosowanie w wielu dziedzinach inżynierii. Ich niezawodność ma decydujące znaczenie dla niezawodności i bezpieczeństwa całych układów sterowania serwomechanizmami. W celu analizy i oceny niezawodności zaworów EHSV, w pracy przedstawiono najpierw ich budowę fizyczną i zasadę działania. Następnie przeprowadzono analizę przyczyn, skutków i krytyczności uszkodzeń (FMECA). Na podstawie tej analizy określono uszkodzenia zaworu EHSV spowodowane wspólną przyczyną (CCF). Wreszcie, zaproponowano metodę, za pomocą której można ilościowo analizować niezawodność i awaryjność EHSV z uwzględnieniem uszkodzeń spowodowanych wspólną przyczyną. Badania wykazały, że awaryjność EHSV przy uwzględnieniu CCF jest niższa niż w wypadku nieuwzględnienia CCF.

PROCHOWSKI L, ŻUCHOWSKI A. **Analiza wpływu pozycji człowieka w samochodzie na ryzyko obrażeń w czasie wypadku drogowego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 360–366.

Przygotowano i przeprowadzono badania eksperymentalne, które mają na celu pogłębienie analizy wpływu pozycji osoby siedzącej na tylnym siedzeniu w samochodzie osobowym na ryzyko jej obrażeń w czasie wypadku drogowego. Ryzyko to jest rozważane i obliczone w aspekcie zajmowania przez człowieka w samochodzie pozycji nieznacznie odbiegającej od zaplanowanej przez producentów urządzeń ochrony indywidualnej. Uwagę skupiono na analizie wymiernych skutków zmiany tej pozycji w obszarze ruchu torsu i głowy oraz odkształcenia szyi. Badania eksperymentalne były symulacją fizyczną czołowego uderzenia samochodu w sztywną przeszkodę. Pokazały one, że stosunkowo niewielkie różnice w początkowej pozycji manekina mogą być przyczyną znacznych różnic w trajektorii ruchu oraz położeniu manekina w kulminacyjnej fazie testu zderzeniowego. Różnice te zostały szczegółowo opisane. Analizie poddano także wartości wskaźników kryterialnych, które dotyczą powstawania obrażeń głowy, szyi i klatki piersiowej. Wyniki obliczeń wskaźników potwierdzają wpływ już niewielkiej zmiany położenia nóg i torsu na ich wartości, a zatem na prawdopodobieństwo powstawania obrażeń u osób jadących na tylnych siedzeniach samochodu osobowego.

ZBROWSKI A, SAMBORSKI T. **Badania napędów elektromechanicznych stosowanych w systemach oddymiania i wentylacji**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 367–376.

W artykule przedstawiono badania eksperymentalne napędów elektromechanicznych pod kątem zastosowania w systemach wentylacji i oddymiania budynków. Wyznaczone charakterystyki (siła, przemieszczenie, prąd i napięcie zasilania w funkcji czasu) pozwalają na dokonanie oceny możliwości ich zastosowania do napędu klap i okien w systemach przeciwpożarowych. W badaniach wykorzystano opracowane stanowisko umożliwiające wyznaczanie charakterystyk pracy napędów liniowych i obrotowych w zmiennych warunkach temperatury.

MAZURKIEWICZ D. **Systemy informatycznego wspomagania eksploatacji i zapewnienia niezawodności przenośników taśmowych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 377–382.

Niezawodność eksploatacyjna systemów transportu przenośnikowego, to zagadnienie dotyczące zapewnienia odpowiedniego poziomu gotowości systemu transportu ciągłego, który w przypadku przenośników taśmowych zależny jest nie tylko od ich walorów użytkowych, determinowanych zastosowanymi rozwiązaniami konstrukcyjnymi i wykonawstwem, ale też zależy w istotnym stopniu od właściwego poziomu ich użytkowania, rozumianego między innymi jako stopień zaawansowania i skuteczności zastosowanych metod i narzędzi diagnostyki przemysłowej. W artykule scharakteryzowano najpopularniejsze systemy diagnostyczne stosowane w eksploatacji przenośnikowych systemów transportu wewnątrzzakładowego. Zaprezentowano również nową metodę komputerowego wspomagania w utrzymaniu ruchu tego typu systemów.

MYSTKOWSKI A, KARBAY VK, MYSTKOWSKA J. **A PLC based robust monitoring model for the labelling machine automation process**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 383–390.

This paper presents a method for improving the labelling process and a robust monitoring model for the labelling machine with the purpose of reducing waste of labels and bottles.The proposed monitoring method is based on a combination of Matlab®-designed and hardware-in-the-loop (HIL) simulationas well as Arena Simulation. The method solves problems with the application of labels during the labelling stage and provides a robust monitoring algorithm that recognizes defective labels before they are stuck onto bottles.The Grafcet optimal algorithm for recognizing defective labels is executed. The Matlab®Stateflow model for monitoring and recognizing defective labels is applied. The proposed algorithms are complete, and optimized solutions are ready for implementation in the existing PLC supervisory control system. Based on HIL simulations, the proposed method ensures an increase of the total production quantity. Statistical data was collected directly from the field, classified using Statfit software, and used in Arena Simulation software to present the difference and benefits before and after using the PLC-based robust monitoring model for the labelling machine automation process.

MYSTKOWSKI A, KARBAY VK, MYSTKOWSKA J. **Model odporny systemu monitorowania w automatyzacji procesu etykietowania z wykorzystaniem sterowników PLC**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 383–390.

W pracy przedstawiono metodę poprawy procesu etykietowania oraz model odporny monitorowania uszkodzeń etykiet w celu zmniejszenia ilości odpadów etykiet i butelek. Opracowanie proponowanej metody monitorowania i wykrywania wad etykiet opiera się na wykorzystaniu kombinacji funkcji środowiska Matlab® oraz symulacji sprzętowej (ang. hardware-in-the-loop, HIL). Nowa metoda rozwiązuje problemy związane z wykrywaniem uszkodzeń przyklejanych etykiet do butelek w przemysłowej linii produkcyjnej oraz zawiera model odporny detekcji wad etykiet. Algorytm systemu monitorowania w procesie etykietowania został przedstawiony za pomocą sieci Grafcet, a następnie zrealizowany w środowisku Matlab Stateflow®. Proponowane algorytmy monitorowania/detekcji zostały zoptymalizowane pod kątem ich realizacji w istniejącym systemie sterowania opartym o programowalne sterowniki logiczne (ang. programmable logic controllers, PLCs). Przeprowadzone symulacje sprzętowe HIL pomyślnie weryfikują opracowane rozwiązania podnoszące efektywność produkcji. Zaproponowany odporny model detekcji uszkodzeń etykiet został zaimplementowany w układzie sterowania linii produkcyjnej i zweryfikowany eksperymentalnie. Zebrane dane statystyczne bezpośrednio z obiektu sterowania zostały opracowane w programie Statfit. Oprogramowanie Arena Simulation zostało wykorzystane do porównania wyników pracy linii produkcyjnej przed i po wprowadzeniu modelu wykrywania uszkodzeń etykiet.

KATUNIN A, PRZYSTAŁKA P. **Detection and localization of delaminations in composite beams using fractional B-spline wavelets with optimized parameters**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 391–399.

In this paper the method of detection and localization of delaminations in composite layered beams using an algorithm based on the wavelet transform of bending modal shapes of beams was presented. For the basis functions the fractional B-spline wavelets with single- and multi-objective optimization of the values of their parameters were selected. The analysis was carried out basing on the results of numerical simulations. Several cases of the delaminations occurrence with respect to their location on the thickness of the beams, different sizes and geometrical features, were analyzed. Results of the conducted analyzes show the high effectiveness of a method in the task of detection of delaminations and a possibility of its application in industrial conditions.

KATUNIN A, PRZYSTAŁKA P. **Detekcja i lokalizacja rozwarstwień w kompozytowych belkach z wykorzystaniem B-splajnowych falek ułamkowych z optymalizowanymi parametrami**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 391–399.

W pracy przedstawiono metodę detekcji i lokalizacji rozwarstwień w kompozytowych belkach warstwowych z wykorzystaniem algorytmu opartego na transformacji falkowej giętnych postaci własnych drgań belek. Jako funkcje bazowe zastosowano ułamkowe falki B-splajnowe z jedno- i wielokryterialną optymalizacją wartości ich parametrów. Analizę przeprowadzono na wynikach obliczeń numerycznych. Przeanalizowano przypadki występowania rozwarstwień w różnej lokalizacji na grubości płyt oraz przypadki z rozwarstwieniami o różnych rozmiarach i postaciach geometrycznych. Wyniki przeprowadzonych analiz wykazały wysoką skuteczność metody w wykrywaniu rozwarstwień i możliwość jej zastosowania w warunkach przemysłowych.

MUTINGI M. **System reliability optimization: A fuzzy multi-objective genetic algorithm approach**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 400–406.

System reliability optimization is often faced with imprecise and conflicting goals such as reducing the cost of the system and improving the reliability of the system. The decision making process becomes fuzzy and multi-objective. In this paper, we formulate the problem as a fuzzy multi-objective nonlinear program. A fuzzy multi-objective genetic algorithm approach (FMGA) is proposed for solving the multi-objective decision problem in order to handle the fuzzy goals and constraints. The approach is able flexible and adaptable, allowing for intermediate solutions, leading to high quality solutions. Thus, the approach incorporates the preferences of the decision maker concerning the cost and reliability goals through the use of fuzzy numbers. The utility of the approach is demonstrated on benchmark problems in the literature. Computational results show that the FMGA approach is promising.

MUTINGI M. **System reliability optimization: A fuzzy multi-objective genetic algorithm approach**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 400–406.

Często spotykanym problemem w optymalizacji niezawodności systemu są niedokładnie określone i sprzeczne cele, takie jak zmniejszenie kosztów systemu przy jednoczesnej poprawie jego niezawodności. Proces podejmowania decyzji staje się wtedy rozmyty i wielokryterialny. W niniejszej pracy, sformułowaliśmy ten problem jako rozmyty wielokryterialny program nieliniowy (FMOOP). Zaproponowaliśmy metodę rozmytego wielokryterialnego algorytmu genetycznego (FMGA), która pozwala rozwiązać wielokryterialny problem decyzyjny z uwzględnieniem rozmytych celów i ograniczeń. Podejście to jest uniwersalne, co pozwala na rozwiązania pośrednie, prowadzące do rozwiązań wysokiej jakości. Metoda uwzględnia preferencje decydenta w zakresie celów związanych z kosztami i niezawodnością poprzez wykorzystanie liczb rozmytych. Użyteczność FMGA wykazano na przykładzie wzorcowych problemów z literatury. Wyniki obliczeń wskazują, że podejście FMGA jest obiecujące.

BRZOZOWSKI K, NOWAKOWSKI J. **Model for calculating compression ignition engine performance**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 407–414.

Optimising the performance of an internal combustion engine requires both empirical and theoretical work. The latter involves reasoning based on results yielded by mathematical models. This paper presents a computationally efficient model of the working cycle for a compression ignition engine. The model enables analysis of the working cycle of an engine with an electronically controlled common-rail type power supply and a controlled exhaust gas recirculation system. The model's parameters are chosen in a two-stage identification process based on the results of the experiments. The first stage of identifying the parameters requires formulating and solving an appropriate dynamic optimisation problem for multiple discrete points describing the engine's operation. To this end a genetic algorithm is used with an additional condition controlling the quality of the solution. Artificial neural networks are used for the second stage of identification. The paper shows an example of using the model to assess the influence of the kinetic combustion phase, which results from the way in which the injection process proceeds on the course of the working cycle. The accuracy of calculations with respect to basic parameters characterising the working cycle is also discussed.

BRZOZOWSKI K, NOWAKOWSKI J. **Model do wyznaczania parametrów pracy silnika o zapłonie samoczynnym**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 407–414.

Doskonalenie parametrów pracy silnika spalinowego poprzez odpowiednie sterowanie cyklem roboczym wymaga stosowania zarówno prac o charakterze doświadczalnym jak i obliczeniowym. W tym drugim przypadku podstawą wnioskowania są wyniki uzyskiwane z modeli matematycznych. Artykuł przedstawia efektywny obliczeniowo model cyklu roboczego silnika o zapłonie samoczynnym. Model umożliwia analizę cyklu roboczego silnika z elektronicznie sterowanym układem zasilania typu common-rail oraz układem sterowanej recyrkulacji spalin. Parametry modelu dobrano w dwuetapowym procesie identyfikacji bazującym na wynikach badań stanowiskowych. Pierwszy etap identyfikacji parametrów wymagał sformułowania i rozwiązania odpowiedniego zadania optymalizacji dynamicznej dla wielu dyskretnych punktów pracy silnika. W tym celu zastosowano algorytm genetyczny z dodatkowym warunkiem kontroli jakości rozwiązania. W drugim etapie identyfikacji do uogólnienia wyników wykorzystano sztuczne sieci neuronowe. W pracy przedstawiono przykład zastosowania modelu w ocenie udziału fazy spalania kinetycznego wynikającej z realizacji przebiegu procesu wtrysku na przebieg cyklu roboczego oraz przedstawiono dokładność obliczeń w odniesieniu do podstawowych parametrów charakteryzujących cykl roboczy.

VAIČIŪNAS G, BUREIKA G. **Approach modelling of constant interfailure process of renewal multi-unit fleet**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 415–421.

While railway companies operate rolling-stock, a substantial part of its expenses goes to maintenance and repair. However, the amount of repair works is directly proportional to the average age of a rolling-stock fleet or its reliability. When renewal an existing fleet of a few dozen rolling-stocks, the installation of the new vehicles reduces the overall failure amount of the fleet proportionally to the number of acquired vehicles. This article provides a concept for creating the model of a passenger rolling-stock's failure intensity according to the mileage. According to this model, a vehicle fleet renewal algorithm can be created and used in order to limit the fluctuation of the fleet's average failure intensity as much as possible and to achieve the most accurate correlation between the number of failures and the fleet's average mileage. Thus a railway company has an opportunity to avoid the unplanned expenses for repairing the vehicles during the unforeseen failure peaks. The SPLINE method is proposed in order to indicate the vehicle failure flow's dependency on the vehicle mileage. After using this method to indicate the variation of the fleet's constant interfailure according to the mileage, the fleet's failure intensity can be modelled according to the algorithm of installing the acquired vehicles for operation.

SZKODA M. **Assessment of reliability, availability and maintainability of rail gauge change systems**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 422–432.

The paper provides a comparative assessment of the reliability of two rail gauge change systems: wagon bogie exchange and SUW 2000 self-adjusted wheel sets. In the applied method of assessment, reliability is treated as a comprehensive feature comprising such system characteristics as reliability itself together with availability and maintainability. The calculations of selected reliability ratios, based on operation data, demonstrated that the SUW 2000 system may be an alternative method for overcoming the barrier of different track gauges compared to the current wagon bogie exchange.

IDZIASZEK Z, GRZESIK N. **Object characteristics deterioration effect on task realizability – outline method of estimation and prognosis**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 433–440.

The article introduces the essence of a potential technical object work range. Vital issues connected with the modeling of an object work range deterioration which are influenced by the destructive processes derived from environment, operation and wear of the object, were discussed/described. Typical destructive processes were described and deterministic and probabilistic models which allow for evaluation and prognosis of an object durability were included in the description. An outline of the approach to object work range deterioration adopted by the authors was presented. An outline of an object condition models for evaluation and prognosis of its durability purposes including a complex issues of random influence of the many factors which affect changes in an object work range and influencing the quality of the performed tasks were shown. In the models including randomness, probabilistic tools/apparatus and fuzzy logic were adopted. This kind of approach in modeling the changes in object durability adopted by the authors aims at bringing the models of object durability change closer to operational reality and at the same time at better utilization of their potential work range while maintaining the assumed level of reliability/safety during operations.

DĄBROWSKI JR, KLEKOTKA M, SIDUN J. **Fretting and fretting corrosion of 316L implantation steel in the oral cavity environment**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 441–446.

Processes of mechanical destruction of implants, dental prosthetics elements, and orthodontic apparatus considerably limit their operating lifetime and the comfort of patients. Processes of destruction of kinematic joint elements caused by fretting and fretting corrosion processes are an important problem, albeit one that is not yet fully understood. This paper presents the results of fretting and fretting corrosion studies conducted on 316L implantation steel, which is used in dentistry, particularly in prosthetic and orthodontic applications. Tests were performed by means of an original device of the authors' own design, with the application of methodology developed by the authors. Fretting and corrosion tests were carried out in phosphate buffered saline (PBS) as well as in the presence of natural saliva and its substitutes. Own compositions of artificial saliva were developed for the purposes of studies. Observations of sample surfaces were performed using a scanning electron microscope (SEM) and a confocal microscope. Test results indicate a significant influence of fretting on the corrosion of 316L steel (fretting corrosion) as well as the important role of the studied fluids (saliva and its studies) in these processes. It was stated that the saliva substitute containing mucin III was characterized by the most favorable tribological characteristics. During fretting tests, intensive phenomena of materials conveyance into the friction contact area were observed.

VAIČIŪNAS G, BUREIKA G. **Podejście do modelowania stałego procesu międzyawaryjnego odnowy wieloelementowej floty pasażerskiej**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 415–421.

Znaczna część wydatków ponoszonych przez przedsiębiorstwo kolejowe z tytułu eksploatacji taboru kolejowego, to wydatki na konserwację i naprawy. Jednakże, należy pamiętać, że liczba prac remontowych jest wprost proporcjonalna do średniego wieku floty taboru lub jej niezawodności. Przy odnawianiu istniejącej floty kilkudziesięciu pojazdów kolejowych, wprowadzenie nowych pojazdów zmniejsza ogólną ilość awarii floty proporcjonalnie do liczby nowo nabytych pojazdów. W artykule przedstawiono koncepcję stworzenia modelu intensywności uszkodzeń taboru pasażerskiego w funkcji przebiegu kilometrowego. Zgodnie z tym modelem, można stworzyć algorytm odnowy floty pojazdów, z wykorzystaniem którego można maksymalnie ograniczyć wahania średniej intensywności uszkodzeń floty oraz osiągnąć najbardziej dokładną korelację między liczbą uszkodzeń a średnim przebiegiem pojazdów floty. W ten sposób przedsiębiorstwo kolejowe ma szansę uniknąć nieplanowanych wydatków na naprawy pojazdów podczas nieprzewidzianych okresów wzmożonej awaryjności. Zaproponowano metodę SPLINE, za pomocą której można określić zależność awaryjności pojazdu od jego przebiegu kilometrowego. Zastosowanie tej metody pozwala ustalić zmiany w stałym procesie międzyawaryjnym zależne od przebiegu, co z kolei pozwala na modelowanie intensywności uszkodzeń floty według algorytmu wprowadzania nowo nabytych pojazdów do eksploatacji.

SZKODA M. **Ocena nieuszkadzalności, gotowości i podatności utrzymaniowej kolejowych systemów przestawczych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 422–432.

Praca dotyczy porównawczej oceny niezawodności dwóch kolejowych systemów przestawczych: systemu wymiany wózków wagonowych i systemu samoczynnie rozsuwanych zestawów kołowych SUW 2000. W zastosowanej metodzie oceny, niezawodność jest traktowana jaka właściwość kompleksowa obejmującą takie cechy systemów jak: nieuszkadzalność, gotowość i podatność utrzymaniową. Przeprowadzone obliczenia wyselekcjonowanych wskaźników niezawodnościowych, oparte na danych eksploatacyjnych wykazały, że system SUW 2000 może stanowić alternatywną metodę pokonania bariery rożnej szerokości toru w stosunku do aktualnie stosowanej wymiany wózków wagonowych.

IDZIASZEK Z, GRZESIK N. **Zarys metody oceny trwałości i niezawodności obiektu z uwzględnieniem czynnika ludzkiego i płaszczyzny liczb zespolonych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 433–440.

W artykule dokonano wprowadzenia w istotę pojęcia tzw. potencjalnego zasobu pracy obiektu technicznego. Opisano istotne zagadnienia związane z modelowaniem zużywania zasobu pracy, na które wpływają procesy destrukcyjne od środowiska, użytkowania i obsług. Wskazano na typowe procesy destrukcyjne i dla nich przedstawiono modele deterministyczne i probabilistyczne umożliwiające ocenę oraz prognozę zużywania potencjalnego zasobu pracy obiektu dla przyjętego poziomu niezawodności lub trwałości obiektu.Przedstawiono też zarys realizowanego przez autorów podejścia w modelowaniu zużywania zasobu pracy obiektu.Pokazano zarys modeli stanu obiektu do oceny i prognozy jego trwałości z uwzględnieniem zagadnień losowego wpływu wielu czynników wpływających na zmianę zasobu pracy obiektu, a tym samym, na jakość realizowanych zadań. W modelach uwzgledniających losowość przyjęto aparat probabilistyczny oraz wykorzystano logikę rozmytą. Tak przyjęte przez autorów podejście w modelowaniu zmian niezawodności/trwałości obiektu, ma na celu lepsze przybliżenie do rzeczywistości eksploatacyjnej, a tym samym lepsze wykorzystanie ich potencjalnego zasobu pracy, przy zachowaniu założonego poziomu niezawodności/bezpieczeństwa w trakcie realizacji działania/uzyskania efektu. Na koniec pokazano nowatorskie na skalę światową podejście, pozwalające na łączenie w jednym modelu technicznych i nietechnicznych aspektów oceny i prognozy zmian jakości obiektów w eksploatacji poprzez wykorzystanie do tego celu płaszczyzny liczb zespolonych.

DĄBROWSKI JR, KLEKOTKA M, SIDUN J. **Fretting i fretting-korozja stali implantacyjnej 316L w środowisku jamy ustnej**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 441–446.

Procesy destrukcji metalicznych implantów, elementów protetyki stomatologicznej i aparatów ortodontycznych znacznie ograniczają ich trwałość eksploatacyjną i komfort pacjentów. Szczególnym zagadnieniem, aczkolwiek dalece niepoznanym, są procesy niszczenia elementów połączeń kinematycznych wywołane procesami frettingu i frettingu – korozji. W pracy przedstawiono wyniki badań frettingu i fretting-korozji stali implantacyjnej 316L – używanej w stomatologii, szczególnie w zastosowaniach protetycznych i ortodontycznych. Badania realizowano za pomocą oryginalnego urządzenia własnej konstrukcji, z wykorzystaniem metodyki opracowanej przez autorów. Badania frettingu i korozji przeprowadzone zostały w buforze fosforanowym (PBS) jak również w obecności śliny naturalnej i jej substytutów. Na potrzeby badań opracowano własne kompozycje sztucznych ślin. Obserwacje powierzchni próbek prowadzone były z wykorzystaniem skaningowego mikroskopu elektronowego (SEM) oraz mikroskopu konfokalnego. Wyniki badań wskazują na znaczący wpływ frettingu na niszczenie korozyjne stali 316L (fretting-korozja), a także na istotną rolę badanych płynów (śliny i jej substytutów) w tych procesach. Stwierdzono, że najkorzystniejszymi charakterystykami tribologicznymi charakteryzował się substytut śliny zawierający mucynę III. W trakcie testów frettingu obserwowano intensywne zjawiska przenoszenia materiałów w strefie kontaktu tarciowego.

ODHIAMBO S, DE MEY G, HERTLEER C, VAN LANGENHOVE L. **Reliability testing of PEDOT:PSS capacitors integrated into textile fabrics**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 447–451.

Textile-based capacitors have been made from polyethylene dioxythiophene, polystyrene sulphonate (PEDOT:PSS) as the electrolyte and pure stainless steel filament yarns as the electrodes. The capacitor is well integrated into the textile structure, small in size and of light weight. Although they experience a self-discharge, the reliability of the PEDOT:PSS capacitors has been investigated by repeating up to 14 cycles of charging and discharging. Initially, the voltage output turns out to be higher with increasing number of cycles. However, after the fifth cycle, degradation of the cell starts occurring and a decreasing behaviour in the voltage output is observed. One can roughly say that these capacitors could be used up to 10–15 cycles.

MICHALSKI R, GONERA J, JANULIN M. **A simulation model of damage-induced changes in the fuel consumption of a wheeled tractor**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 452–457.

The existing diagnostic systems are applied to monitor and optimize a tractor's performance and effectiveness, while there is no consumption monitoring systems in terms of damage. A malfunction can be analyzed at different levels of complexity, including systems, kinematic pairs and components. Based on the generated consequences, defects can be classified into the following groups. Are the following classes of damage: damage to functional, emission, unsafe deteriorating dynamics, which are assigned to certain effects. The study was prepared simulation model wheeled tractor, which incorporates traction characteristics describe physical phenomena associated with the operation of the tractor and having an impact on the process of degradation under certain loading cycles. An algorithm for determining fuel consumption during simulated defects in a wheeled tractor is presented in this paper. Most tractor malfunctions affect fuel consumption. Fuel consumption is one of the key diagnostic parameters in evaluations of a vehicle's technical condition. The magnitude of changes in fuel consumption varies subject to the type of defect. The effects of simulated malfunctions of a wheeled tractor on its fuel consumption are discussed in this paper.

BURDZIK R. **Implementation of multidimensional identification of signal characteristics in the analysis of vibration properties of an automotive vehicle's floor panel**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 458–464.

The article provides a proposal of software application of a method and an algorithm developed to identify signal characteristics in the analysis of vibration properties of an automotive vehicle's floor panel. Due to the complexity resulting from nonlinear and random nature of vibration phenomena in automotive vehicles, the analysis in question is multidimensional. The property table being established consists of numerous measures and estimators, both dimensional and dimensionless ones, in the domains of amplitudes, time, frequency and time-frequency. The foregoing enables observation and separation of signal components in multiple domains, but it also makes it possible to define signal measures depending on stationary and non-stationary characteristics as well as accurate time positioning of resonant frequencies. Multicriterial approach to identification of vibration enables determining the table of vibration properties measures of floor panel. The table is numerical form of characteristics properties of the vibration signal.

LAI M-H. **Optimal number of Minimal Repairs under a Cumulative Damage Model with Cumulative Repair Cost Limit**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 465–471.

In this paper, we consider a repair number counting replacement policy under a cumulative damage model, in which the policy includes the concept of a cumulative repair cost limit. The system experiences two kinds of shocks: a type I shock causes a random amount of damage to the system leading to a serious failure when the total damage exceeds a failure level; or a type II shock causes the system into minor failure which can be corrected by minimal repair. When a minor failure occurs, the repair cost will be evaluated and minimal repair is executed if the accumulated repair cost is less than a predetermined limit L. The system is replaced anticipatively at *n-th* minor failure, or at the *j-th* minor failure ($j < n$) at which the accumulated repair cost exceeds a predetermined limit L, or any serious failure. In order to assess the performance of the proposed maintenance policy and to minimize the long-term expected cost per unit time, a mathematical model for the maintained system cost is derived. By minimizing that cost, the optimal number n^* is also verified finite and unique under certain conditions. Analyses based on numerical results are conducted to highlight the properties of the proposed maintenance policy in respect to the different parameters.

ODHIAMBO S, DE MEY G, HERTLEER C, VAN LANGENHOVE L. **Badanie niezawodności kondensatorów PEDOT: PSS wbudowanych w tkaniny**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 447–451.

Kondensatory tekstylne wytwarza się z mieszaniny poli(3,4-dioksyetylenotiofenu) z polistyrenem sulfonowanym (PEDOT: PSS), pełniącej rolę elektrolitu oraz włókien ciągłych z czystej stali nierdzewnej, pełniących funkcję elektrod. Kondensatory tego typu są dobrze zintegrowane ze strukturą tkaniny, są lekkie i mają niewielkie rozmiary. Chociaż kondensatory PEDOT:PSS ulegają samorozładowaniu, przeprowadzono badania ich niezawodności powtarzając 14 cykli ładowania i rozładowywania Początkowo napięcie wyjściowe zwiększało się wraz ze wzrostem liczby cykli. Jednakże po piątym cyklu, dochodziło do degradacji ogniwa i obserwowano zmniejszanie się napięcia wyjściowego. Można orientacyjnie powiedzieć, że omawiane kondensatory nadają się do użytku przez maksymalnie 10–15 cykli.

MICHALSKI R, GONERA J, JANULIN M. **Model symulacyjny zmian zużycia paliwa ciągnika kołowego w aspekcie uszkodzeń**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 452–457.

Stosowane obecnie elektroniczne systemy w ciągnikach kołowych służą do monitorowania i optymalizacji efektów pracy pod kątem jego wydajności i efektywności, natomiast brak jest systemów monitorujących zużycie paliwa w aspekcie uszkodzeń. Uszkodzenie można rozpatrywać na różnych poziomach złożoności maszyny, np. układów, zespołów węzłów konstrukcyjnych lub elementów. W pracy przyjęto klasyfikację uszkodzeń ze względu na ich skutki. Wyróżniono następujące klasy uszkodzeń: uszkodzenia funkcjonalne, emisyjne, zagrażające bezpieczeństwu, pogarszające dynamikę, którym przyporządkowano określone skutki. W pracy przygotowano model symulacyjny ciągnika kołowego, w którym uwzględniono charakterystyki trakcyjne opisujące zjawiska fizyczne związane z funkcjonowaniem ciągnika i mające wpływ na proces jego degradacji w określonych cyklach obciążeń. Przygotowano algorytm służący do określania zmian zużycia paliwa przy symulowanych uszkodzeniach ciągnika kołowego. Większość z uszkodzeń ciągnika kołowego ma wpływ na zużycie paliwa. Zużycie paliwa może być jednym z podstawowych parametrów diagnostycznych podczas oceny stanu technicznego pojazdu. W zależności od rodzaju uszkodzenia zmiany zużycia paliwa mogą być różne. W pracy przedstawiono przykładowe przebiegi symulacyjne niektórych uszkodzeń ciągnika kołowego na zużycie paliwa.

BURDZIK R. **Implementation of multidimensional identification of signal characteristics in the analysis of vibration properties of an automotive vehicle's floor panel**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 458–464.

The article provides a proposal of software application of a method and an algorithm developed to identify signal characteristics in the analysis of vibration properties of an automotive vehicle's floor panel. Due to the complexity resulting from nonlinear and random nature of vibration phenomena in automotive vehicles, the analysis in question is multidimensional. The property table being established consists of numerous measures and estimators, both dimensional and dimensionless ones, in the domains of amplitudes, time, frequency and time-frequency. The foregoing enables observation and separation of signal components in multiple domains, but it also makes it possible to define signal measures depending on stationary and non-stationary characteristics as well as accurate time positioning of resonant frequencies. Multicriterial approach to identification of vibration enables determining the table of vibration properties measures of floor panel. The table is numerical form of characteristics properties of the vibration signal.

LAI M-H. **Optymalna liczba napraw minimalnych w świetle Modelu Sumowania Uszkodzeń przy ograniczonym łącznym koszcie napraw**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 465–471.

W przedstawionym artykule omawiamy politykę wymiany systemu opartą na modelu sumowania uszkodzeń polegającą na obliczaniu liczby napraw. Polityka ta obejmuje koncepcję limitu łącznego kosztu napraw. System może być narażony na działanie dwóch rodzajów szkodliwych czynników: czynniki I-ego typu powodują losowo określony zakres uszkodzeń systemu, prowadząc do poważnej awarii, gdy łącznie uszkodzenia przekraczają poziom awarii; lub czynniki typu II-ego powodujące drobne uszkodzenia, które można skorygować poprzez minimalną naprawę. Gdy dochodzi do niewielkiego uszkodzenia, wtedy szacuje się koszt naprawy i realizuje minimalną naprawę, jeśli łączny koszt naprawy jest niższy od uprzednio ustalonego limitu L. System zostaje prewencyjnie wymieniony albo przy *n-tej* drobnej awarii albo przy *j-tej* drobnej awarii ($j <n$), przy której łączny koszt naprawy przekracza uprzednio ustalony limit L lub też przy jakimkolwiek poważnym uszkodzeniu. W celu oceny skuteczności proponowanej polityki obsługiwania i zminimalizowania przewidywanego długoterminowego kosztu przypadającego na jednostkę czasu, wyprowadzono model matematyczny kosztów dla obsługiwanego systemu. Poprzez minimalizację tych kosztów, określono również optymalną liczbę napraw n^*, która w pewnych warunkach jest liczbą skończoną i niepowtarzalną. W oparciu o wyniki numeryczne, przeprowadzono analizy mające na celu naświetlenie właściwości proponowanej polityki obsługiwania w odniesieniu do różnych parametrów.

HRYNIEWICZ O, KARPIŃSKI J. **Prediction of reliability – the pitfalls of using Pearson's correlation**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 472–483.

Pearson's coefficient of linear correlation r is the measure of dependence which is the most popular among practitioners. In the paper we have shown, using comprehensive computer simulations, that its application is very limited when we search for informative variables that can be used for the prediction of reliability. We have shown that Kendall's coefficient of association τ is much better for this purpose.

HRYNIEWICZ O, KARPIŃSKI J. **Prognozowanie niezawodności – pułapki związane z używaniem współczynnika korelacji Pearsona**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 472–483.

Współczynnik korelacji liniowej r Pearsona jest najbardziej popularną wśród praktyków miarą zależności statystycznej. W artykule na podstawie wyników wyczerpujących symulacji komputerowych pokazano, że w przypadku poszukiwania zmiennych mogących służyć do prognozowania niezawodności zakres jego stosowalności jest bardzo ograniczony. Wyniki badań symulacyjnych pokazują, że temu celowi lepiej służy współczynnik asocjacji τ Kendalla.

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.

Electron multiplier (EM) is a kind of highly reliable and long-lifetime vacuum electronic device applied widely in spectrometry, space exploration and atom frequency standard. It is a critical device which might constrain the related technology. A challenge remains for researcher and engineer how to predict the life span of EM. Firstly, degradation mechanism of EM is investigated. It shows that the secondary emission ratios of each multiplier electrode reduces gradually with operating time, which results in the degradation of the key performance index of EM, i.e. the gain of electric current. So an accelerated degradation test (ADT) methodology using dual stresses is proposed to predict the life span of EM. Secondly, the ADT plan with dual stresses is designed and carried out by the corresponding test system established. Finally, the data analysis procedure is presented, and its validity is investigated by model verification. The presented method can sharply reduce testing time and cost because of using accelerated stress which can accelerate degradation process of EM. This method can also provide a new way to lifetime and reliability prediction for other products with long lifetime and high reliability.

WANG Y, ZHANG C, CHEN X, TAN Y. **Metoda prognozowania cyklu życia powielacza elektronów oparta na przyspieszonych badaniach degradacji**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 484–490.

Powielacz elektronów (EM) to elektroniczne urządzenie próżniowe o wysokiej niezawodności i długim cyklu życia, które znajduje szerokie zastosowanie w spektrometrii i badaniach przestrzeni kosmicznej, a także w atomowych wzorcach częstotliwości. Jest to urządzenie krytyczne, które może stanowić ograniczenie dla technologii, w której jest wykorzystywane. Wyzwaniem dla naukowców i inżynierów pozostaje pytanie, jak przewidzieć żywotność EM. W pierwszej kolejności w artykule zbadano mechanizm degradacji EM. Badanie pokazało, że współczynniki emisji wtórnej elektrody powielacza maleją stopniowo wraz z upływem czasu pracy, co prowadzi do degradacji kluczowego wskaźnika wydajności EM, to znaczy wzmocnienia prądu elektrycznego. W oparciu o ten fakt, zaproponowano metodę prognozowania żywotności EM zasadzającą się na metodologii przyspieszonych badań degradacji (ADT) z wykorzystaniem podwójnych naprężeń. Następnie zaprojektowano i zrealizowano plan ADT z podwójnymi naprężeniami za pomocą odpowiedniego systemu testowego. Na koniec przedstawiono procedurę analizy danych, a ich wiarygodność zbadano poprzez weryfikację modelu. Przedstawiona metoda może znacznie zredukować czas i koszty badań dzięki wykorzystaniu przyspieszonych naprężeń, które mogą przyspieszyć proces degradacji EM. Metoda ta może również umożliwić nowy sposób przewidywania niezawodności i cyklu życia produktów o długim cyklu życia i wysokiej niezawodności.

STELMASIAK Z. **Uniformity of Diesel oil dosage in dual fuel engines**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 491–495.

Adaptation of compression ignition engines to dual fuel supply can be accomplished both in case of modern engines equipped with common rail system and older engines equipped with classic injection system (piston pumps). Due to big differences in price of gaseous and liquid fuels there is observed a natural tendency to use very small initial doses. At current level of introduction of gaseous fuels to powering of traction engines there is a need to provide alternating fuelling of the engine with the Diesel only and in dual fuel system. It requires usage of original injection systems in the dual fuel engines, what largely restricts possibility of reduction of the initial doses. In the paper are presented investigations concerning uniformity of the dosing by in-line piston pump of the P56-01 type and two types of injectors with common rail system. The investigations have shown that the P56-01 pump adjusted for nominal doses shows big non-uniformity of the dosage in area of small doses. Improved uniformity of the dosage can be attained in result of adjustment for a smaller doses, what allows reduction of the dose to about 15-20% of nominal dose and improves smoothness of engine operation. Also in case of the injectors in common rail system, reduction of the doses is limited due to worsening of uniformity of the dosing from one cycle to another, and failure of the dosing. It results from the fact, that minimal doses in dual fuel system are smaller than the ones present in case of idling speed when the engine is run on the Diesel oil only. In case of the injectors in common rail system, minimal initial doses possible to be obtained are equal to 10-15% of the nominal dose.

STELMASIAK Z. **Równomierność dawkowania oleju napędowego w silnikach dwupaliwowych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 491–495.

Adaptacja silników o zapłonie samoczynnym do zasilania dwupaliwowego może być dokonywana zarówno dla nowoczesnych silników z systemami common rail jak i starszych silników wyposażonych w klasyczną aparaturę wtryskową (pompy tłoczkowe). Ze względu na duże różnice cen paliw gazowych i ciekłych występuje naturalna tendencja do stosowania bardzo małych dawek inicjujących. Na obecnym poziomie wprowadzania paliw gazowych do zasilania silników trakcyjnych istnieje konieczność zachowania przemiennego zasilania silnika samym olejem napędowym i w systemie dwupaliwowym. Wymaga to zastosowania w silnikach dwupaliwowych oryginalnej aparatury wtryskowej co ogranicza w znacznym stopniu możliwość zmniejszania dawek inicjujących. W pracy przedstawiono badania równomierności dawkowania tłoczkowej pompy rzędowej P56-01 oraz dwóch typów wtryskiwaczy układu common rail. Badania pokazały, że pompa P56-01 wyregulowana dla dawek znamionowych wykazuje dużą nierównomierność dawkowania w zakresie małych dawek. Poprawę równomierności dawkowania można uzyskać przez regulacją pompy dla dawek mniejszych, co pozwala zmniejszyć dawkę do około 15-20% dawki znamionowej i poprawia równomierność pracy silnika. Również w przypadku wtryskiwaczy common rail zmniejszanie dawek jest ograniczone z powodu pogorszenia równomierności dawkowania z cyklu na cykl i zaniku dawkowania. Wynika to z faktu, że minimalne dawki w systemie dwupaliwowym są mniejsze od występujących dla wolnych obrotów przy zasilaniu samym olejem napędowym. W przypadku wtryskiwaczy common rail minimalne dawki inicjujące jakie można uzyskać wynoszą 10-15% dawki znamionowej.

Article citation info:
RYBAK P. Operating loads of impulse nature acting on the special equipment of the combat vehicles. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 347–353.

Piotr RYBAK

OPERATING LOADS OF IMPULSE NATURE ACTING ON THE SPECIAL EQUIPMENT OF THE COMBAT VEHICLES

OBCIĄŻENIA EKSPLOATACYJNE O CHARAKTERZE UDAROWYM DZIAŁAJĄCE NA WYPOSAŻENIE SPECJALNE WOZÓW BOJOWYCH

Providing the combat vehicles with high operation effectiveness, safety and reliability during execution of complex tasks makes a priority. Therefore the armament and the military equipment have to meet very high requirements in that aspect when used in various conditions. This paper presents basic sources of dynamic loads affecting the combat vehicles. Attention is paid to the loads of impact nature as they mostly affect the effectiveness and reliability of a vehicle, electronic equipment and psychophysical condition of the combat vehicle crew. These loads result from off-road drives, firing the gun, the influence of the land mines or IED, hitting by enemy's missile. As a result, some fragments of the experimental and model tests on combat vehicles are presented. Results of these tests can be helpful when designing internal vehicle equipment including special equipment. Particularly in the aspect of normative requirements for that class of vehicles and their special equipment.

Keywords: *special equipment, combat vehicle, main battle tank, impact loads, research.*

Zapewnienie wozom bojowym wysokiej skuteczności działania, bezpieczeństwa oraz niezawodności podczas realizacji złożonych zadań jest traktowane priorytetowo. A zatem uzbrojenie i sprzęt wojskowy musi spełniać bardzo wysokie wymagania w tym aspekcie podczas eksploatacji w różnych warunkach. W pracy przedstawiono podstawowe źródła obciążeń dynamicznych działające na wozy bojowe. Uwagę skupiono na obciążeniach mających charakter udarowy, gdyż one głównie wpływają na sprawność i niezawodność pojazdu, urządzeń wewnętrznych i stan psychofizyczny załogi. Obciążenia te wynikają z jazd terenowych, strzelania z armaty, oddziaływania miny lub IED, trafienia pociskiem przeciwnika. W rezultacie przedstawiono niektóre fragmenty z badań eksperymentalnych i modelowych wozów bojowych. Wyniki tych badań mogą być pomocne przy projektowaniu urządzeń wewnętrznych pojazdu w tym urządzeń specjalnych. Szczególnie w aspekcie wymagań normatywnych dla tej klasy pojazdów oraz ich urządzeń specjalnych.

Słowa kluczowe: *wyposażenie specjalne, wóz bojowy, obciążenie udarowe, badania.*

1. Introduction

Combat (caterpillar and wheeled) vehicles, as a basic mean of execution of tasks by the army, are designed to execute special tasks in heavy-duty operation conditions. When riding on the roads and in the wilderness they are subject to the influence of dynamic reactions in a constant manner [5, 6, 11]. These loads have very complex structure and differ from each other with many factors, including values, nature, duration, intensity as well as the direction of action. The level of dynamic loads is determined by mutual interactions of a complex system, which consists of the following elements, namely: combat vehicle – internal equipment – ambient environment (the ground) – crew members [1, 12, 15]. A diagram of mutual interactions of components of the aforementioned system are presented on fig. 1.

The main factors causing dynamic loads while driving, affecting motor vehicles, including internal equipment and people located inside, are as follows:

- high driving resistance of significant variation frequency – type of the ground and resulting force reactions,
- dynamic loads resulting from driving at high speeds on various roads (including when crossing natural and artificial obstacles) – fig. 2a,
- influence of inertia (sudden acceleration and delay of motion, driving along a road curve, skidding etc.) – fig. 2b,
- engine, driving system and chassis.

In the caterpillar vehicles, due to the complexity of chassis, generated loads result from:

- cooperation of caterpillar bands with driving wheels,
- waving of upper sections of caterpillars,
- execution of serpentine motion, short jumps – as an element of active defence.

There are additional sources of dynamic loads in the combat vehicles resulting from combat operations. These are loads of impact nature, of various direction and intensity of action. They include:

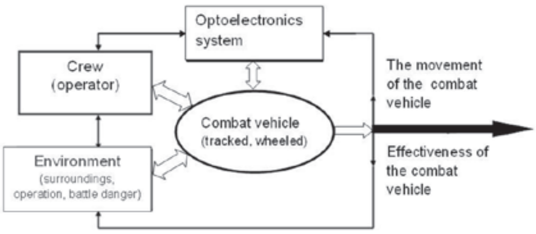


Fig. 1. Diagram of interactions in the system: combat vehicle – human – internal equipment – surroundings a) dynamic overcoming of a slope [3], b) overcoming of a cross ditch

Fig. 2. Combat vehicles during a ride in typical road conditions

- Firing the gun (fig. 3); level of impact depends on the gun calibre and type, missile type, condition of resistance-returners and other factors. The load during firing amounts to a value of several hundreds kN in a reaction of a second and generates the recoil. This force affects the tank tower, therefore it affects the internal equipment, the tank body and the crew. An approximate value of the tank gun recoil can be defined according to the Vallier's hypothesis [19], from the following dependence:

$$R_0 = \frac{0.5 \cdot M_o \cdot w^2_{\max}}{\lambda - L_k + w_{\max} \cdot t_p} \ \text{[kN]}, \tag{1}$$

  where: M_o – recoil unit weight [kg], λ– assumed recoil length [m], w_{max} – maximum free recoil velocity [m/s], L_k – free recoil path at the end of the post-exhaust period of the gun powder gas effect, t_p – time of completion of the post-exhaust gun powder action [s].
- A hit of a missile or a fragment in the armour without piercing. An effect of such hit is presented on fig. 4. Value of energy of a hitting element can amount to several or over a dozen of MJ. According to [4, 7] the impact energy can be defined on the basis of the following dependence:

Fig. 3. Firing the tank gun

$$E_p = \frac{m_p \cdot V_p^2}{2} \ \text{[MJ]} \tag{2}$$

  where: m_p – missile weight, V_p – outlet missile velocity.
- Influence of firing factors after the explosion of mines and improvised explosive devices (IED). In case of an impact wave load generated by explosion of a mine, according to a diagram shown on fig. 5, a value of the maximum pressure affecting a vehicle can be estimated according to [Kozłow A. G., Tału K. A., *Tank Structure and Calculation*. Moscow 1958] from the following dependence:

$$p = 60\psi \frac{m_{MW}{}^{0,87}}{r^{2,6}}(1+\cos\Theta) \ \text{[Pa]}, \tag{3}$$

  where: m_{MW} – explosive weight, r – distance from a centre of explosion, Θ – an angle of the impact wave contacting the tank bottom, ψ – coefficient considering a mine depth in the ground and the loss of energy on the influence of the impact wave on the ground and the caterpillar (if explosion takes place under a caterpillar).

Listed loads refer to vehicles operating in the direct threat area and they mostly concern the infantry combat vehicles, wheeled armoured soldier carriers and other construction applications on their chassis.

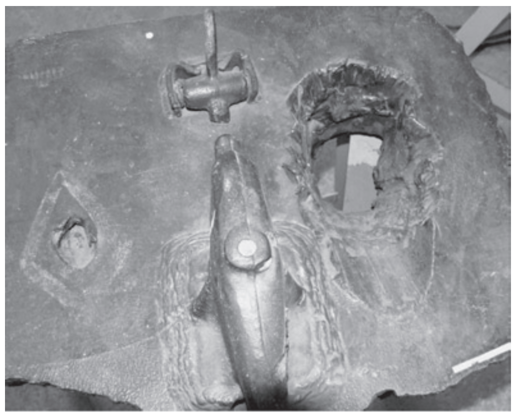


Fig. 4. An armour after missile impact (armour after tests in the Motor Vehicle Institute of the Military Technical University) – a photo from own collections

Fig. 6 presents an arrangement of special equipment in a combat vehicle body against the C centre of mass.

The equipment is located at a significant distance from a vehicle, depending on a vehicle type from 800 mm to even 2500 mm. Such equipment location has a significant influence on a dynamic load level.

When designing a new vehicle or modernizing and existing one, a number of design works are carried out in order to improve its effec-

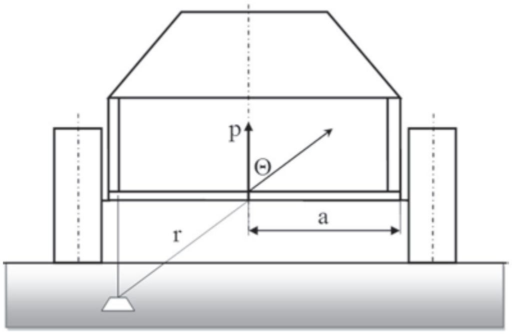


Fig. 5. Diagram of influence of a mine on a vehicle

Fig. 6. Arrangement of special equipment in a combat vehicle

tiveness. Multi-variant and multi-aspect model studies for expected loads are carried out. However, the quality, durability and reliability of introduced solutions and equipment operation can be evaluated only when the experimental tests, meeting the normative requirements, are carried out. [10]. This paper presents selected fragments from the experimental tests on combat vehicles subject to various impact loads. Their results make a good basis and can be helpful when designing the equipment of increased resistance to mechanical impacts.

It should be underlined that works related to the research and evaluation of effects of dynamic load influence on motor vehicles (both short-term and long-term ones) have been carried out for years both in national and foreign centres, presented among others [6, 8, 9, 13]. However, published results refer to a different aspect (load influence on people) and other vehicle category (light vehicles). As far as heavy vehicles are concerned, information about the works in this area is hardly available and not available for public. Therefore this paper tries to present an issue of influence of loads of impulse nature on a combat vehicle (tank) and its internal equipment during various operation conditions. Addressing that issue also results from a fact that the tank crew without properly operating devices of the internal equipment (including special equipment) does not represent a fully valuable mean of combat.

Due to used combat vehicles and scope of performed tests (parameters, conditions), obtained results are characterized by certain sensitivity and therefore they are of a quality nature.

2. Experimental tests on the influence of impulses on a combat vehicle body

2.1. Tests during off-road drives

The load tests were carried out in the premises of the Military Technical University at two stages. The first stage included test drives at pre-set speeds across single triangle prisms of pre-set geometrical parameters. Fig. 7a and b present selected courses of vertical accelerations of a frontal part of a tank body for various initial conditions of a caterpillar chassis. The second stage was carried out during drives on the ground road of average corrugation level.

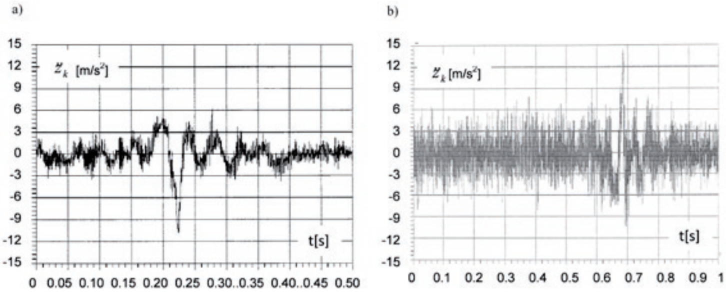


Fig. 7. Courses of vertical body accelerations, crossing a triangle prism of a height of 170mm and length of lo at the speed v=5.56m/s: a) and b) reflect various conditions of a caterpillar chassis

Table 1. Maximum vertical acceleration values of the front part of a vehicle body during off-road drives on the ground road of average corrugation level.

No.	Driving speed V [m/s]	Vertical accelerations of a vehicle body $\ddot{z}_k$ [m/s²]
1.	2,78	9,9
2.	4.17	14,7
3.	5,56	16,6

Table 1 specifies vertical acceleration values registered in the front part of a vehicle body for various driving speeds.

The analysis of obtained test results clearly indicates which factors determine the dynamic load level. Condition of a caterpillar chassis, speed acceleration and road type as well as a level of corrugation have a significant influence on the dynamic load increase. For wilderness of high corrugation level of random location and higher driving speeds (from 8 to 15 m/s), the level of loads, the impulse ones in that case, can exceed the values presented in the table by many times.

If the analysis assumes a drive on a frozen and ploughed field or on the stone or rock (rubble) ground, then a type of a vehicle chassis is important in the aspect of dynamic load level. Results of performed tests indicate the advantage of the caterpillar chassis (characterized by an ability to smooth the bumps of the ground) over the wheeled chassis.

2.2. Tests carried out at the influence of mine explosion on a vehicle

Introduction of non-contact reaction mines and improvised explosive devices (IED) among anti-armour means has increased the issue of resistance of combat vehicles and their equipment. The pressure of the explosion impact wave pressure dispersing at the supersonic velocity is the main firing factor of mines and IED. The wave affects an encountered obstacle. In that case it affects a combat vehicle body. As a result, it causes deformation of vehicle components and usually such vehicle can be eliminated from further actions. That type of firing means is called “humanitarian”, as their main firing factor does not affect directly the combat vehicle crew. It just affects the vehicle structure and internal equipment.

A necessary condition for minimizing the loss in vehicles designed conventional and non-conventional battle fields is to carry out the tests leading to define a level of loads and identification of effects of mine and IED impacts. Theoretical and practical tests should be performed on supporting structures and sensible equipment in almost continuous way and as a result they should be performed for complete vehicles. Easy theoretical tests are as more significant as they are verified by the experimental tests. It was one of the reasons why the experimental tests were performed and their results are presented below.

The main purpose of performed experimental tests in the military training ground conditions was to estimate a level of impact conditions affecting the combat vehicle body and equipment during an explosion of a non-contact anti-bottom mine. It generated fragmentary purposes that can be brought in order to obtain:

- information on the distribution of pressure in a highly limited space (between the vehicle bottom and the ground), among others, in the aspect of verification of mathematical model of post-explosion impact wave dispersion between the vehicle bottom and the ground and the caterpillar mechanism components);
- data on a level of loads affecting the caterpillar combat vehicle body and significant components of internal equipment;
- information on a level of pressure in the crew compartment of a combat vehicle;
- information on the effects of explosion of non-contact anti-bottom mines on the internal vehicle equipment.

A scope of tests included a measurement and registration of the following signals:

- pressure on the vehicle body components, in selected points
- vehicle bottom deformation,
- vertical acceleration of a vehicle body and a driver.

The tests were performed on combat vehicles representative for tank class type of vehicles. A profile of one of those vehicles, used in the tests, is presented on fig. 8. A source of loads included the mines with plastic PMW-8 explosive (weight: m_{MWi}) formed in a semi-spherical way. A shape of explosive and the way of arrangement under the vehicle are presented on fig. 9. The mine was places on the ground at a distance of h_i from the bottom, in the longitudinal axis of a standing vehicle, near the first or between the first and the second supporting wheels.

Fig. 8. Test object profile

Fig. 9. An example of explosive position under the vehicle in its longitudinal axis

Test results

Fig. 10 presents selected moments from the tests. While the fig. 11 and 12 present time courses of pressures affecting the vehicle body components, deformations of its bottom and vertical accelerations during explosion of an explosive of weight of m_{MWi}.

Fig. 13 and 14 present time courses of pressures registered on the surface of the combat vehicle bottom, in points distant from the explosion epicentre by R_1 and R_2 respectively, for two tests, both with an explosive of weight of m_{MW3} ($m_{MW3} > m_{MW1}$).

Presented courses of pressures of the impact wave for the next tests in corresponding measurement points are of very similar nature and very similar peak values. It indicates high recurrence of results.

a)

b)

Fig. 10. Selected moments from experimental tests on the influence of explosives on the combat vehicles

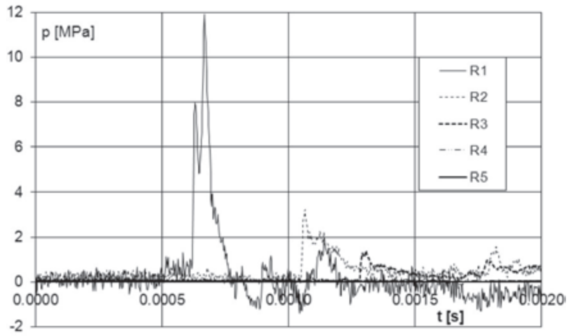


Fig. 11. Pressures in the vehicle body points at a distance of Ri from the source of explosion of an explosive of weight of m_{MW1}, where R1<R2<...<R5

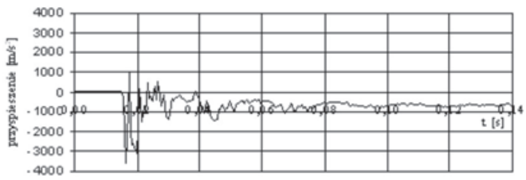
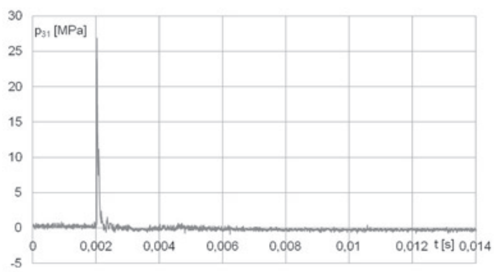
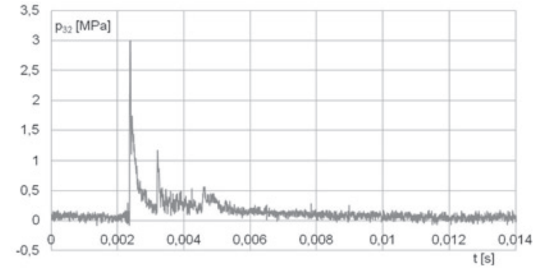


Fig. 12. A course of vertical accelerations of the vehicle body sections during explosion of an explosive of weight of m_{MW1}



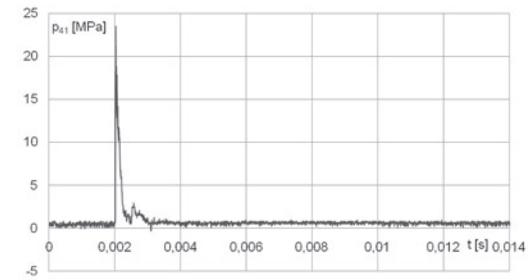
a) R_1

Fig. 13. Courses of pressures in selected points of the bottom for an explosive of m_{MW3}, clearance h, test 3

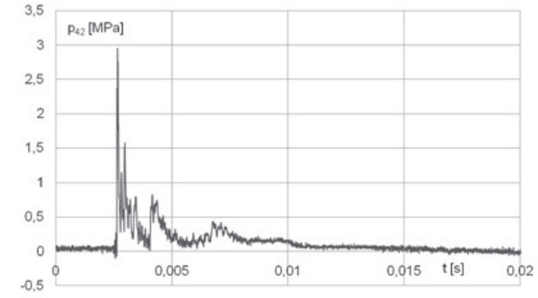


b) R_2

Fig. 13. Courses of pressures in selected points of the bottom for an explosive of m_{MW3}, clearance h, test 3 (continued)



a) R_1



b) R_2

Fig. 14. Courses of pressures in selected points of the bottom for an explosive of m_{MW3}, clearance h, test 4

The influence of such high pressures results in a single or multiple mechanical impacts. Its amplitude can exceed normative values acceptable for values for special equipment or other equipment. The effect of influence of explosives on supporting structures of the combat vehicles and the equipment and the internal equipment depend on many factors and each of them can have a dominating influence as well as an inconsiderable one. However the most important include: the explosive weight and its distance to the vehicle body. Other factors as explosive type, vehicle clearance, location towards the supporting structure, chassis structure can be of secondary significance.

2.3. Tests carried out when firing the main armament

When firing the gun (fig. 15), the load level depends on, among others, the following factors: the gun calibre, missile type.

The gun recoil force affects through the resistance-returners on the vehicle tower and internal equipment installed in it (such as viewfinder, rangefinder, ballistic converter, day observation and night vision instruments, communication equipment and other equipment). Results obtained from the experimental tests [16, 17, 18] – examples of courses of accelerations affecting the gun, vehicle tower components and centre of mass of the vehicle are presented on fig. 16 and 17. Courses presented on fig. 16 refer to firing an explosive projectile, while the courses on fig. 17 refer to firing a kinetic energy penetrator.

Fig. 15. Firing the gun in perpendicular direction towards the longitudinal axis

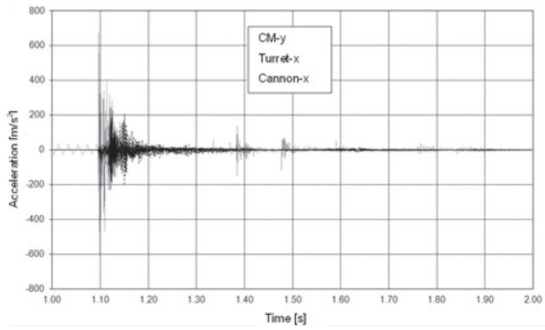


Fig. 16. Courses of accelerations affecting the vehicle tower when firing the gun with explosive projectile

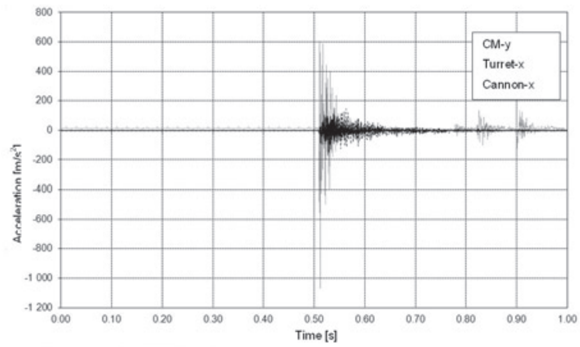


Fig. 17. Courses of accelerations affecting the vehicle tower when firing the gun with kinetic energy penetrator

Significant values of accelerations affecting the aforementioned elements, as a result on the internal equipment, can be noticed. The maximum values of the longitudinal accelerations of the gun amount to a level close to 2500 m/s^2, of the vehicle tower to over 700 m/s^2 the centre of mass to over 800 m/s^2, which is located below the axis of the gun trunnion. It should be underlined that the value of impact affecting the components of the vehicle body and tower depends not only on a type of missile but also on a technical condition and quality of resistance-returners.

3. Model tests

Usually, elimination from further operations and often destruction make the effect of missile acting on a vehicle. Model tests are one of the methods of shaping the resistance of the supporting structures and internal equipment. They are executed on the models that fully reflect the real vehicles. Models can be used many times and analysed in multiple variants until obtaining a supporting structure that fully meets the resistance requirement. The variants that are the most frequent in practice were chosen from various possible ones. Variants of missile impact on the tank tower or a front section of the tank body were analysed. The missile does not cause the piercing of the armour but gets stuck in its material or rebounds. Then the whole or a significant part of the missile energy in a dynamic (impulse) way is transferred directly or indirectly to the vehicle chassis. As a result, the impulse affects the internal equipment. The vehicle load level depends on a gun calibre, barrel length, missile type and initial velocity, impact angle, impact location, position assumed for structure node analysis etc. Courses of longitudinal accelerations, presented on fig. 18, refer to a case of hitting a tower of a tank, weight of 40 000 kg, with a rebounding missile fired from a gun of 120 mm calibre.

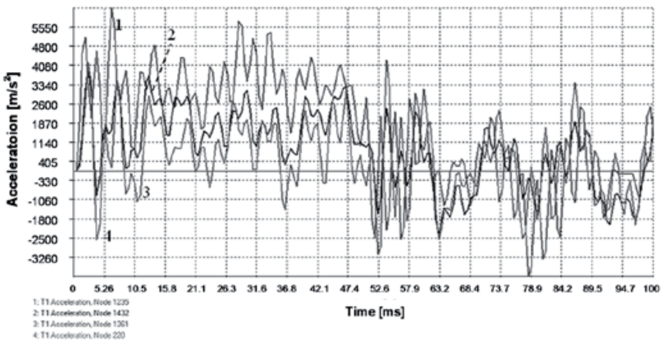


Fig. 18. Longitudinal accelerations in selected points of the tank body after a hit of a non-piercing missile: 1 – tower centre of mass, 2 – tank body centre of mass, 3 – a driver, 4 – floor under a driver's seat

Fig. 19. Caterpillar combat vehicle during rides in complex operating conditions

When using caterpillar combat vehicles, sometimes in unexpected conditions, the load can be more complex, as partially presented on fig. 19 – firing from an own gun when overcoming a counterscarp.

That type of load can consist of forces from the road roughness, firing an own gun, impact of the enemy's missile, explosion of antitank mine or improvised explosive.

4. Final conclusions

Dynamic loads affecting special equipment, as shown, are characterized by a high variability both in relation to a value as well as to direction of impact (vertical, longitudinal and transverse).

Sometimes the maximum values of mechanical impacts affecting special equipment of the combat vehicles significantly exceed the level of normative values and they can be even higher in the future. It is a consequence of permanent competition between a missile and armour.

Installation of special equipment fixed to the supporting structure of combat vehicles forces designers and manufacturers of the aforementioned equipment to imply solutions increasing their impact resistance and low sensitivity to forcing values, direction of action and frequency band.

References

1. Borkowski W, Rybak P. Modelling of impact strength on combat vehicles. Journal of Kones Powertrain and Transport 2007; 1: 121-130.
2. Borkowski W, Rybak P, Michałowski B, Hryciów Z. Badania odporności udarowej kadłuba pojazdu specjalnego. Teka Komisji Motoryzacji PAN o/Kraków 2008: 29-36.
3. Borkowski W, Rybak P. Dynamic load in operation of high-speed tracked vehicles. Journal of Kones Powertrain and Transport 2009; 4(16): 17-22.
4. Burdziński Z. Teoria ruchu pojazdu gąsienicowego. Warszawa: WKiŁ, 1972.
5. Carlucci DE, Jacobson SS. Ballistic. Theory and design of guns and ammunition. CRC Press © 2008 by Taylor & Francis Group, LLC, 2008.
6. Jamroziak K, Kosobudzki M, Ptak J. Assessment of the comfort of passenger transport in special purpose vehicles. Eksploatacja i niezawodnosc – Maintenance and Reliability 2013; 1: 25-30.
7. Koszycki T, Kraszewski E. Konstrukcja i obliczanie czołgu cz. I. Warszawa: WAT, 1970.
8. Larcher M. Simulation of the effects of air blast wave. Luxemburg Office of Official Publications of the European Communities 2007.
9. Larsen MD, Jorgensen K C. Landmine protection of armoured personnel carrier M113, 6th European Ls-Dyna Users Conference, Gothenburg 2007.
10. Ministerstwo Obrony Narodowej. Normy Obronne NO-06-A103 i NO-06-A107. Warszawa: MON, 2005.
11. Prochowski L, Rybak P. Analiza wpływu napięcia gąsienicy na obciążenia dynamiczne załogi czołgu. Biuletyn WAT 1992; 11: 55-73.
12. RTO Technical Report TR-HFM-090. Test Methodology for Protection of Vehicle Occupants against Anti-Vehicular. Landmine Effects. France: RTO/NATO, 2007.
13. Rust M. Passive Protection Concept. 33 Protection Technologies, 2009.
14. Rybak P, Badania eksperymentalne oddziaływań niekontaktowych min przeciwdennych na strukturę nośną pojazdu gąsienicowego. Archiwum Motoryzacji 1997; 3-4: 117-129.

15. Rybak P, Papliński K. Obciążenia udarowe kadłuba wozu bojowego a granice odporności człowieka. Journal of Kones Powertrain and Transport 2007; 1: 331-342.
16. Rybak P, Borkowski W, Wysocki J, Hryciów Z, Michałowski B. Badania modelowe lekkiego czołgu na bazie wielozadaniowej platformy bojowej. Biuletyn Naukowo-Techniczny Ośrodka Badawczo-Rozwojowego Urządzeń Mechanicznych –,,OBRUM" Szybkobieżne Pojazdy Gąsienicowe 2011; 2: 39-58.
17. Rybak P, Borkowski W, Wysocki J, Hryciów Z, Michałowski B. Badania eksperymentalne lekkiego czołgu na bazie wielozadaniowej platformy bojowej. Biuletyn Naukowo-Techniczny Ośrodka Badawczo-Rozwojowego Urządzeń Mechanicznych –,,OBRUM" Szybkobieżne Pojazdy Gąsienicowe 2011; 2 : 59-60.
18. Rybak P, Borkowski W, Hryciów Z, Michałowski B, Wysocki J. Czołg lekki na bazie wielozadaniowej platformy bojowej" Anders". Praca zbiorowa pod redakcją A. Najgebauera ,,Technologie podwójnego zastosowania" – wybrane technologie Wojskowej Akademii Technicznej. Warszawa 2012; WAT: 467-477.
19. Wasserman E. Zasady projektowania broni artyleryjskiej. Warszawa: WAT, 1968.

Piotr RYBAK
Military University Of Technology
Faculty of Mechanical Engineering
Institute of Motor Vehicles and Transport
ul. Gen. S. Kaliskiego 2, 00-908 Warsaw, Poland
E-mail: prybak@wat.edu.pl

Article citation info:
YANG Y-J, HUANG H-Z, 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.

Yuan-Jian YANG
Hong-Zhong HUANG
Yu LIU
Shun-Peng ZHU
Weiwen PENG

RELIABILITY ANALYSIS OF ELECTROHYDRAULIC SERVO VALVE SUFFERING COMMON CAUSE FAILURES

ANALIZA NIEZAWODNOŚCI SERWOZAWORU ELEKTROHYDRAULICZNEGO NARAŻONEGO NA USZKODZENIA SPOWODOWANE WSPÓLNĄ PRZYCZYNĄ

The electrohydraulic servo valve (EHSV) is widely used in many engineering fields. Its reliability is of great importance to the reliability and safety of entire servo control systems. With the aim of analyzing and evaluating reliability of EHSV, this paper firstly presents the physical structure and functional principle of EHSV. It is followed by the Failure Mode, Effects and Criticality Analysis (FMECA). From the analysis, the common cause failures (CCF) in the studied EHSV are identified. Lastly, a method that can quantitatively analyze reliability and failure rate of EHSV with considering the common cause failures is proposed. It is observed from the study that the failure rate of the EHSV with CCF is lower than the failure rate without considering CCF.

Keywords: *EHSV, common cause failures (CCF), FMECA, β-factor model.*

Serwozawory elektrohydrauliczne (EHSV) mają szerokie zastosowanie w wielu dziedzinach inżynierii. Ich niezawodność ma decydujące znaczenie dla niezawodności i bezpieczeństwa całych układów sterowania serwomechanizmami. W celu analizy i oceny niezawodności zaworów EHSV, w pracy przedstawiono najpierw ich budowę fizyczną i zasadę działania. Następnie przeprowadzono analizę przyczyn, skutków i krytyczności uszkodzeń (FMECA). Na podstawie tej analizy określono uszkodzenia zaworu EHSV spowodowane wspólną przyczyną (CCF). Wreszcie, zaproponowano metodę, za pomocą której można ilościowo analizować niezawodność i awaryjność EHSV z uwzględnieniem uszkodzeń spowodowanych wspólną przyczyną. Badania wykazały, że awaryjność EHSV przy uwzględnieniu CCF jest niższa niż w wypadku nieuwzględnienia CCF.

Słowa kluczowe: *serwozawór elektrohydrauliczny (EHSV), uszkodzenia spowodowane wspólną przyczyną (CCF), analiza przyczyn, skutków i krytyczności uszkodzeń (FMECA), model współczynnika β.*

1. Introduction

The electrohydraulic servo valve (EHSV) is a core component of servo control systems. Due to its advantages such as high level of control precision, quick response, light weight, small volume and high immunity to load variations, EHSV has been applied in many fields, such as astronavigation, aviation, navigation, and military equipment [5, 6, 13]. At the same time, EHSV is one of the most failure prone components, and has a direct and significant impact on the performance and reliability of the entire servo control system. Thus, it is very important to analyze the failure mode, failure effects, failure mechanism and failure rate of EHSV.

The Failure Mode, Effects and Criticality Analysis (FMECA) is effective for reliability analysis and has been used in many products. It can be applied in different life stages of product to find the defects and weak components and to provide basis information for the further reliability analysis [6, 18].

In most studies, it is assumed that failures of different components are independent random events. Such assumption is valid in most cases for electronic devices, but invalid for mechanic products. Common causes failures are multiple failures which are exist at the same time and are a direct result of a shared root cause [20]. Considering that common cause failure would result in a more reliable and accurate analysis [15].

2. Working principle of EHSV

There must be a bridge component in combination of electric and hydraulic device. This interface connection in servo control system is achieved by electrohydraulic servo valve. Such servo valve converts low power electrical signals into motion of a valve which in turn controls the flow and pressure to a hydraulic actuator [16].

The two-stage nozzle flapper electrohydraulic servo valve is the most widely used one[5], as shown in Fig1. Thus, this paper takes it as an example to introduce the working principle of EHSV and conducts the reliability analysis in the ensuing sections. The torque motor, consisting of permanent magnet, armature, spring pipe and feedback rod, is used as electric-mechanic transducer. The nozzle flapper is the first stage hydraulic amplifier, and the spool valve is the second stage hydraulic amplifier.

A servo valve has a hydraulic pressure inlet and an electrical input for the torque motor. The input current controls the flapper position. A small flapper motion creates an imbalanced pressure in one direction or the other on the ends of the spool of the second stage. Obviously the spool will tend to move in response to this imbalance and allow flow Q_L to the actuator. Since continued imbalance in pressure would quickly move the spool to its limits of travel, a form of feedback connects the motion of the spool to the effective displacement of the flap-

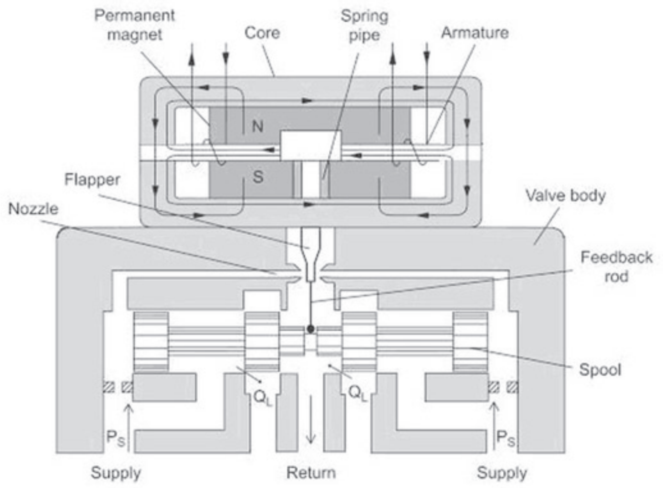


Fig. 1. The schema of EHSV

per. A very small spool displacement will result in a large flow at high pressures typically used.

From reliability engineering point of view, the studied electrohydraulic servo is a series system mainly consisting of the torque motor, the nozzle flapper amplifier, and the spool valve amplifier. The reliability block diagram of the EHSV is shown in Fig 2.

Fig. 2. The reliability block diagram of EHSV

3. FMECA for EHSV

Every product or system has failure modes. It is extremely significant to provide designers or operators with safety assessment methods that help to minimize the adverse effects of failures. Failure mode, effects and criticality analysis (FMECA) is one of the most established and powerful methods for identifying and evaluating system failure. As an engineering tool, it has a fundamental role in any safety or reliability study [19].

The purpose of the FMECA is to provide a systematic, critical examination of potential failure modes of equipment and their causes, to estimate the reliability of systems, to analyze the effect (the consequence of a failure mode) of each failure mode on a system, and to identify corrective actions, i.e., design modifications [21]. FMECA is a bottom-up, inductive analysis method which starts from the lowest Indenture level, and it permits to analyze a system in order to identify potential failure modes, their cause and effect on performance and, when applicable, their effect on the safety of humans, on environment and on the system [4]. FMECA extends FMEA by including a criticality analysis which is used to quantify failure effects and severity.

Risk Priority Number (RPN), a quantitative index, is used to analysis the risk associated with potential problems identified during the failure mode and effects analysis, and to rank the failure modes and effect in the criticality analysis. The calculation of the RPN is based on severity (S), occurrence (O) and detection (D) [3] as follows:

$$RPN = S \cdot O \cdot D$$

Severity quantifies the likelihood of the strength of a failure mode impacts on the system. Occurrence represents the probability that a failure mode will occur. Detection is the estimate of possibility of detecting before it reaches end-users or customers.

In this study, data of a certain type of EHSV have been collected form 6 experts from the research institute where electrohydraulic servo valve has been widely used. On the basis of collected data, FMECA and the evaluation criteria of indices are applied. Fig.3 depicts the framework for FMECA of EHSV.

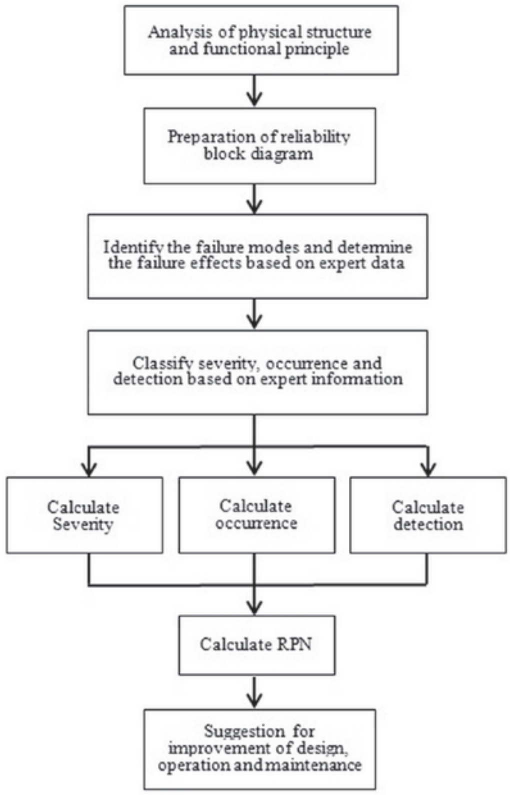


Fig. 3. The framework for FMECA of EHSV

For each of these indices (S, O, D) in the critical analysis, a detailed analysis is needed to identify their appropriate values, related to the type of application and environment. Table 1 shows the evaluation criteria of the three indices in this study. S, O and D are defined in the range of 1 to 4, so the value of RPN is 1 to 64.

Table 1. The evaluation ceriteria for severity, occurrence, and detection

Score	Severity	Occurrence	Detection
1	Insignificant: A negligible effect	Failure is unlikely	High: Can be detected by operator
2	Minor: A minimal effect	Relatively few failures	Moderate: Can be detected by regular detection
3	Critical: A great effect	Occasional failures	Low: Hard to detect, usually need disassembly
4	Catastrophic: causes system failure	Repeated failure	Non-detection: Impossible to detect

According the reliability block diagram of EHSV, as shown in Fig.2, the torque motor, the nozzle flapper amplifier, and the spool valve amplifier are defined as the lowest indenture level. Table 2 shows a part of results of the FMECA analysis. As mentioned earlier, the RPN was used to rank all the failure modes. Furthermore, consid-

Table 2. A part of the FMECA for the EHSV

Lowest Indenture Level	Failure Mode	Failure Cause	Failure Effect	Severity (S)	Occurrence (O)	Detection (D)	RPN
Torque Motor	Coil breakage	Overload or Wear	system failure	4	1	4	16
	Ball end wear	Wear	instability and degradation in performance	2	2	3	12
	Spring pipe fatigue	Fatigue	System failure	4	1	3	12
	Feedback rod bending	Wear	zero deviation increase	3	1	3	9
Nozzle Flapper Amplifier	Nozzle clogging	Oil contamination	zero deviation increase	3	3	2	18
	Orifice clogging	Oil contamination	zero deviation increase	3	3	2	18
Spool Valve Amplifier	Valve core wear	Wear	leakage and degradation in perfomance	3	1	2	6
	Jam fault of spool valve	Oil contamination or Wear	system failure	4	2	2	16

ering the failures modes, effects, based on the RPN values, the plans of improvement and maintenance will be discussed.

4. CCF analysis

4.1. Definition of CCF

Common cause failure is a specific type of dependent failure. The evidence that dependent failures are significant was presented by G. T. Edwards and I. A. Watson in their study [7], and they demanded that the design and operation of some systems must include a concerted approach against the dependent failures. Dependent failures include all definitions of failures that are not independent, encompass common cause failures and cascade failures.

A set of definitions of dependent failure (DF), common cause failure (CCF), cascade failures (CF) and common mode failure (CMF) are given as follows, and Fig.4 shows the relationship between them [1, 2, 8 ,11, 12, 14].

- Dependent failure (DF): The failure of a set of events, the probability of which cannot be expressed as the simple product of the unconditional failure probabilities of the individual events.
- Common cause failure (CCF): This is a specific type of dependent failure where simultaneous (or near-simultaneous) multiple failures result from a single shared cause.
- Cascade failures (CF): These are all dependent failures that do not share a common cause, and they propagate failures.
- Common mode failure (CMF): This term is reserved for common-cause failures in which multiple equipment items fail in the same mode.

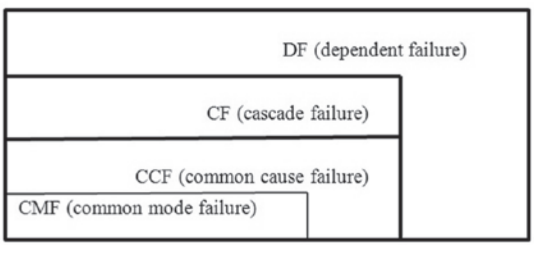


Fig. 4. The relationship between DF, CF,CCF and CMF

The root causes of CCF could be:

- The same design, manufacturer or assembly technology
- The same environmental conditions
- The same personnel dealing with the operation or maintenance or installation and constructions
- A human error

4.2. An overview of methodology for quantitative evaluation of CCF

Common methods for evaluation of common cause failures include:

- Basic parameter model
- Beta factor method
- Multiple Greek letter method
- Alpha factor method

(1) Basic parameter model

The basic parameter model refers to the straightforward definition of the probabilities of the basic failure events. The symmetry assumption is the probability of failure of any given basic event within a common cause component group depends only on the number and not on the specific components in that basic event. The total probability of failure for a component Pc in a common cause group of m components is:

$$P_c = \sum_{k=1}^{m} \binom{m-1}{k-1} \cdot P_k \tag{1}$$

where $\binom{m-1}{k-1} = \frac{(m-1)!}{(k-1)!(m-k)!}$.

Pk is probability of a basic event involving k specific components, $1 \le k \le m$. Ideally, the values can be calculated from data, but unfortunately the complete data is normally not available. Other models putting less stringent requirements on the data have been developed [2].

(2) Beta factor method

The beta factor method was introduced in 1974 by Fleming [9]. The beta factor method assumes that Pc, which is the total probability of failure for a component, can be expanded into an independent failure contribution P_{IF} and a common cause failure contribution P_{CCF} :

$$P_c = P_{IF} + P_{CCF} \tag{2}$$

A parameter β is defined as the fraction of total failure rate attributable to dependent failure:

$$\begin{aligned} &\beta = \frac{P_{CCF}}{P_c} = \frac{P_{CCF}}{P_{IF} + P_{CCF}} \\ &\Rightarrow P_{CCF} = \beta \cdot P_c \\ &\Rightarrow P_{IF} = (1-\beta) \cdot P_c \end{aligned} \tag{3}$$

The strength of the β factor model lies on data including historical data collected from both experiment and field. If β factor is not known, a general value of 0.1 can be used.

The beta factor method is the least demanding among the above methods and is used in this study. Because it only requires the estimation of common cause parameter in addition to the independent fail rate to the model the total component failure rate.

(3) Multiple Greek letter method

The multiple Greek letter parameters consist of the total failure probability and a set of failure fractions. The failure probability includes the effects of all independent and common cause contributions to that component failure, whereas the failure fractions are used to quantify the conditional probabilities of all the possible ways which a common cause failure of a component can be shared with other components in the same group, given the condition that the component has failed [10].

The following equation $m-1$ with parameters ($\rho 2, \rho 3, \ldots, \rho k$) is the general expression for the multiple Greek letter method, m is common cause group size:

$$P_k = \frac{1}{\binom{m-1}{k-1}} \cdot \left(\prod_{i=1}^{k} \rho_i \right) \cdot (1 - \rho_{k+1}) \cdot P_c \tag{4}$$

$\rho k+1$ is the conditional probability of the failure of at least one additional component, given that k components has failed, $\rho k=1$, $\rho k+1=0$, $k=1,\ldots, m$.

(4) Alpha factor method

The general expression for the alpha factor method is the following [17]:

$$P_k = \frac{k}{\binom{m-1}{k-1}} \cdot \frac{\alpha_k}{\alpha_t} \cdot P_c \tag{5}$$

where

$$\alpha_t = \sum_{k=1}^{m} k \cdot \alpha_k$$

$$k = 1, \ldots, m$$

α_k is the fraction of the total failure probability of events that occur in the system and involve the failure of k components because of a common cause.

4.3. CCF Analysis for EHSV

If R_{EHSV} is the reliability of the EHSV, λ_{EHSV} is the failure rate of the EHSV, R_T is the reliability of the torque motor, λ_T is the failure rate of the torque motor, R_S is the reliability of the spool valve amplifier , λ_s is the failure rate of the spool valve amplifier , R_N is the reliability of the nozzle flapper amplifier, λ_N is the failure rate of the nozzle flapper amplifier.

If the failure events of the three components are independent, R_{EHSV} is calculated based on the reliability model of series systems as follows:

$$R_{EHSV} = R_T \cdot R_S \cdot R_N \tag{6}$$

When it is assumed that all the failure of the three main components obey the exponent distribution, λ_{EHSV} is calculated as:

$$\lambda_{EHSV} = \lambda_T + \lambda_S + \lambda_N \tag{7}$$

It assumes that:

$$\lambda_T = \lambda_S = \lambda_N = \lambda \tag{8}$$

The failure rate λ_{EHSV} can be given by:

$$\lambda_{EHSV} = \lambda_T + \lambda_S + \lambda_N = 3\lambda \tag{9}$$

However, according to the FMECA results, as shown in Table2, there is a common cause failure of nozzle flapper amplifier and nozzle flapper amplifier due to oil contamination. At the same time, there is another common cause failure for all the components of EHSV which is ignored in the FMECA but really exists due to the same design, manufacturer, assembly, environmental conditions and hours of use. The fault tree with considering the two common causes is shown in Fig. 5.

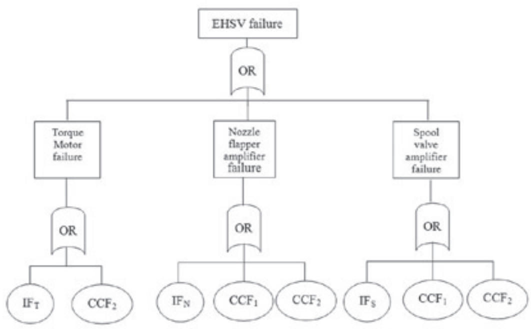


Fig. 5. The fault tree of EHSV with CCF

IF_T: Independent failure of torque motor, and the failure rate of IFT is λ_{IF}^{T};

IF_N: Independent failure of nozzle flapper amplifier, and the failure rate of IFN is λ_{IF}^{N};

IF_S: Independent failure of spool valve amplifier, and the failure rate of IFS is λ_{IF}^{S};

CCF_1:Common failure of nozzle flapper amplifier and spool valve amplifier due to oil contamination, and the failure rate of CCF_1 is λ_{CCF1};

CCF_2:Common failure of torque motor, nozzle flapper amplifier and spool valve amplifier due to same design, manufacturer, assembly, environmental conditions and the hours of use, and the failure rate of CCF_2 is λ_{CCF2};

The minimal cut set of the above fault tree is
{ IFT, IFN, IFS, CCF1,CCF2 }

It is assumed that all the failure rates are constant. The failure rate of EHSV and each component with common cause failures can be calculated by:

$$\lambda'_{EHSV} = \lambda_{IF}^{T} + \lambda_{IF}^{N} + \lambda_{IF}^{S} + \lambda_{CCF1} + \lambda_{CCF2} \quad (10)$$

$$\lambda_{T} = \lambda_{IF}^{T} + \lambda_{CCF2} \quad (11)$$

$$\lambda_{N} = \lambda_{IF}^{N} + \lambda_{CCF1} + \lambda_{CCF2} \quad (12)$$

$$\lambda_{S} = \lambda_{IF}^{S} + \lambda_{CCF1} + \lambda_{CCF2} \quad (13)$$

In this study, β_1 is defined for CCF1, and β_2 is defined for CCF2 based on the beta method, it also assumes $\lambda_T = \lambda_S = \lambda_N = \lambda$ as Eq.(8).

We can get:

$$\lambda_{CCF1} = \beta_1 \cdot \lambda$$

$$\lambda_{CCF2} = \beta_2 \cdot \lambda$$

$$\lambda_{IF}^{T} = (1-\beta_2)\lambda$$

$$\lambda_{IF}^{N} = (1-\beta_1-\beta_2)\lambda$$

$$\lambda_{IF}^{S} = (1-\beta_1-\beta_2)\lambda$$

The failure rate λ_{EHSV} can be given as:

$$\lambda'_{EHSV} = \lambda_{IF}^{T} + \lambda_{IF}^{N} + \lambda_{IF}^{S} + \lambda_{CCF1} + \lambda_{CCF2} = (3-\beta_1-2\beta_2)\lambda \quad (14)$$

In order to make it easy to compare Eq. (7) with Eq. (14), we assume $\lambda = 0.001, \beta_1 = 0.3, \beta_2 = 0.6$. According Eq. (7) and Eq. (14), we can get $\lambda_{EHSV} = 0.003$, $\lambda'_{EHSV} = 0.0015$. The calculation of the failure rate of the EHSV with CCF is lower than the failure rate without considering CCF. This means that the assumption that components of a series system fails independently tends to underestimate the reliability of system.

5. Conclusion

The major works of this paper contains three aspects. First, the structure and working principle of the two-stage nozzle flapper electrohydraulic servo valve are analyzed. Second, the method of failure mode, effects and criticality analysis is chosen to conduct the reliability analysis for EHSV, and rank the main failure modes. Third, in the criticality part, the beta method is used to calculate the failure rate of EHSV with common failures. The comparison shows that failure rate of EHSV with CCF is lower than the failure rate without CCF.

The assumption that all the failures of EHSV obey the exponent distribution makes it easy to compare the calculations with or without common cause failure. Nevertheless, the assumption may not reasonable, and the failure rates of the components of ESHV are not constant in fact. In order to evaluate the reliability of ESHV, the lifetime models of components of EHSV will be analyzed by fusing accelerate life testing data, accelerate degradation testing data and field information in future work.

Acknowledgments

This research was partially supported by the National Natural Science Foundation of China under the contract number 11272082.

References

1. Borcsok J, Holub P. Different approaches for probability of common cause failure on demand calculations for safety integrity systems. Proceedings of International Conference on Computer Systems and Applications 2008; 1052-1057.
2. Borcsok J, Schaefer S, Ugljesa E. Estimation and evaluation of common cause failures. Proceedings of Second International Conference on Systems 2007; 41-49.
3. Bowles JB, Pelaez CE. Fuzz logic prioritization of failures in a system failure mode, effects and critically analysis. Reliability Engineering and System Safety 1995; 50(2): 203-213.
4. Catelani M. FMECA technique on photovoltaic module. Proceedings of Instrumentation and Measurement Technology Conference, 2011; 1-6.
5. Chen B, Lin ML. The present research and prospects of electro-hydraulic servo vale. Chinese Hydraulics & Pneumatics 2005; 33(6): 5-8.
6. Dai C G, Wang X H, Zhang X. Fuzzy comprehensive evaluation in FMEAC of electro-hydraulic servo valve. Journal of Beijing University of Aeronautics and Astronautics 2011; 37(12): 1575-1578.
7. Edwards GT, Watson IA. A study of common mode failures SRD-R-146. United Kingdom Atomic Energy Authority, Safety and Reliability Directorate, 1979.
8. EPRI NP-3837. A Study of Common Cause Failures. Phase 2: A Comprehensive classification System for component fault analysis. Los Alamos Technical Associates Inc, 1985.
9. Fleming KN. A reliability model for common mode failures in redundant safety systems. General Atomic Report, GA-13284, 1974.
10. Fleming KN, Mosles A, Deremer RK. A systematic procedure for the incorporation of common cause events into risk and reliability models. Nuclear Engineering and Design 1986; 93(5): 245-273.
11. Humphreyes P, Johnston B D. Dependent Failure Procedure Guide SRD-R-418. United Kingdom Atomic Energy Authority, Safety and Reliability Directorate, 1987.
12. Johnston BD, Crackett J. Common Cause Failure Reliability Benchmark exercise. SRD-R-383. United Kingdom Atomic Energy Authority, Safety and Reliability Directorate, 1985.
13. Li T, Yang SL. Common fault analysis of electro-hydralic press servo valve on certain plane. Chinese Aviation Maintenance & Engineering 2011; 29(2): 73-75.
14. Mauri G. Integrating Safety Analysis Techniques, Supporting Identification of Common Cause Failures. Doctor of Philosophy thesis, University of York, Department of Computer Science, 2000.

15. Mi J, Li Y, Huang H-Z, Liu Y, Zhang X-L. Reliability analysis of multi-state system with common cause failure based on Bayesian networks. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2013; 15(2): 169-175.
16. Mondal N, Datta BN. A study on electro hydraulic servovalve controlled by a two spool valve. International Journal of Emerging Technology and Advanced Engineering 2013; 23(3): 479-484.
17. Mosles A, Siu NO. A multi-parameter, event-based common cause failure model. Proceedings of the Ninth international Conference on Structural Mechanics in Reactor Technology 1987: 147-152.
18. Rausand M. Reliability centered maintenance. Reliability Engineering and System Safety 1998; 60(2): 121-132.
19. U.S.A. Depatment of Defense. MIL-STD-1629 Procedures for Perfoming a Failre Mode, Effects and Criticality Analysis. 1980.
20. Vaurio JK, An implicit method for incorporating common-causefailures in system analysis. IEEE Transactions on Reliability 1998; 47(2): 173-180.
21. Wei BC. A unified approach to failure mode, effects and criticality analysis (FMECA). Proceedings of Annual Reliability and Maintainability Symposium 1991; 260-271.

Yuan-Jian YANG
Hong-Zhong HUANG
Yu LIU
Shun-Peng ZHU
Weiwen PENG

School of Mechanical, Electronic, and Industrial Engineering
University of Electronic Science and Technology of China
No. 2006, Xiyuan Avenue, West Hi-Tech Zone, Chengdu
Sichuan, 611731
P. R. China
E-mail: hzhuang@uestc.edu.cn

Article citation info:
PROCHOWSKI L, ŻUCHOWSKI A. Analysis of the influence of passenger position in a car on a risk of injuries during a car accident. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 360–366.

Leon PROCHOWSKI
Andrzej ŻUCHOWSKI

ANALYSIS OF THE INFLUENCE OF PASSENGER POSITION IN A CAR ON A RISK OF INJURIES DURING A CAR ACCIDENT

ANALIZA WPŁYWU POZYCJI CZŁOWIEKA W SAMOCHODZIE NA RYZYKO OBRAŻEŃ W CZASIE WYPADKU DROGOWEGO*

Experimental tests have been prepared and carried out in order to deepen the analysis of the influence of a position of a rear seat passenger in a passenger car on a risk of injuries during a road accident. This risk is considered and calculated in the aspect of a car passenger position that slightly deviates from the one planned by manufacturers of individual protection devices. Attention was focused on the analysis of measurable effects of the position modification in the area of torso and head movement and neck deformation. The experimental tests included a physical simulation of a frontal collision of a car and a rigid obstacle. They showed that relatively small changes to the initial dummy position could be a reason of significant differences in the movement trajectory and dummy position at the final stage of the crash test. These differences are described in details. The analysis also included the criteria index values that refer to the head, neck and chest injuries. Index calculation results confirm the influence of a small change to the leg and torso position on the index values, thus on the probability of injuries of the rear seat passengers in a passenger car.

Keywords: *road transport, crash tests, safety of passengers.*

Przygotowano i przeprowadzono badania eksperymentalne, które mają na celu pogłębienie analizy wpływu pozycji osoby siedzącej na tylnym siedzeniu w samochodzie osobowym na ryzyko jej obrażeń w czasie wypadku drogowego. Ryzyko to jest rozważane i obliczone w aspekcie zajmowania przez człowieka w samochodzie pozycji nieznacznie odbiegającej od zaplanowanej przez producentów urządzeń ochrony indywidualnej. Uwagę skupiono na analizie wymiernych skutków zmiany tej pozycji w obszarze ruchu torsu i głowy oraz odkształcenia szyi. Badania eksperymentalne były symulacją fizyczną czołowego uderzenia samochodu w sztywną przeszkodę. Pokazały one, że stosunkowo niewielkie różnice w początkowej pozycji manekina mogą być przyczyną znacznych różnic w trajektorii ruchu oraz położeniu manekina w kulminacyjnej fazie testu zderzeniowego. Różnice te zostały szczegółowo opisane. Analizie poddano także wartości wskaźników kryterialnych, które dotyczą powstawania obrażeń głowy, szyi i klatki piersiowej. Wyniki obliczeń wskaźników potwierdzają wpływ już niewielkiej zmiany położenia nóg i torsu na ich wartości, a zatem na prawdopodobieństwo powstawania obrażeń u osób jadących na tylnych siedzeniach samochodu osobowego.

Słowa kluczowe: *transport drogowy, testy zderzeniowe, bezpieczeństwo pasażerów.*

1. Introduction

The objective of improvement of the means of transport is to increase people mobility, including a possibility of driving at higher and higher speeds. At the same time, threats related to the use of that means become more and more serious [6, 11, 13, 16, 18, 20]:

- constant excessive number of accidents and victims;
- the majority of people occupy a position in a car that is unexpected by manufacturers of passive safety equipment;
- dynamic loads of passengers in the rear rows of seats are often several times higher than in case of the front seat passengers.

These problems line out the area of considerations in this paper and refer to the issue of safe vehicle usage, presenting the example with passenger cars. Using a vehicle should be combined with awareness of effects of improper use of passive safety devices (safety belt, seat). For example, a passenger position, unexpected by a manufacturer of safety devices, i.e. the so-called *Out-of Position* (OoP), can create possible threats for passengers. Studying these threats and reasons of their occurrence when using a car make a subject of studies and analyses in this paper. The analysis of reasons of threats related to a position occupied in a car has a significant social aspect, including a scope of driver training.

The papers [4, 18, 19, 20] present a comparative evaluation of loads that affect the passengers in the front and rear seats of a car during a head-on collision of a car and an obstacle. It was found there that dynamic loads of people in the rear seats are often several times higher than the front seat passengers. These differences result from occupied positions, available space and efficiency of applied individual protection devices.

In contemporary passenger cars, rear seat passenger protection equipment might not be able to provide the same safety level as for the front seat passengers. Relations in injuries of adult passengers in the first and the second row of seats become reverse to the ones that were characteristic for cars 20–30 years ago [5, 7, 12, 16].

The authors of this paper [15] have carried out the analysis of the influence of a rear seat passenger's position on the motion trajectory and dynamic loads during a head-on collision of a car and an obstacle. Tests were performed and measurements were made in order to define a course and dynamic load values, resulting from the action of the force of inertia and reaction of the seat belts.

A dummy was placed in the rear seat of a passenger car. It had a different initial position in the following tests. Detailed specification of dimensions and distances characterizing these positions are given in [15]. The effects of the initial position modification, observed on

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

the trajectory of the torso and head movement and on the dynamic loads, occurring during a head-on collision of a car and an obstacle, were analysed.

The objective of this paper is to deepen the analysis of the effect of the rear seat passenger's position in a passenger car on a proper operation of the protection equipment. That operation is evaluated on the basis of analysis of the passenger displacement and a possibility of injuries during a road accident. The attention was focused on the torso and head movement analysis. Relation between the torso and the head displacement strongly affects the spine deformation on the neck section. The authors want to answer the question whether small changes in the initial leg and hip positions against the seat have a significant and measurable influence on the operation of protective devices, including the influence on the condition of dynamic loads affecting the passenger's torso, neck and head during a frontal collision of a car and an obstacle.

Fig. 1. A diagram of the initial dummy position and its view in the tests 2P (a), 3P (b) and 11P (c) in time t=0 ms, 40 ms, 80 ms and 120 ms

2. Test preparation and measurement conditions

A Hybrid III size M50 dummy (50-centile male) made the object of the studies. It was placed in the rear seat of a medium class passenger car, including three following positions in the next tests:

– classic one, where a passenger is resting his back on a rear seat back, with the head in a torso line and legs slightly extended forward, accordingly to a space available in a car, i.e. a distance to a preceding seat (test marked as 2P);
– position as above but feet are withdrawn towards a seat, the head and the neck are visibly inclined towards in relation to the torso (test marked as 3P);
– position with hips moved away from the back but feet withdrawn towards the seat (test marked as 11P).

The experimental tests were a physical simulation of a frontal collision of a car and a rigid obstacle. The impact speed amounted to 48 km/h. Time t = 0 s is a beginning of a car contact with an obstacle.

Figure 1 specifies single video frames that were chosen for a several characteristic moments from a course of a collision of a car and an obstacle. They show relatively small differences in the initial dummy position and at the same time clear differences in the dummy positions at the final stage of the crash test (t=120 ms). These differences are described in details further on.

3. Dummy kinetics in the tests 2P, 3P, 11P

The following figures, prepared on the basis of the frame-by-frame analysis of crash test videos, show succeeding dummy positions from the moment of t = 0 to 200 ms. Silhouettes were drawn with 20ms interval in two sets:

– time interval 0-100 ms, i.e. 6 following positions in the forward dummy movement;
– time interval 120-200 ms, which includes 5 subsequent positions in the reverse dummy movement from the maximum inclined position to the position occupied at the final stage of the test.

Figures 2 a, b, c allow for general characteristics of the dummy motion in individual tests, including torso displacement in the dummy movement on a seat and the head movement. Figures clearly illustrate the scope of the head movement at two stages of that motion: forward and reverse. Torso and head displacements in the tests 2P and 3P are similar but not the same. However the dummy movement in the test 11P is definitely different than in two previous tests. Figure 2c shows ineffective operation of the seat belt. As a result there was a movement of hips from a seat to the space between the rear seat and the front seat as well as pushing and deformation of the front seat as a result of the leg pressure.

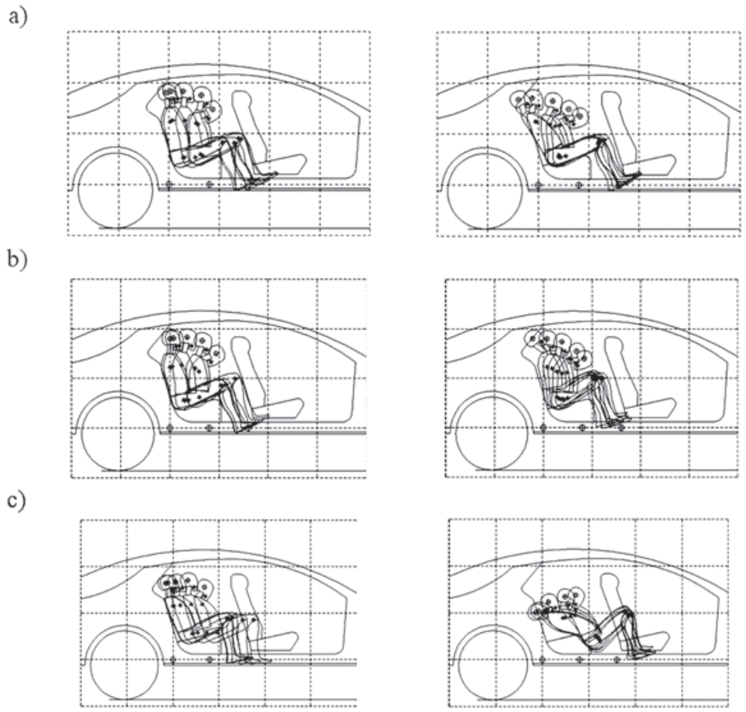


Fig. 2. Dummy displacements against the passenger cabin interior in the tests 2P (a), 3P (b) and 11P (c)

4. Analysis of results in the tests 2P, 3P and 11P

4.1. Analysis of torso movement and affecting force

Analysis of the influence of the dummy position (silhouette) on its kinetics during a road accident was started from identification of leg (thigh) position change and kinematics. That area of analysis refers to clearly visible differences of the initial state that might affect the behaviour of the dummies. Results of that analysis are presented in this paper [15]. They unequivocally indicate that initial leg position and forces resulting from that position affect the dummy's torso movement. The dummy's torso movement is characterized below by showing the longitudinal displacement of its centre and changes of the torso inclination angle (fig. 3 and table 1). Inclination angle change, against the initial position for t = 0 s, characterizes the torso rotation around the centre of the hip joint. The way of determining the change of the torso position angle is shown on fig. 4. This rotation is most of all a result of action of forces of inertia (caused by a delayed movement of the car body after a car hit san obstacle) and reaction in the seat belt. The dummy's position before the test defines the initial conditions of that movement.

On the basis of figure 3, a mutual relation between the components of the torso movement i.e. its longitudinal displacement and inclination angle was evaluated in the following part of this paper. These relations depend on the initial torso position, as it results from the comparison of courses on figures 3 and 5. Excessive range of the torso movement (shift and rotation) results in a high range of head movement, including the area of risk of hitting the car body components.

The figure 5 shows that changes of the torso inclination angle have a distinctly different course in the analysed tests. Positive values of the inclination angle indicate its forward movement from the position at the moment t = 0 s (according to the direction of the pre-impact car speed vector). Negative values indicate that the angle of the torso inclination backwards is higher than its initial value (for t = 0 s) in the analysed test. It was possible when the dummy moved on the seat and its hips moved away from the backrest.

The change of the leg position in the test 3P compared to the test 2P affects the torso angle movement, visible on the specification shown on figure 5. The initial leg position assumed in the test 3P resulted in limitation of its longitudinal shift (knees hitting the front seat [15]). It caused the longitudinal hip movement limitation and further action of the force of inertia in the centre of mass of the torso resulted in a significant increase of the torso inclination angle at the final stage of the crash test. The extreme value of the torso inclination angle in 3P is twice higher than in 2P. The angle movement of the torso resulted in

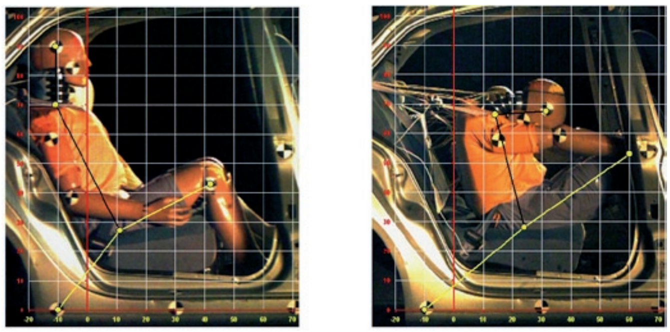


Fig. 4. The way of defining the torso inclination angle using an example from the test 3P (time 0 and 120 ms)

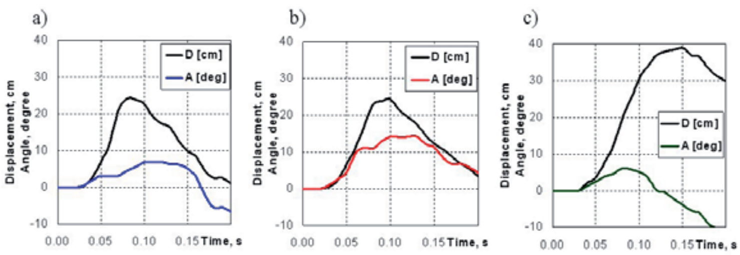


Fig. 3. Longitudinal displacement of the centre of mass of the torso (D) and change of the torso inclination angle (A) in the tests 2P (a), 3P (b) and 11P (c)

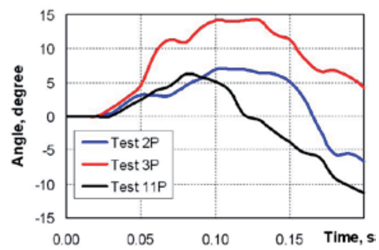


Fig. 5. A process of changes in the torso inclination angle in the time function

Table 1. Specification of characteristic values in the torso movement

Value	Test 2P	Test 3P	Test 11P
Maximum longitudinal displacement of the centre of the torso, m	0, 24	0,24	0,39
Change of the torso inclination angle value in the forward movement, degrees	+7	+14	+6
Change of the torso inclination angle value in the reverse movement, degrees	-13	-10	-17

significant acceleration values and force of inertia in the centre of its mass (fig. 6), that are 15 % higher than the ones occurring in the test 2P and result in injury risk increase.

While the initial hip movement away from the rear seat backrest in the test 11P resulted in torso displacement 60 % higher than in 2P and 3P (table 1). The dummy's hips slid down from the seat (fig. 1 and 3c) under the lap belt, which could not operate properly in case of that initial dummy position. Obviously, the dummy motion, occurred in the test 11P, resulted in high dynamic leg and stomach loads, though the extreme torso acceleration values were significantly lower than in the tests 2P and 3P (fig. 6). The process of sliding down the dummy's hips from the seat results in decrease of the acceleration paste in the centre of the torso mass, as the process of braking its longitudinal motion by the seat belts (at the final stage of the impact) becomes weaker, as it can be seen on figure 6 during 50÷60 ms.

The paper [2] identifies the following location of fatal injuries for the rear seat passengers, during frontal collision: head – 25 %, neck – 8 %, chest – 35 % and belly 30 %. The Hybrid III dummy had no sensors in the belly but significant forces in the lap belt, which slid down from the hips on the soft belly in the test 11P, confirm a high injury risk. So it is another important threat signal, which results from an occupied position in towards the passive safety equipment.

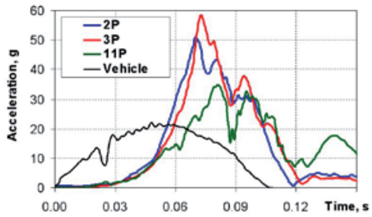


Fig. 6. Resultant acceleration in the centre of the torso

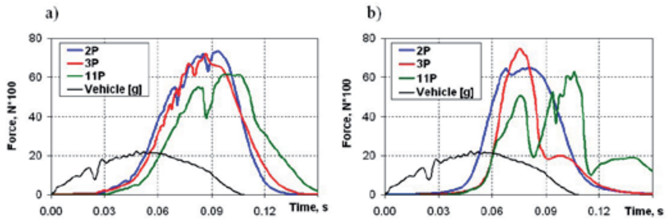


Fig. 7. Comparison of changes in the force, measured on the shoulder seat belt (a) and the lap belt (b)

The courses of resultant torso acceleration in the tests 2P and 3P are similar but the extreme values are about 15 % higher than in the test 3P. It can result from the time accordance (t = 70÷80 ms) of occurrence of the highest torso displacement and rotation in this test (compare fig. 3 and 6). So, despite similar courses of resultant acceleration in the centre of the torso, the change of the leg position angle in the test 3P towards their position in 2P affects the angular motion of the dummy's torso.

The results of the measurement of forces in the shoulder belt (fig. 7a) have similar courses in the tests 2P and 3P. While in the test 11P, the force values in the shoulder belt are clearly smaller than in the previous tests due to the reasons described above. Forces in the lap belt (fig. 7b) have different courses in all three tests. In the test 2P, the lap belt reacts on the dummy clearly earlier than in the tests 3P and 11P. At the final stage of the dummy's movement forward (time 80÷100 ms), the force in the lap belt decreases when the knees hit the front seat. On the basis of other observations, it should be also mentioned that the belt displacement on the grommet (between the lap and the shoulder sections) practically do not occur in these tests and it allows for a separate consideration of reaction of the both seat belt sections.

The resultant torso accelerations in the tests 2P and 11P have various courses (fig. 6). However, due to a high scope of longitudinal dummy displacement, forces of inertia and belt reactions in 11P are lower than in 2P. The maximum force values in the lap and the shoulder sections of the seat belt are compared on figure 8.

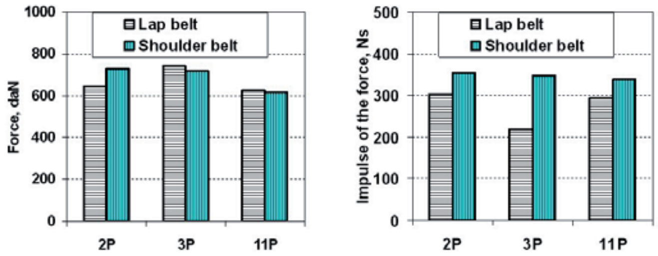


Fig. 8. Maximum values of the force and the impulse of the force acting in the lap and shoulder portions of the seat belt strap (tests 2P, 3P and 11P)

Additionally, the force impulse was calculated for the quantitative evaluation of differences in reaction of the seat belts in individual tests:

$$I = \int_{t=0}^{t_K} F dt \tag{1}$$

where:

F – belt stretching force,

t_K - belt stretching time, F(t)>0.

Figure 8 shows that relations between the maximum force value and the force impulse are different. It confirms the idea of using the force impulse as a supplement of description of the belt load during collision. The force impulse values are clearly smaller for the lap belt than for the shoulder belt, particularly in the test 3P, when the dummy's knees hit the backrest of the front seat, resulting in the belt relief.

4.2. Head movement analysis

Initial differences in the leg and torso positions (classic position 2P and leg position change 3P and hip position change 11P) affected the head movement, and first of all it affected changes of its position towards the torso and the seat backrest. Determination of the head rotation angle changes is based on identification of position of the markers placed on the head in subsequent video frames of the experiment, in a way shown before on figure 4.

On the basis of the frame-by-frame analysis of the test videos, values and a course of changes of longitudinal displacement of the centre of the head mass and its rotation (angular position change) during collision of a car and an obstacle (fig. 9) were determined. That movement is shown towards the initial head position, i.e. for t = 0 s. In all tests, the head displacement is very high and amounts to 0,44 m in the test 11P and 0,49 m in the test 2P (table 2). A classic position of the body on the seat (test 2P) leads to the highest head displacements. In the test 11P, a significant scope of the torso movement resulted in limited head displacement.

Obtained results were referred to the paper [3], where many results of the tests on the scope of the human head movement in the crash tests were presented. The tests were performer during 1980-1990. Based on that, an empirical dependence was determined:

$$a = 0{,}94 \cdot (0{,}83 \cdot \Delta v + 18) \tag{2}$$

where:

a [cm] – a scope of the head movement towards the car speed vector,

$\Delta v\,[km/h],\ a[cm]$ – head velocity decrease value during the test.

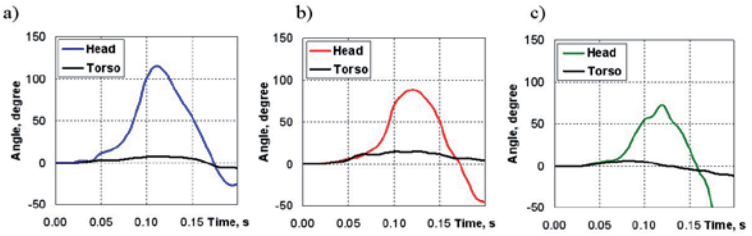


Fig. 9. Longitudinal displacement and rotation angle of the head towards the initial position in the tests 2P (a), 3P (b) and 11P (c)

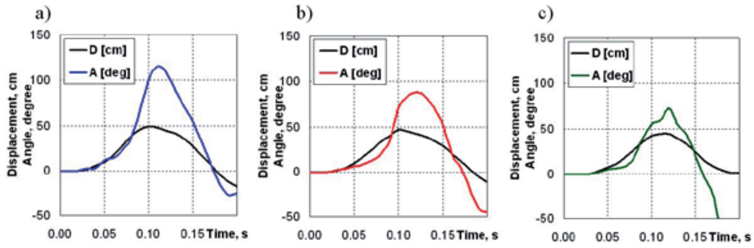


Fig. 10. Comparison of changes in the head and torso position angle changes in the tests 2P (a), 3P (b) and 11P (c)

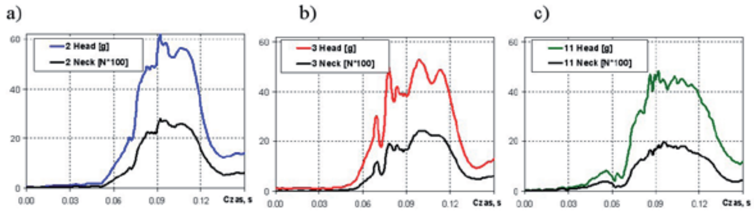


Fig. 11. Resultant head acceleration and resultant force in the neck in the tests 2P (a), 3P (b) and 11P (c)

Table 2. Specification of characteristic values describing the head movement

Value	Test 2P	3P	11P
Maximum longitudinal head displacement, *m*	0,49	0,47	0,44
Maximum change of the head inclination angle value towards the initial position, degree	115	88	72

On the basis of (2), longitudinal head displacement was calculated for the conditions of performed tests and the following value was obtained *a* = 0,54 m. So the distance defined on the basis of experimental tests carried out during 1980–1990 [3] is higher by 10–17 % than the one currently observed in the tests of that type (table 2). It can be interpreted as a good proof of progress in operation of the individual protection equipment.

Conclusions from the comparison of results shown on figure 10, where courses of the head and torso position angle changes in the time function in subsequent tests are presented, are interesting. In the tests 2P and 3P, where the torso movement is strongly (but properly) limited (table 1) by effective influence of the belt reaction forces, the extreme values of the angular head and torso movement occur at the same stage of the car and obstacle collision process, i.e. 115–125 ms (fig. 10 a, b). While in the test 11P, where the belt did not operate properly, the extreme values of the torso rotation angle occurred already in 80–90 ms (fig. 10 c), and at the later stage of the impact the torso slides on the seat. During that slide the hips slide out from under the belt to the space between the rear seat and the front seat. The torso inclination angle decreases but the head rotation angle value continues to increase during the time up to 120 ms.

4.3. Relation between courses of dynamic loads affecting the head and the neck (force in the neck)

Performed measurements confirmed that relatively small changes in the passenger position and location towards the individual protection equipment in a car have a significant influence on the risk of head and neck injuries during a road accident. These injuries occur as a result of forces of inertia caused by the presence of high head acceleration values. Courses of resultant head acceleration $a_G(t)$, compared on figure 11, show that position change results in:

- delay of the beginning of the intense acceleration value increase process (evaluated at the level of 10 g) by about 10 ms; the beginning of the intense increase of the value $a_G(t)$ in the tests P2 and P3 was observed in time of about 60 ms and 70÷72 ms in the test 11P;
- decrease of extreme head acceleration values by about 6-10 % in the test 11P towards the values measured in the tests 2P and 3P;
- delay of the beginning of the intense neck dynamic load increase process by about 5 ms (evaluated value obtaining time at the level of 30 % of the extreme values); that beginning was estimated for 68÷70 ms in the tests 2P and 3P and 72÷75 ms in the test 11P;
- decrease of the extreme values of the resultant force in the neck by 20÷24 % in the test 11P towards 2P and 3P.

5. Analysis of biomechanical index values

Calculations of biomechanical index values make the final element of the evaluation of the results of the passenger position change towards the seat and the seat belts. Three indexes were used *HIC*, N_{ij}, *CAcc*. Head Injury Criterion (*HIC*) calculated on the basis of head acceleration, make the basis for the injury risk evaluation. Acceleration of the centre of the head, measured in three mutually perpendicular directions, was used in the following way:

$$HIC = \max\left[\frac{1}{(t_2-t_1)}\int_{t_1}^{t_2} a_G(t)dt\right]^{2,5}\cdot(t_2-t_1) \tag{3}$$

where:

$a_G(t)$ – resultant acceleration of the centre of the head in *g;*

$\Delta t = t_2 - t_1$ - time interval in seconds when the highest head acceleration values occur.

Calculated values, shown on figure 12, do not exceed the criterion HIC=1000 (and it means 50 % injury risk at AIS scale at the level of AIS2+ or 24 % at the level AIS3+ [17]). The highest value occurred in the test 2P. The highest head load level in this test is confirmed by results shown on figures 10 and 11 and in the table 2. They result from a free head movement towards the body in a classic passenger position on a seat in a car. In the remaining two tests, where a small torso movement away from the backrest was applied and its higher angle of deviation backward from the vertical position than in 2P, a free head movement was limited and it resulted in a lower dynamic load.

The head movement towards the body can result in the neck spine injuries, and their type depends on the direction of the head movement under the influence of the force of inertia and the pressure of the torso.

The risk injury is evaluated by using the index N_{ij}, and its value is calculated as follows [14]:

$$Nij = \frac{F_z}{F_{zc}} + \frac{M_{OCy}}{M_{yc}} \tag{4}$$

where:

F_z, M_{OCy} – axial force (F_T, F_C – tension, compression) and corrected moment of bending the neck towards the axis Oy (M_E, M_F – flexion, extension),

F_{zc}, M_{yc} – critical values of forces F_T and F_C and moments M_E and M_F .

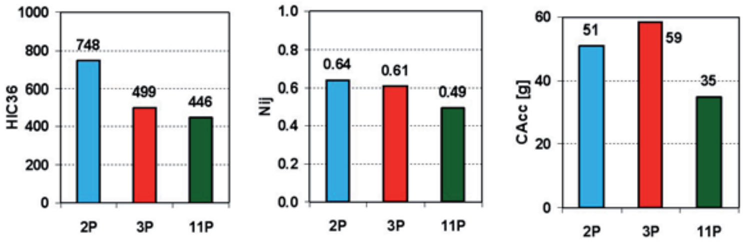


Fig. 12. Biomechanical index values, calculated on the basis of the results of measurements for three positions of the rear seat passengers in a passenger car

The criterion index value N_{ij} should be lower than 1. The value N_{ij} =1 means 30 % risk of occurrence of medium injuries at the level of AIS2+ or 18 % risk of very severe injuries AIS4+ [1]. In performed tests, the index values are high and highest ones in 2P and 3P. The definitely lower value of the neck injury index in the test 11P (lower by 20 %) is conditioned by excessive torso movement and it allowed for reduction of dynamic loads of the head and the neck, but it occurred at the expense of excessive dynamic loads of the passenger's legs, belly and the lower section of the spine.

During chest injury risk evaluation, the values of its maximum acceleration *CAcc,* were used that occurred in time of at least 3 ms. The value *CAcc* = 60 g means 20 % risk of very severe injuries AIS4+ [9] . The risk of chest injuries is additionally evaluated on the basis of the maximum value of the force of stretching the shoulder belt during collision. Considering that they amounted to the level of 600-700 daN (compare Fig. 7), so according to the paper [8], a risk of severe injuries (AIS3+) for people at the age of 30 years amounts to 5-10 %, but for people at the age of over 50 years it is as high as 60-100 %.

Calculated values of biomechanical indexes confirm that there is a high risk of injuries for passengers. The influence of the position change on the level of that risk is visible but not unequivocal. It indicates that passive safety equipment is incompatible with passenger position change in cars. Despite the fact that the latest model of Hybrid III dummy model was used for the tests, it does not allow for detection of all threats occurring due to a limited efficiency of operation of individual protection equipment.

6. Summary

Test and measurement results were presented where three different initial dummy's positions on a rear seat of a passenger car were applied. The results of the change to the initial position towards the individual protection equipment caused by the changes to the torso and the head movement trajectory and dynamic load values, occurring during frontal collision of a car and an obstacle, were analysed. The subsequent figures show general characteristics of the dummy movement in individual tests, including torso displacement in its movement on a seat and the head movement. Already small changes to the initial torso position towards the classic position (test 2P) result in its displacement at the final stage of the accident by over 80 % (compare test 11P). In all tests the head displacement is very high and amounts to 0,44 m in the test 11P and up to 0,49 m in the test 2P. A classic body position on a seat (compare the test 2P) leads to the highest head displacements that are by over 10% higher than in 11P. In the last test a significant range of the torso movement resulted in limitation of the angular head movement.

It should be strongly highlighted that relatively small changes to a passenger position and location towards the individual protection equipment in a car have a serious influence on a risk of injuries during a road accident. That risk was evaluated for each position change. Calculated values of the probability of the head, neck and chest injuries are high and directly indicate that there is a need to adjust protection devices to changing positions of the rear seat passengers, including changes to their position while driving.

***Acknowledgement**: The research work was done as part of N N509 559640 Project, funded by the National Science Centre, Poland*

References

1. Adam T, Untaroiu CD. Identification of Occupant Posture Using a Bayesian Classification Methodology to Reduce the Risk of Injury in a Collision. Transportation Research, 2011; Part C 19: 1078–1094.
2. Aduma S, Oota K, Nagumo H, Okabe T. Development of New Airbag System for Rear-Seat Occupants. The 21th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Stuttgart, Germany, 2009.
3. Araszewski M, Roenitz E, Toor A. Maximum Head Displacement of Vehicle Occupants Restrained by Lap and Torso Seat Belts in Frontal Impacts. SAE Technical Paper Series 1999; 1999-01-0443.
4. Bidez MW, Hauschild HW, Syson SR, Mergl KM. Lap-Shoulder Belt Performance as a Function of Occupant Size, SAE Technical Paper Series 2005; 2005-01-1705.
5. Bilston LE, Dud W, Brown J. A matched-cohort analysis of belted front and rear seat occupants in newer and older model vehicles shows that gains in front occupant safety have outpaced gains for rear seat occupants. Accident Analysis and Prevention 2010; 42: 1974–1977.
6. Bose D, Crandall JR, Untaroiu CD, Maslen EH. Influence of pre-collision occupant parameters on injury outcome in a frontal collision. Accident Analysis and Prevention 2010; 42: 1398–1407.

7. Evans L, Frick MC. Seating Position in Cars and Fatality Risk. USA, American Journal of Public Health 1988; Vol. 78, No. 11: 1456-1458.
8. Foret-Bruno J-Y, Trosseille X, Lecoz Y, Bendjellal F, Steyer C, Phalempin T, Villeforceix D, Dandres P and Got C. Thoracic Injury Risk in Frontal Car Crashes with Occupant Restrained with Belt Load Limiter. SAE Technical Paper 983166, 1998.
9. Hackney JR, Kahane CJ. The New Car Assessment Program: Five Star Rating System and Vehicle Safety Performance Characteristics. SAE Technical Paper 950888, 1995.
10. Jurecki R, Stańczyk TL. The Test Methods and the Reaction Time of Drivers. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2011; 3: 84-91.
11. Kisilowski J, Zalewski J. Chosen Problems of Road Accidents Analyses in Poland in the Period Between 1995 and 2004. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2008; 1: 37-43.
12. Martin JL, Lardy A, Compigne S. Specificities of Rear Occupant Protection: Analysis of French Accident Data. IRCOBI Conference, pp. 315-318, Hanover, Germany 2010.
13. Merkisz J, Tarkowski S. Selected Aspects of Using Deck Recorders in Automotive Vehicles. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2011; 2: 50-58.
14. Pike JA. Neck Injury Biomechanics, SAE International PT-141, 2009.
15. Prochowski L, Żuchowski A. The Analysis of the Influence of The Rear Seat Passenger Position on the Kinematics and Dynamic Loads on a Torso and Legs During a Road Accident. Journal of KONES, 2013; Vol.20. 335-342
16. Smith KM, Cummings P. Passenger seating position and the risk of passenger death in traffic crashes: a matched cohort study. Injury Prevention 2006; 12: 83–86.
17. Wicher J. Bezpieczeństwo samochodów i ruchu drogowego. WKŁ, Warszawa 2012.
18. Żuchowski A, Prochowski L. Analysis of the Influence of the Seat Taken in a Car on the Dynamic Load Affecting People During the Obstacle Impact by the Passenger Car. Automotive safety problems, Wydawnictwo Politechniki Świętokrzyskiej 2012; Vol.1: 113-126.
19. Żuchowski A, Prochowski L. The Analysis of Safety of Rear Seat Occupants of a Passenger Car during the Road Accident. The Archives of Automotive Engineering, Warsaw 2011; 3: 83-98.
20. Żuchowski A. Risk of Injury for the Front and Rear Seat Passengers of the Passenger Cars in Frontal Impact. Journal of KONES Powertrain and Transport 2012; Vol.19, No.3: 507-518.

Leon PROCHOWSKI
Andrzej ŻUCHOWSKI
Institute of Motor Vehicles and Transportation
Faculty of Mechanical Engineering
Military University of Technology
ul. Gen. S. Kaliskiego 2, 00-908 Warsaw 49, Poland
E-mails: lprochowski@wat.edu.pl, e-mail: azuchowski@wat.edu.pl

Article citation info:
ZBROWSKI A, SAMBORSKI T. Study on electromechanical drives used in ventilation and smoke extraction systems. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 367–376.

Andrzej ZBROWSKI
Tomasz SAMBORSKI

STUDY ON ELECTROMECHANICAL DRIVES USED IN VENTILATION AND SMOKE EXTRACTION SYSTEMS

BADANIA NAPĘDÓW ELEKTROMECHANICZNYCH STOSOWANYCH W SYSTEMACH ODDYMIANIA I WENTYLACJI*

The paper presents experimental tests for electromechanical drives for use in ventilation and smoke extraction systems of buildings. The determined characteristics (force, shift, current and voltage as a function of time) enable the assessment of applicability of these drives as the drives of flaps and windows in fire protection systems. The study was conducted using a developed test stand facilitating the determination of characteristics of work of linear and rotary drives in changeable temperatures.

Keywords: *test system, electromechanical drive, smoke extraction and ventilation, fire protection system.*

W artykule przedstawiono badania eksperymentalne napędów elektromechanicznych pod kątem zastosowania w systemach wentylacji i oddymiania budynków. Wyznaczone charakterystyki (siła, przemieszczenie, prąd i napięcie zasilania w funkcji czasu) pozwalają na dokonanie oceny możliwości ich zastosowania do napędu klap i okien w systemach przeciwpożarowych. W badaniach wykorzystano opracowane stanowisko umożliwiające wyznaczanie charakterystyk pracy napędów liniowych i obrotowych w zmiennych warunkach temperatury.

Słowa kluczowe: *system badań, napęd elektromechaniczny, oddymianie, wentylacja, system ppoż.*

1. Introduction

Electromechanical actuators (EMAs) are much more commonly used in industrial applications, where there is a must for multiple forces and loads to be reliably exerted. The purpose of electromechanical solutions is to convert electrical energy into mechanical energy of the actuator.

The dynamic development of EMAs stems from their growing popularity in the most demanding sectors of modern technology. Due to the risk and criticality of functions executed in actuators in applications concerning human and technical object safety, EMAs used must meet the highest durability and reliability requirements. For that purpose, it is necessary to conduct detailed tests enabling the determination of operating characteristics of EMAs working under different conditions of force and environmental excitations. Laboratory tests performed in a relevant environment resembling real-life operating conditions are the foundation for the acquisition of specialised knowledge about the correct operation and durability of EMAs. The work parameters recorded during the tests enable the development of modern diagnostic systems which are an essential element of safe control systems. The information from the investigations provides valuable data sets used in the diagnosis and operating capacity prediction models. Improving test methods, the reliability of recreated operating conditions, and the accuracy of recorded work parameters of EMAs are the basic conditions for ensuring a high level of technical safety.

2. Application of electromechanical actuators

EMAs are finding increasing use in new areas of application in the aviation (both civil and military), aerospace and military industry. As for the construction of aircrafts, electric actuators are responsible for the positioning of the elements of a tailplane, i.e. ailerons, flaps, spoilers, or aerodynamic brakes. In the case of spacecrafts, electric actuators are used for the positioning of antennas and robotic arms, but there are also plans for such actuators to be used for the construction of rockets and their thrust vector control [25]. R&D works aimed at the application of EMAs for the positioning of rudders of ships and submarines are also conducted [20, 38].

In latest aircrafts (i.e. Boeing 787 or Airbus 380), EMAs operate landing gear breaks, and in the case of military aircrafts (i.e. F-35 Joint Strike Fighter), they are used for the positioning of weapons [1]. The nearest future will see the intensification of R&D aimed at the construction of aircrafts in which electric systems will entirely replace hydraulic systems [2]. The elimination of the central hydraulic system will reduce the weight of an aircraft, which in turn will lead to the decrease in the power consumption.

EMAs are, however, far more complicated than hydraulic actuators, which stems from the necessity to integrate in one module numerous precise components including i.e. electric motors (usually two), differential gears, reduction gears, screw mechanisms and built-in control system elements [25]. Despite the high level of complexity, these systems have to comply with severe and restrictive requirements, stating that the faults may not occur more often than $1x10^5$ (strike-fighter aircraft F/A 18) or even $18x10^6$ (fighter F-35 AB) flight hours [6].

Taking into consideration the issues of safety, actuators are the critical components of an aircraft system (fig. 1).

Even a small, undetected fault can lead to very serious consequences. This is confirmed by results of aircraft accidents, in which such failures as e.g. the failure of the actuator of the horizontal rudder, turned out to be the cause of the crash. Cases in which excessive, uncontrolled wear of servomotors stood behind the grounding of the entire fleet are also recorded. Therefore, actuators used in aviation undergo detailed functional tests, whose results enable the develop-

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

Fig. 1. Actuators with screw gear controlling the positioning of the flaps in the IL-76 transport aircraft

ment of reliable structures and support better understanding of various causes of defects of complex structures of EMAs.

Collecting knowledge from the experiments enables fast and reliable detection and identification of faults, in which such additional methods as FMEA and FTA supporting the design of suitable algorithms tailored to specific actuators, are used [14]. The developed actuator fault prediction methods are an essential element of contemporary monitoring systems that use artificial intelligence, fuzzy logic and neural networks. The analysis of possible faults allows the selection of detection algorithms - heuristic methods, model-based methods and experimental data sets [13, 17]. They provide early detection of small defects, and enable the monitoring of gradual degradations.

The developed diagnostic systems for EMAs support proper planning of renovations. The traditional approach of overhaul after given hourly service is an unsatisfactory solution. The observed wear, especially in the case of military aircrafts, depends on the individual style of piloting and the operating environment. The reasons for the occurrence of dangerous situations lie in the wear which, due to the unique style of piloting under e.g. battlefield conditions, is a much quicker phenomenon than anticipated. [17]. Therefore, it is reasonable to replace periodic maintenance with the strategy to carry out repairs based on the current assessment of the technical condition of the object, conducted using an intelligent system forecasting further operating capacity. Systems for the diagnosis of actuators will enable the elimination of redundant systems, typical for aircraft structures, which in turn will contribute to a significant reduction in their weight and the reduction in the overall dimensions of new aircrafts and other systems responsible for technical safety [13].

Due to the responsibility of the tasks performed, the accurately archived set of many years of experience, and the high implementation potential, studies on actuators used in aerospace set the standard for tests for these devices also in other fields of their application, like mechanical engineering, shipbuilding and construction industries, for instance.

3. Actuator tests

Study on EMAs enables the development of safe systems with damage tolerance, the self-diagnosis, the early detection and localisation of defects, and the assessment of the degree of degradation of devices and actuators. The development of the advanced research on EMAs is mainly driven by the needs of the aerospace industry. For this purpose, extensive field and laboratory tests are conducted, and their results are the basic information used in the developed diagnostic and monitoring systems.

The first experiments were carried out on real objects, in which the hydraulic actuators were replaced with an electromechanical system. Such tests were performed under the Electrically Powered Actuation Design (EPAD) programme executed in cooperation with the U.S. Air Force, the U.S. Navy and NASA. In the study, an EMA was installed in the F/A-18B fighter to control the left aileron. The tests conducted during 22 flights for the total of 25 hours allowed the registration of work parameters of the actuator, and enabled the comparison with the operation of the typically used hydraulic system located in the right wing [12]. The main problems detected concerned the overheating of the engine, the inefficient energy dissipation during braking, and the low durability of the casing of the ball screw nut.

So as to enable the execution of cheaper laboratory tests resembling real operation conditions, a mobile test stand, making the tests to be performed in the flight mode, on board of an aircraft, was developed. The stand was developed in collaboration with the NASA Prognostic Center of Excellence and the California Polytechnic State University [3, 16, 27]. The installation of the developed stand does not require any modifications in the design of the aircraft structure. The operation on board of an aircraft makes it possible to perform tests under the influence of real loads exerted by varied values of vectors of velocities and accelerations occurring during the flight. The stand comprises three actuators: two tested actuators and a load exertion actuator. One of the tested actuators is in working order, whereas an intentional defined damage is caused to the other. The actuators which are being tested are, by means of magnetic coupling, connected to the load system. The types of faults caused concern the damage to the screw mechanism (loss of threading, deformation of balls, return channel ball jam, excessive backlash), the electric motor (overheating, vibration), and the sensors (indication error or signal loss). [5] The sensor system of the test stand enables the recording of the following parameters: the amplitude and frequency of vibrations of a tested actuator, the value of the load set, the temperature, the voltage and current of power, and the length and angular position of the screw. The characteristics are correlated in time with the information recorded by the on-board flight data recording system.

The tests enable the recording of the characteristics of the monitored parameters during the occurrence of the determined faults, and their comparison with the model derived from a working motor. The test system is designed to work with the Boeing C-17 Globemaster military transport aircraft and the UH60 Blackhawk helicopter. The stand can also be used for the testing of diagnostic systems monitoring the work of the tested actuators. The device became widely used as a test platform that allows the testing of diagnostic and control systems operating in conditions where the fault occurred. The stand was used for numerous investigations and verification tests for diagnostic and prognosis systems based on models, experimental data, and hybrid solutions combining analytical models with functions defined experimentally [4, 15, 25].

An alternative to mobile test systems are stationary stands enabling the performance of tests on actuators of larger size and load capacity. Tests on the Moog MaxForce 833-023 actuator presented in [1] were performed using a hydraulic system. The tested actuator was connected to a hydraulic load actuator by means of a rotating robotic arm. The stand enables the three directional measurement of accelerations impacting the actuator, the registration of the temperature of the stator, the determination of the position of the tip of the piston rod, and the two directional measurement of the force on the piston rod. The control system enables different levels, types and profiles of loads to be set for a tested actuator.

Another stationary test stand built in cooperation between the Impact Technologies and the NASA Ames Research Center is used for the experimental study on diagnosis and prognosis methods for EMAs with screw gear [1]. The dynamic loads are set using a large Mogg 866 EMA enabling the generation of longitudinal force of up to 50 000 N. The control system ensures the possibility of varied shaping of the load profile. The load can be exerted in a rectangular, trapezoidal, sinusoidal and triangular manner. The structure of the stand enables the performance of long-term tests with data recording at up to 64 kHz.

A simplified version of the stand is a solution using a pneumatic actuator to exert a load on the piston rod of the tested EMA [1]. Using a pneumatic system, it is possible to set loads which do not exceed the value of 4500 N. The system is equipped with three independent controllers. Two of them work in a subordinate system and control the operation of the proportional air pressure regulator and the servo motor, and cooperate with the master PXI controller.

A slightly different concept of EMA tests was used in the method using Hall sensors for the diagnosis of the technical state of screw gears. The non-contact measurement technique facilitates the monitoring of the wear of an actuator based on the detected defects of balls in the screw mechanism [9]. The Hall sensor installed on the nut makes it possible to generate a sinusoidal signal, proportional to the distance of the ball from the sensor surface. In the case of detection of a defective ball, the signal is distorted and has smaller amplitude. The study was conducted on the stand in which the gear was driven by a 30 kW electric motor. The axial load of a tested mechanism was set using a hydraulic actuator connected with the front of the screw. The level of the load dependent on pressure in the actuator controls the controller of the proportional valve.

Tests on EMAs using stepper motors can be performed on a test stand developed at the Micromechanics and Photonics Institute at the Warsaw University of Technology. The stand is intended for the determination of operating characteristics for precise, linear, high resolution actuators, and their positioning precision [43]. The determination of border movement characteristics is performed using feed impulses of known frequency, through the loading of the pusher with defined force, and the precise detection of the pusher's movement. The stand has two load exertion fields: the control field and the actuator load field, and two measurement fields for the measurement of the load force and the shift of the pusher. The tested actuator is loaded using a DC motor with electronic commutation, coupled with the pusher by means of a tension roller mechanism with a built-in force strain gauge sensor. The border start characteristics are determined through the programmed loading of the actuator with the predefined force, and the search for the maximum beat frequency at which the actuator will work in a stable manner. The measure for actuator's stable operation is the linear shift of the pusher stemming from the set number of control impulses.

Apart from tests performed by R&D organisations, studies on EMAs are also conducted by industrial R&D centers, in which case the results of such investigations are confidential and unavailable in print. The example of industrial tests on EMAs are experiments conducted by the Honeywell International Inc. on linear EMAs used i.e. in aviation, transport, process control [28]. The tests are conducted in dynamic load conditions reflecting the character of the actuator's operation resulting from the moving of masses or the presence of different load forces, in order to determine the inertia and delay the reaction of the actuator.

Control over technological processes involving the transmission of gases and liquids (chemical, mining, refining industry) requires the use of remotely controlled valves driven by rotary actuators. The manufacturers of industrial automation including valve actuators developed stationary (PV 1405 AUMA) and mobile (PV 1236 AUMA) test devices equipped with an integrated diagnosis system allowing the verification of operating parameters of rotary actuators in a wide range of loads (up to 6 000 Nm).

Stands for testing rotary actuators used in rail transport were also developed by the Kyalsi Engineering Inc.

4. Study on electromechanical actuators in ventilation systems of buildings

Out of the technical areas in which EMAs are broadly applied, the ones of great importance are construction and fire protection. Reliability of EMAs is also directly related to the issues of fire safety, particularly fires safety of public buildings.

In order for rescue operations to be carried out effectively, the area where the fire broke out needs to be cut off, and the smoke removed from the fire exit route [24] The closing of the fire screen bulkhead, the opening of smoke vents or ventilation windows is performed using EMAs [21, 22, 32, 42]. The screw, chain and spindle mechanisms used in actuators allow effective and fast positioning of the effector (flaps or windows) so as to transfer the smoke and hot air outside a burning building (Fig. 2).

Fig. 2. Linear EMA with screw gear controlling the opening the smoke extraction flap located on the building's roof

The proper functioning of fire ventilation is a prerequisite for efficient evacuation of a building [10, 30, 31], especially in high-rise buildings, in which it is difficult for rescue actions to be performed from the outside. In the case of floors located over 50 m above the ground, the only chance for rescue is the evacuation through the building's corridors and staircases [11, 13, 18, 19, 37].

EMAs used in modern building automation systems constitute a unique group of electricity receivers operating in fire protection system. Their reliable operation, in the case of fire, is essential for an effective evacuation of people trapped in a burning building.

In automatic flap control systems, the electric current, once transmitted to each object, is distributed via proper installation tailored to the used actuator voltage and current standard. Actuators for smoke extraction systems most commonly use 24V DC [39]. This requires special units equipped with batteries responsible for emergency power backup to be mounted. Reliable and safe transmission of energy from the unit to the receiver requires the construction of an additional electrical installation [35, 36].

The high temperature at the time of fire leads to decreases in the electrical conductivity of wires, which results in lower quality of the supplied power manifesting in the excessive voltage drop and the worsening of fire protection conditions [34, 40]. The low value of voltage results in the decreased torque of the electric motors used in actuators, which can significantly hinder the effectiveness of a rescue operation. Proper cooperation between the power supply system and the actuator is possible only if the required level of electricity consumption is maintained, regardless of the loads and temperature [33] in the building. They have to operate reliably under extreme conditions which occur during a fire.

As the evacuation of people from burning buildings is the most important element of rescue actions, formal requirements for the construction of a building and the electrical appliances installed in it, including EMAs, and working at the time of the fire, are set.

The EN 12101-2 standard [26] specifies the requirements for smoke flaps mounted on the roof and smoke flaps installed in the facade of the building, popularly known as smoke extraction windows.

According to the standard, the window smoke extraction system (window + drive, the so-called Natural Smoke and Heat Exhaust Ventilation – NSHEV) should constitute a complete CE-labeled solution, in accordance with the 93/68/EC Directive.

Due to the direct relationship with the consequences of accidents, technical rescue and fire protection have a particularly important place in a technical safety system. Therefore, the introduction of any new solution must be preceded by a detailed study at the stage of R&D and certification as well.

5 Test methodology

In order to ensure the required reliability associated with the performed functions including the removal of toxic fumes [8, 28], electric drives designed for use in safety systems have to undergo a series of tests [7] which will confirm the required level and stability of relevant operating parameters.

One of the areas of the activity of the Institute for Sustainable Technologies - National Research Institute in Radom includes conducting research related to improving technical object safety [23]. For several years, the Institute has been conducting, in cooperation with the Science and Research Center for Fire Protection - National Research Institute in Józefów, R&D works related to the development of procedures and construction of specialised test and research apparatus supporting the safe operation systems [44], including technical rescue technical [17], and fire protection [29].

The study on ventilation systems performed jointly by the above listed research organizations resulted in the development of a concept of a model test system for EMAs used in fire protection systems.

The developed concept takes into account the directive of the Minister for Internal Affairs and Administration on 20th June 2007, which lists products facilitating public safety, as well as health, life and property protection, and delineates the principles of their release. According to this regulation, the test object is assessed at room temperature. The forced load and the registration of characteristics take place after the natural cooling or heating of the object, previously subjected to thermal risks.

The original test methodology, in contrast to the existing legislation, assumes the on-line registration of the operating characteristics during the impact of the thermal load (Fig. 3). This ensures the assessment of the test object in conditions similar to actual operating conditions. Continuous recording of the basic work parameters of EMAs provides opportunities for extended, in relation to the defined by the existing rules, analysis of the impact of extortions on the behaviour of a tested object.

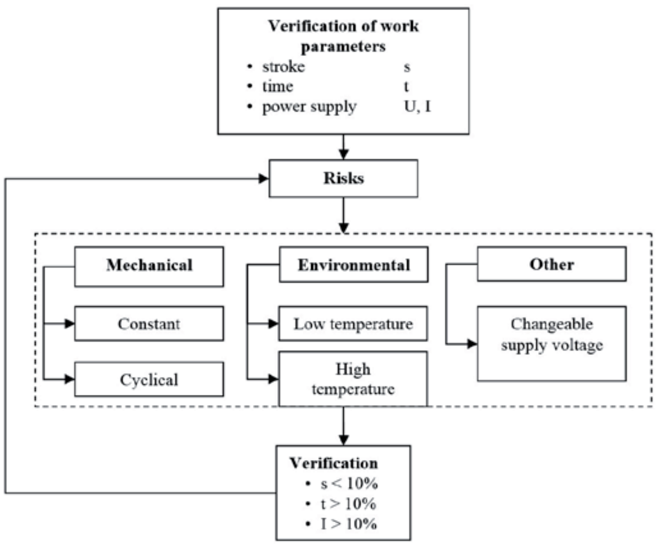


Fig. 3. Draft of selected test procedures for electromechanical drives

The assessment concerns, inter alia, such work parameters as the working stroke, the power consumption and the accompanying time of movement during the opening and closing. The criteria which allow the drive to be used in fire protection systems [7] concern the permissible changes in the following three parameters:

- The time of movement in both directions cannot be changed by more than 10%,
- The stroke cannot change by more than 5%,
- The increase in the power consumption cannot exceed the value of 10% for the two directions of movement.

6. Test stand

The basic element of the developed test system for drives used in smoke extraction and ventilation systems is a test stand, which allows drive tests to be conducted under varying conditions of temperature. Those tests are particularly executed for drives whose actuators perform linear or rotary motion. The developed stand offers a wider range of test possibilities, compared to existing solutions, which is due to the combination of force and thermal excitation systems in one compact unit with an integrated control and data acquisition system.

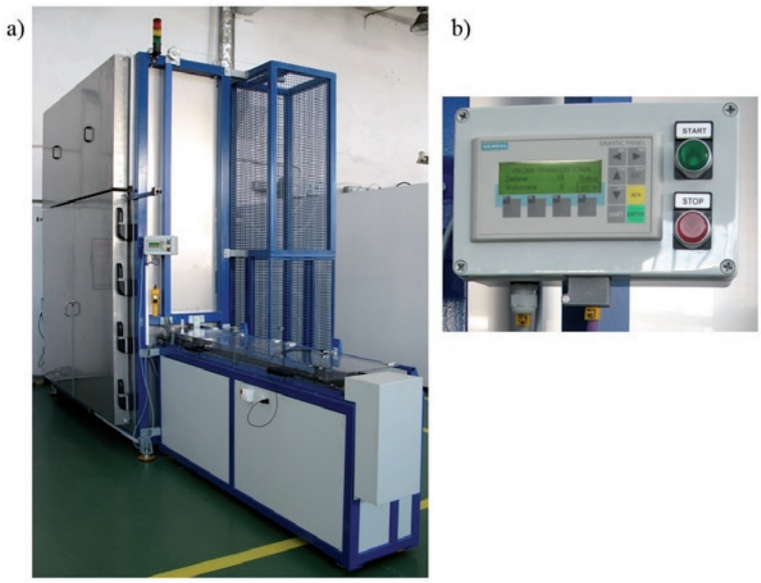


Fig. 4. Actuator test stand: a) general view, b) control panel

The device, designed and developed at the ITeE-PIB (Fig. 4) is composed of a thermal chamber connected to actuators for mechanical excitations. The basic element of the device intended for maintaining stable temperature and humidity conditions at the time of an actuator test, is a thermal chamber equipped with an adjustable, universal system allowing any type of actuator to be mounted.

The developed structural solutions allow the working elements of linear and rotary actuators to be loaded at the time of their movement using gravity or mechatronic load systems. The use of a gravity load system ensures load stability while conducting durability tests, where the required number of cycles can even reach 10 000. The mechatronic system, on the other hand, enables the execution of the endurance tests with predefined dynamics.

The basic parameters of the developed test stand are as follows:

- linear load up to 5 kN
- torque load up to 30 Nm
- shift at the time of:

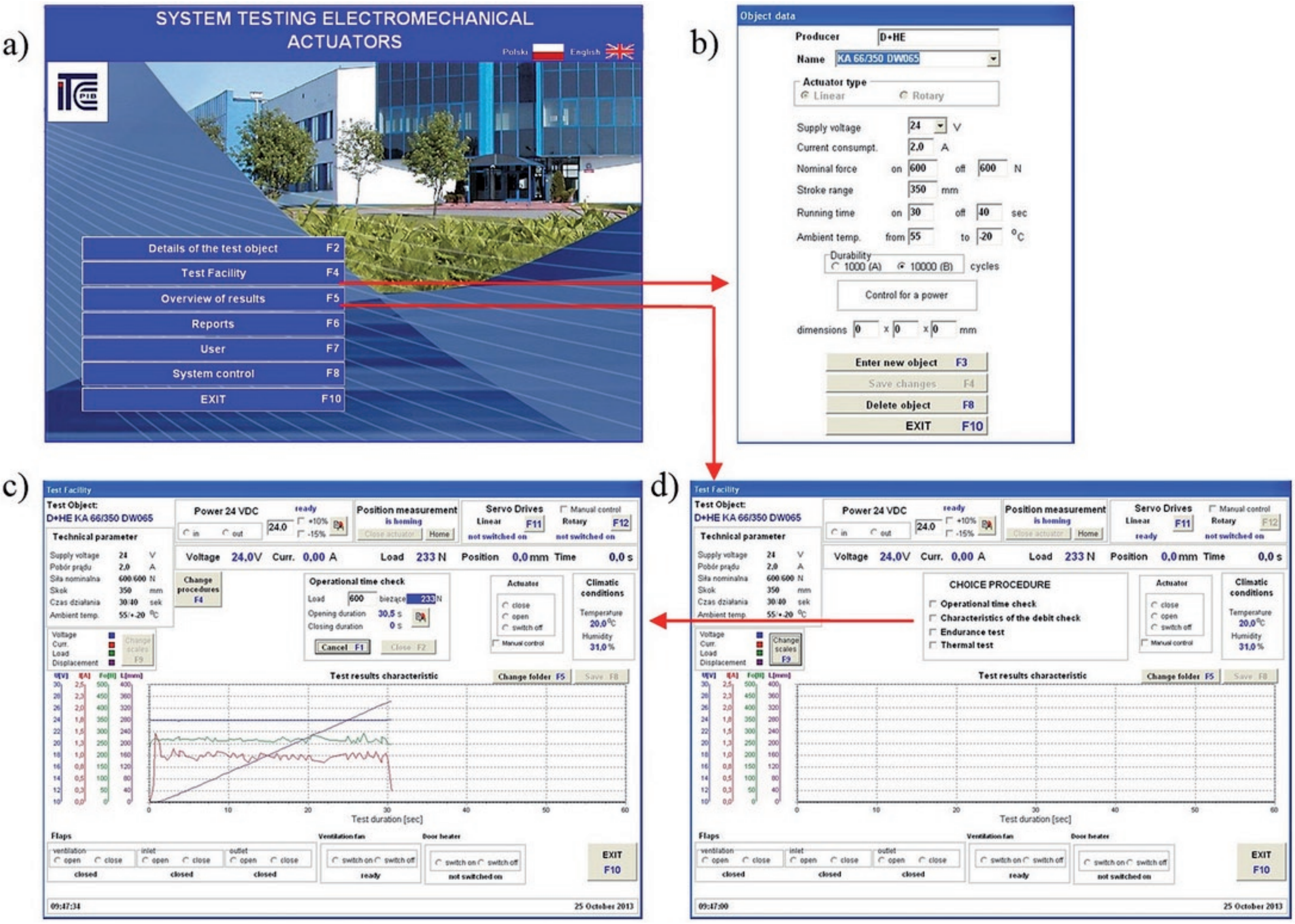


Fig. 5. View of the control system: a) main, b) identification of the object, c) measurements, d) selection of procedures

– linear movement up to 1,2 m
– rotational movement up to 4,7 rad
– test temperature from 243 to 363 K
– actuator supply:
 electric actuator 24 V, 230 V AC/DC
 pneumatic actuator up to 1 MPa
– measurement and recording: force or moment of force, displacement, temperature, humidity, voltage, current and power.

Selected test procedures and data acquisition are controlled via a control and measurement system with specialised two-layered software. The first of the layers includes a graphic user interface (Fig. 5) communicating with the test system by means of a PC with Windows operating system enabling the identification of a test object, the selection of a test procedure, and the setting of test parameters. The latter layer, on the other hand, concerns the performance of test procedures selected by the user, controlled directly by the PLC.

The stand is particularly intended for the execution of certification tests for EMAs used in fire protection systems, therefore an unambiguous identification of a test actuator during the entire certification procedure is required. For that purpose, a procedure for data input was developed, as it is a prerequisite for the tests to be launched.

Once the user inputs the data into the object, he then can verify and record the basic actuator work parameters which are crucial from the point of view of its future field of application [30,40] (Fig. 5c). These parameters include the following:

– control voltage and current,
– ejection and insertion time,

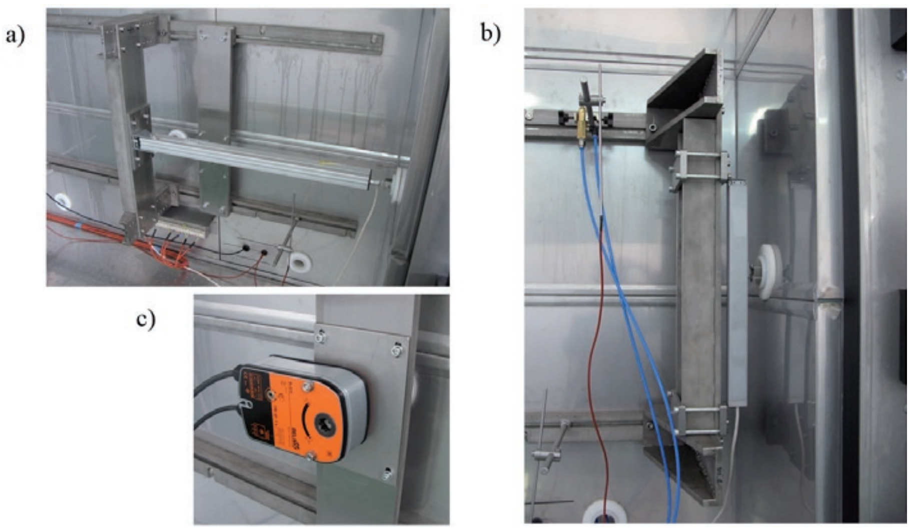


Fig. 6. View of actuators ready to be tested: a) spindle, b) chain, c) rotary

Table 1. Test conditions

Type of a drive	Symbol	Load	Temperature [K]
Spindle	W	460 N	253 ±1 293 ±1 328 ±1
Chain	L	105 N	
Rotary	O	15 Nm	

– stroke,
– nominal force.

The next step in the test process is the selection of a proper test procedure (Fig. 5d). The system generates the parameters of a selected procedure by default, allowing the user to make a final decision concerning their value.

7. Performance of tests

The tests were performed for electromechanical drives of different architecture and purpose. Spindle drives (Fig. 6a) are used for the opening of smoke extraction flaps, façade windows, skylights, glass roof structures, pyramids and roof windows. The structure of these actuators ensures high stability at the time of operation. The actuators' standard equipment includes limit position switches and safety switches in the case of overload. Chain drives (Fig. 6b) are used in the case of façade windows, roof windows and ventilation flaps. They are equipped with limit position switches, closing and opening force controllers, and chain length regulation system. Rotary actuators (Fig. 6c), those with a return spring, enable self execution of the movement of the shutter to the position required by the fire safety system. They are used to adjust the position of air dampers in ventilation and air-conditioning systems.

Table 2. Drive work parameters (temperature: 293 K)

Symbol	Stroke [mm, rad]	Max. power consumption [A]		Time [s]	
		Opening	Closing	Opening	Closing
W	500 ±0.5	0.78 ±0.01	0.20 ±0.01	72.3 ±0.5	59.1 ±0.5
L	346 ±0.5	0.55 ±0.01	0.09 ±0.01	55.8 ±0.5	70.8 ±0.5
O	1.57 ±0.01	0.26 ±0.01	0.03 ±0.01	30.0 ±0.5	20.5 ±0.5

Table 3. Drive work parameters (temperature: 328 K)

Symbol	Stroke [mm, rad]	Max. power consumption [A]		Time [s]	
		Opening	Closing	Opening	Closing
W	500 ±0.5	0.84 ±0.01	0.25 ±0.01	71.1 ±0.5	55.3 ±0.5
L	346 ±0.5	0.42 ±0.01	0.09 ±0.01	53.2 ±0.5	69.1 ±0.5
O	1.57 ±0.01	0.28 ±0.01	0.12 ±0.01	29.5 ±0.5	19.5 ±0.5

Table 4. Drive work parameters (temperature: 253 K)

Symbol	Stroke [mm, rad]	Max. power consumption [A]		Time [s]	
		Opening	Closing	Opening	Closing
W	500 ±0.5	0.85 ±0.01	0.30 ±0.01	78.3 ±0.5	63.2 ±0.5
L	346 ±0.5	0.65 ±0.01	0.15 ±0.01	58.3 ±0.5	72.5 ±0.5
O	1.57 ±0.01	0.26 ±0.01	0.06 ±0.01	31.5 ±0.5	24.8 ±0.5

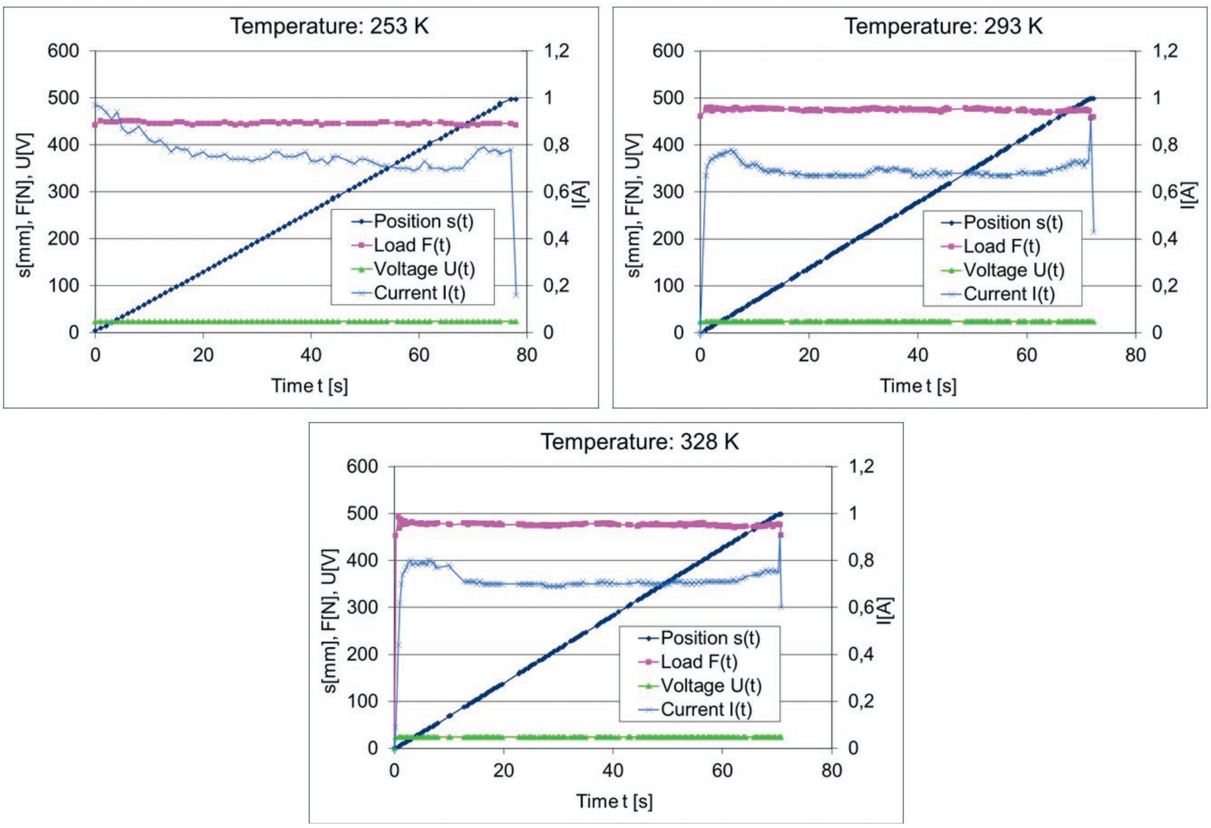


Fig. 7. Sample characteristic of W actuator's work

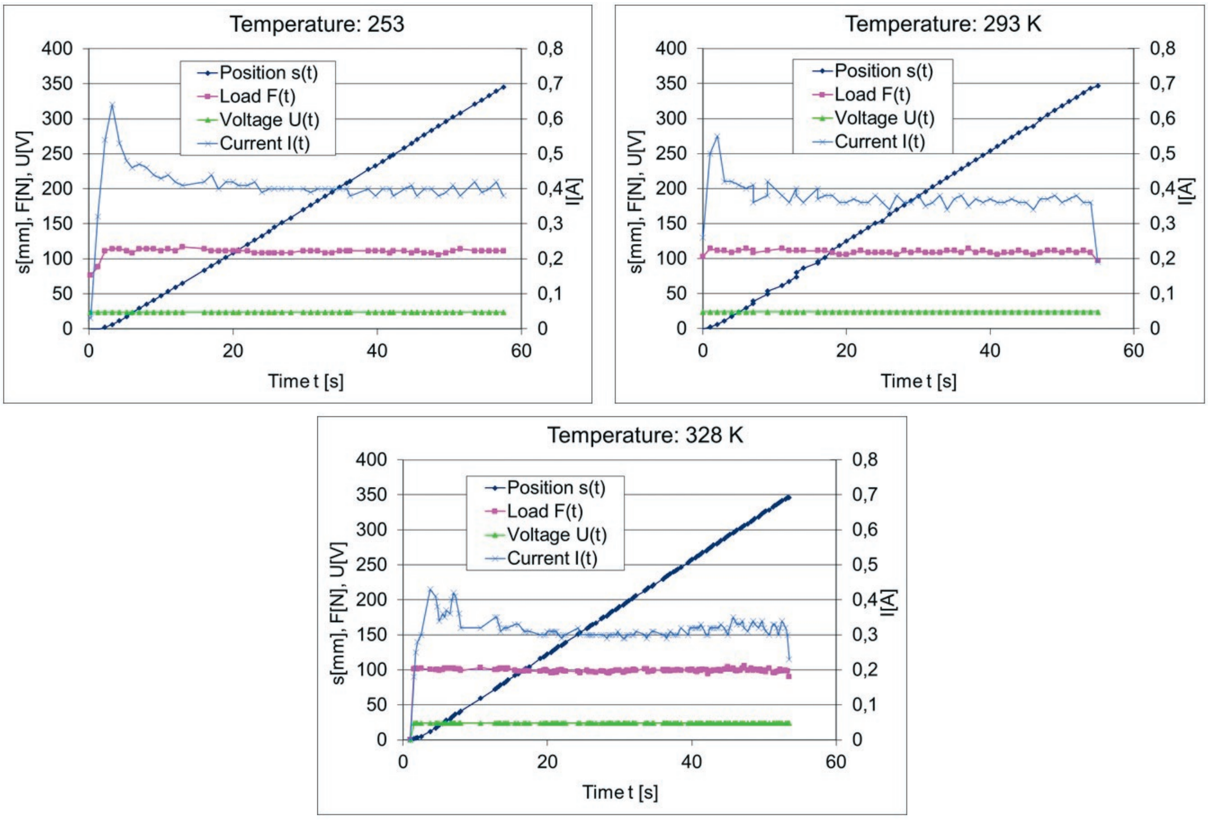


Fig. 8. Sample characteristic of L actuator's work

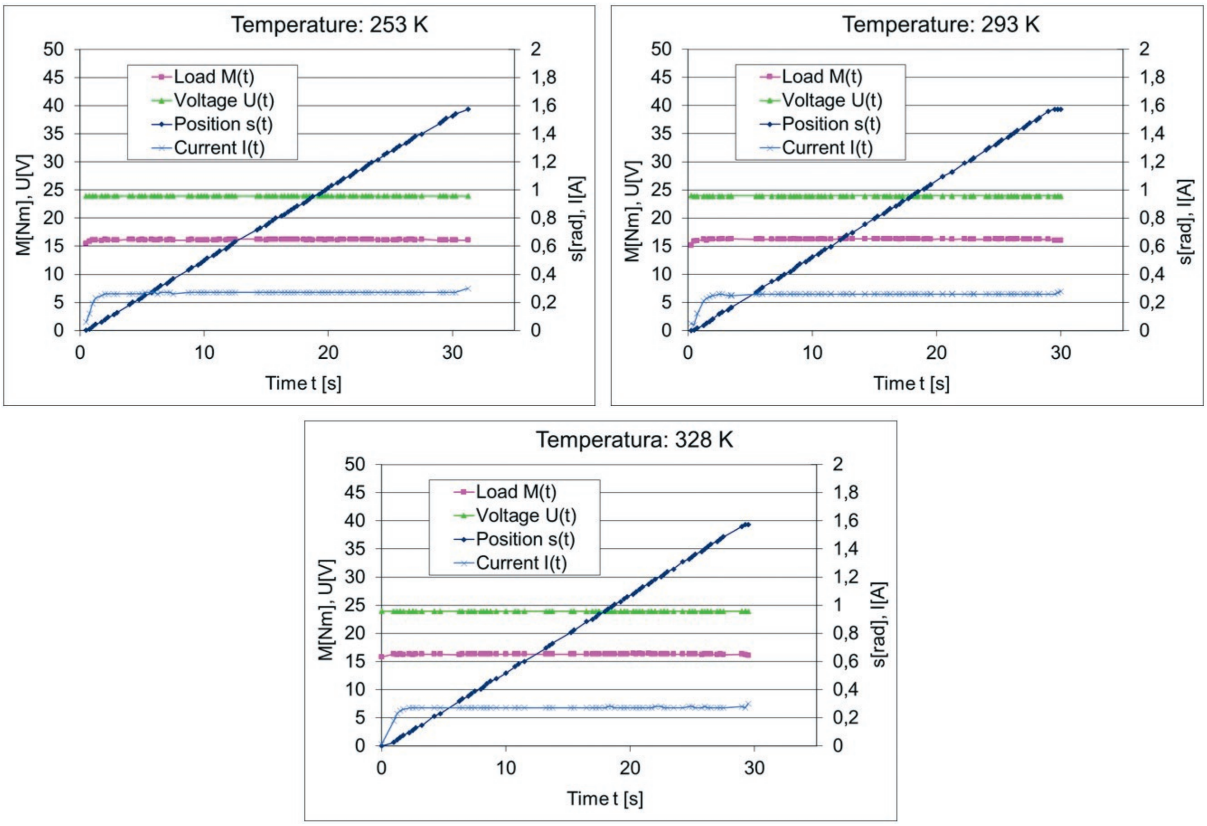


Fig. 9. Sample characteristic of O actuator's work

The way the tested actuator was mounted in the test stand depended on its type and structure, and corresponded to the real character of its work. The temperature of the actuator was set based on the readings from two measuring points located in close proximity to the actuator. The readings were verified using a third sensor placed in the central part of the test zone in the working chamber.

The tests conducted aimed at the determination of the influence of the temperature on the time of movement of a working element of the

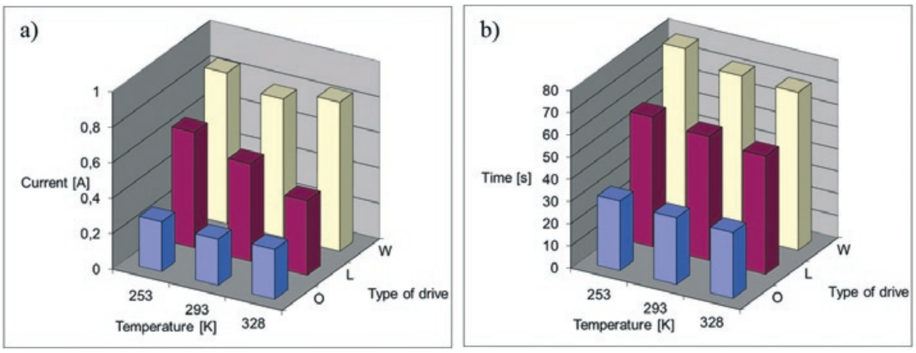


Fig. 10. Comparison between power consumption (a) and time of movement (b) depending on work temperature

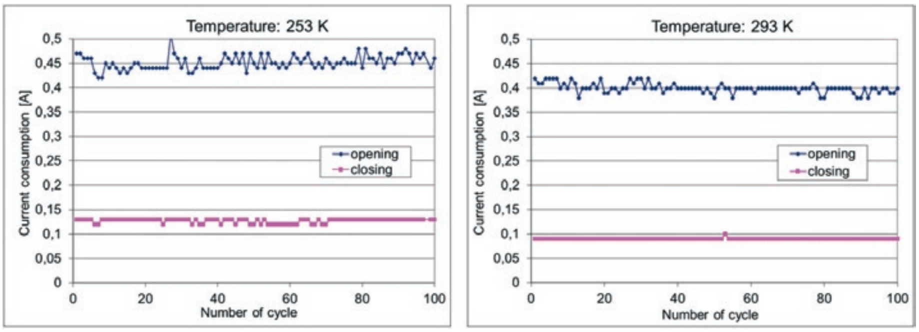


Fig. 11. Comparison of average power consumption for a chain drive in a set movement

drive, and its impact on the power consumption at the time of constant load in changeable temperature conditions (Tab. 1).

The tested actuators were supplied with 24V DC. The tests were conducted at set temperatures after a two hour conditioning period ensuring constant thermal conditions in the entire volume of the object.

8. Test results

During tests, basic parameters deciding on the possibility to use tested actuators in smoke extraction systems were recorded. These parameters included the supply voltage and current, and the shift at the time of loading a working element in the unction of time (Fig. 7, Fig. 8, Fig. 9).

Due to low dynamics of tested actuators, the recording of measurement data was performed in a discrete manner, with the sampling frequency of ca. 2Hz, which was sufficient to conduct a necessary analysis.

The software used in the test system enables the observation of the recorded characteristics in an on-line mode, which allows their current control. The characteristics presented were designed in an external software for the analysis of test results, and developed based on the data archived in *.csv files.

Analysing the changes in recorded work parameters of the actuators, a clear division into three areas can be made. The first area includes the time from the activation to the beginning of the movement of fixed nature; the second – the movement; and the third – the deceleration zone within the internal safety system.

Actuators W and L, during the initial movement, are characterised by increased consumption of power with non-linear waveform. Additionally, the level of value of power consumption depends on the temperature in which the test is conducted. The power features of the actuator O are characterised by a stable waveform over the entire work range of the device. Fluctuations and single peaks visible in the graphs of power characteristics indicate the changeability of internal resistance to movement occurring during the operation of the actuator.

For all tested actuators, the shift is described by linear characteristics, in which the value of the shift is directly proportional to the time of their activation. This means, that the shift of the actuator is of uniform nature, regardless of the momentary changes in resistance to movement.

Constant functions describing the graphs of characteristics of external loads and the supply voltage maintained, unambiguously document the stable conditions of the test process.

In order to increase the readability of the obtained results, based on the recorded characteristics, mean values for individual parameters (Tab. 2, Tab. 3, Tab. 4), deciding on the possibility to use the actuators tested in fire protection systems, were determined.

9. Analysis of the results

According to the test procedure define by the existing regulations [7], the verification of the controlled drive work parameters is performed in conditions corresponding to the temperature of the surrounding. In this case, for all tested objects, no changes exceeding 5% of the initial value (for the temperature of 293K) were recorded. Analysing the results obtained for changeable temperature conditions (in the range specified by the manufacturer) it can however be noticed, that changes in tested values take place at a significantly higher level (Fig. 10)

In the case of spindle and rotary drives, the range of changeability of the maximum power consumption (8.2%), and the time of movement during the opening (7.7%) did not exceed the allowed 10% of the initial value. In the case of chain drives, however, a significant increase in power consumption exceeding 15% was observed for subzero temperatures.

This tendency is confirmed by the analysis of power consumption in the area of operation set at the time of durability tests performed for extreme work temperatures (Fig. 11).

In this case, similar tendencies could be observed as well. At the time of tests conducted at the temperature of 253K, an increase in the mean power consumption was recorded both at the time of opening (12%) and closing (42%).

The tests performed also enabled the assessment of the efficiency of the operation of the overload protection used in the control system of the tested actuators. In some cases, at higher temperatures, the activation of the overload protection was observed, even when there was no increase in load over its nominal value. This constitutes an error, which totally excludes the possibility to use an actuator in applications were higher temperatures may occur as a result of fire.

10. Conclusions

Providing the required level of safety enabling reliable activation of the smoke extraction installation and the evacuation of people is possible thanks to the use of electric actuators characterised by a stable level of electricity consumption, regardless of the temperature of the surrounding. In case of fire, this feature is extremely important particularly due to the fact that together with the increase in temperature, the electrical conductivity of wires decreases, which results in an excessive voltage drop.

The test stand used for the experiment enabled the performance of the tests for typical electric actuators of different architecture in conditions of simultaneous power and thermal exposure. The registered characteristics document the existing differences in work parameters of individual actuators. The induced thermal exposure enabled the determination of the range of parameter changeability, depending on the temperature of the work environment.

The results obtained for the tested drives indicate the lowest power consumption for drives at the temperature of 293 K. When the temperature drops, a significant increase in power consumption can be observed, while at increased temperatures, the power consumption increases slightly only in the case of rotary and spindle drives.

It was noticed, that the time of activation of all tested objects shortens together with the increase in temperature.

The increased power consumption at lower temperatures can be dangerous particularly in the case of multi-storey buildings, when the fire breaks during winter time on its lower levels. In such a situation, cold actuators of smoke extraction flaps located on a building's roof or attic may require such parameters of the electric power, which cannot be provided by the installation, which due to its overheating in the area of fire, is no longer efficient.

Another particularly valuable element of the tests conducted were the results of the effectiveness of work of overload protection. An in-depth verification of this significant parameter can be conducted only in the case of tests performed at increased temperatures.

Based on the tests performed and the results recorded for three, out of many commercially available actuators, it is therefore reasonable to conduct extended certification tests, in which operating characteristics of the actuators are recorded on-line, simultaneously for different temperatures. This constitutes a significant modification to the existing regulation, according to which characteristics are recorded only for the temperature of the surrounding after the heating or cooling of a tested actuator.

The main direction of works necessary for maintaining the reliability of smoke extraction systems is the diagnosis of complete systems in conditions and the scale resembling their real application conditions, which was provided by the developed test stand and methodology.

The test results recorded confirm the possibility of additional application of the stand as a device for verification tests for diagnosis and prognosis systems. The implementation of additional acceleration sensors extending the diagnosis possibilities of the stand will be the scope of further research concerning the development of the scientific works undertaken.

Due to the unique possibilities of reconstructing the changeable conditions of temperature and humidity, the stand can also be used to test actuators used in aviation, maritime and land transport.

Acknowledgements: *Tests were performed within the framework of the "Innovative Systems of Technical Support for Sustainable Development" Strategic Programme POIG.01.01.02-14-03/09 UE/NCBiR*

References

1. Balaban E., Bansal P., Stoelting P., Saxena A., Goebel K., Curran S.: A diagnostic approach for electro-mechanical actuators in aerospace systems. Proceedings of the Aerospace conference IEEE, March 2009: 7-14
2. Balaban, E., Saxena, A., Goebel, K., Byington, C., Watson, M., Bharadwaj, S., Amin, S. Experimental data collection and modeling for nominal and fault conditions on electro-mechanical actuators. Proceedings of the Annual Conference of the Prognostics and Health Management Society 2009: 1-15.
3. Balaban, E., Saxena, A., Narasimhan, S., Roychoudhury, I., Goebel, K., & Koopmans, M.: Airborne Electro-Mechanical Actuator Test Stand for Development of Prognostic Health Management Systems. Proceedings of Annual Conference of the PHM Society. October 10-16, 2010: 1-13.
4. Balaban, E., Saxena, A., Narasimhan, S., Roychoudhury, I., Goebel, K. Experimental Validation of a Prognostic Health Management System for Electro-Mechanical Actuators. The American Institute of Aeronautics and Astronautics AIAA Infotech@ Aerospace Conference, 2011: 1-12.
5. Bodden, D., Clements N., Schley B., Jenney G.: Seeded Failure Testing and Analysis of an Electro-Mechanical Actuator. Proceedings of the Aerospace Conference IEEE, March 2007: 1-8.
6. Byington, C., Watson, M., Edwards D.; Stoelting P.: A model-based approach to prognostics and health management for flight control actuators. Proceedings of the IEEE Aerospace Conference, 2004: 3551 -3562.
7. Dziennik Ustaw - rok 2007 nr 143 poz. 1002.
8. Gałaj J., Jaskółowski W., Karpovič Z., Šukys R., Badanie wpływu impregnacji ogniochronnej na skład i ilość produktów toksycznych powstałych podczas spalania drewna sosnowego. Bezpieczeństwo i Technika Pożarnicza 2011; 3: 55-62.
9. Garinei A., Marsili R.: A new Diagnostic Technique for Ball Screw Actuators. Measurement 2012; 45: 819–828.
10. Granovskiy E. A., LyfarI V. A, Vorona A. P., Barbuca M., Jarosz W., Modelowanie ewakuacji podczas pożarów budynków - analiza ryzyka. Bezpieczeństwo i Technika Pożarnicza 2011; 4: 59-63.
11. He Yaping, Wang Jian, Wu Zhenkum, Hu Lu, Xiong Yi, Fan Weicheng: Smoke venting and fire safety in an industrial warehouse. Fire Safety Journal 2002; 37: 191-215.
12. Jensen S., Jenney G., Dawson D.: Flight test experience with an electromechanical actuator on the F-18 Systems Research Aircraf. Digital Avionics Systems Conference DASC, vol. 1, 2000: 1-10.
13. Jie Ji, Kaiyuan Li, Wei Zhong, Ran Huo. Experimental investigation on influence of smoke venting velocity and vent height on mechanical smoke exhaust efficiency. Journal of Hazardous Materials 2010; 177: 209-215.
14. Koopmans M., Hooven R., Tumer I. Reliability based design recommendations for an electromechanical actuator test stand". Annual Conference of the Prognostics and Health Management Society, 2010: 1-15.
15. Koopmans, M. T., Tumer, I. Y. Electromechanical Actuator Test Stand Coupling Design to Support Actuator Prognostic Model Development. Proceedings of the ASME Design Engineering Technical Conferences, 2011: 1285-1296.
16. Koopmans, M. T., Tumer, I. Y. Function-based analysis and redesign of a flyable electromechanical actuator test stand. Proceedings of the ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, 2010: 1-12.
17. Kozioł S., Matras E., Zbrowski A.: Modular system for testing the stability and safety of specialised height rescue vehicles. Solid State Phenomena 2013; 199: 55-60.
18. Kozioł S., Zbrowski A.: Metoda poprawy bezpieczeństwa budynków poprzez zastosowanie odzysku ciepła w układach wentylacji pożarowej. Bezpieczeństwo i Technika Pożarnicza 2012; 3: 39-45.
19. Kubicki G.; Wiche J.: Safety Way® – pierwszy inteligentny system ochrony przed zadymieniem dróg ewakuacji. Ochrona Przeciwpożarowa

2010; 3: 24-26.
20. Longoria, R. G., LeSage, J., Shutt W.: Modeling and Requirements Formulation for Submarine Control Surface Actuation Systems. American Society of Naval Engineers Day 2010: 8-9.
21. Matuszewski P, Żybura R.: Testowanie przeciwpożarowych klap odcinających a bezpieczeństwo budynku. Chłodnictwo i Klimatyzacja 2010; 6: 67-68.
22. Mazur D. Elektryczne napędy - nowe możliwości. Świat Szkła 2006; 1: 24-46.
23. Mazurkiewicz A. (kierownik): Projekt Strategiczny POIG.01.01.02-14-03/09 UE/NCBiR „Innowacyjne systemy wspomagania technicznego zrównoważonego rozwoju gospodarki".
24. Mizieliński B. Systemy oddymiania budynków, WNT, Warszawa 1999.
25. Narasimhan S., Roychoudhury I., Balaban E. , Saxena. A.: Combining Model-Based and Feature-Driven Diagnosis Approaches A Case Study on Electromechanical Actuators. 21st International Workshop on Principles of Diagnosis, 2010: 1–8.
26. Norma EN 12101-2: Systemy kontroli rozprzestrzeniania się dymu i ciepła, cz.2: Wymagania techniczne dotyczące klap dymowych.
27. Owens K i inni: Electromechanical Actuator Test Apparatus. US Patent No 7775120B2. 17 Aug, 2010.
28. Półka M., Analiza toksyczności produktów rozkładu termicznego i spalania uzyskanych z wybranych materiałów epoksydowych. Bezpieczeństwo i Technika Pożarnicza, 2010; 3: 73-82.
29. Samborski T., Zbrowski A., Kozioł S.: Modelowy system do badań osłon przeciwwietrznych z czujkami dymu. Technologia i Automatyzacja Montażu 2012; 3: 33-39.
30. Sawicki J. Wytyczne sterowania urządzeniami przeciwpożarowymi. Bezpieczeństwo i Technika Pożarnicza nr 2007; 2: 75-90.
31. SFPE Handbook of Fire Protection Engineering, 2008 Edition.
32. Sitarz G.: Zastosowanie pneumatyki w budownictwie - Klapy Dymowe. Pneumatyka 2013; 1: 26-27.
33. Skiepko E. Ciągłość dostawy energii i przekazu sygnału w instalacjach funkcjonujących w warunkach pożaru. Elektro Info 2012; 10: 12-57.
34. Skiepko E. Instalacje elektryczne funkcjonujące w warunkach pożaru. Elektro Info 2006; 3: 54-57.
35. Skiepko E. Wymagania dla instalacji elektrycznych funkcjonujących w czasie pożaru. Elektro Info 2009; 4: 60-66.
36. Skiepko E. Zasady projektowania sterowań instalacji do odprowadzania dymu i ciepła. Elektro Info 2008; 1-2: 82-88.
37. Sypek G., Wiche J. Temperatura a systemy ochrony przed zadymieniem pionowych dróg ewakuacji. Ochrona Przeciwpożarowa 2009; 12: 39-40.
38. Tesar D., Krishnamoorthy G. Intelligent Electromechanical Actuators to Modernize Ship Operations. Naval Engineers Journal, 2008; 120: 77-88.
39. Wiatr J., Jaskółowski W. Wpływ pożaru na wartość napięcia zasilającego urządzenia elektryczne, które muszą funkcjonować w czasie pożaru. Elektro Info 2012; 1-2: 76-83.
40. Wiatr J., Skiepko E. Zasady doboru przewodów elektrycznych w instalacjach oddymiających. Elektro Info 2011; 3: 80-85.
41. Wiatr J. Wpływ temperatury pożaru na wartość napięcia zasilającego urządzenia elektryczne oraz skuteczność ochrony przeciwporażeniowej urządzeń, które muszą funkcjonować w czasie pożaru. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska 2012; 2: 37-44.
42. Wicks E. System and method for smoke free elevator shaft. United States Patent 5718627. Feb. 17, 1998.
43. Wierciak, J., Szykiedans, K., Rawski, J., Lisicki J. System pomiarowy do wyznaczania charakterystyk elektrycznych siłowników liniowych. Pomiary Automatyka Kontrola 2007; 9bis: 527-530.
44. Zbrowski A., Samborski T., Kozioł S. Research and test apparatus for certification testing in technical rescue systems. Problemy Eksploatacji 2010; 3: 127-138.

Andrzej ZBROWSKI
Tomasz SAMBORSKI
Institute For Sustainable Technologies. National Research Institute
ul. Pułaskiego 6/10, 26-600 Radom, Poland
E-mails: andrzej.zbrowski@itee.radom.pl tomasz.samborski@itee.radom.pl

Article citation info:
MAZURKIEWICZ D. Computer-aided maintenance and reliability management systems for conveyor belts. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 377–382.

Dariusz MAZURKIEWICZ

COMPUTER-AIDED MAINTENANCE AND RELIABILITY MANAGEMENT SYSTEMS FOR CONVEYOR BELTS

SYSTEMY INFORMATYCZNEGO WSPOMAGANIA EKSPLOATACJI I ZAPEWNIENIA NIEZAWODNOŚCI PRZENOŚNIKÓW TAŚMOWYCH

Operational reliability of conveyor transport systems is a problem that relates to the provision of an adequate level of availability of a continuous transport system, which in the case of belt conveyors depends not only on their usability, as determined by their design and manufacture, but also on the appropriate level of their use, understood, among others, as the degree of sophistication and effectiveness of the methods and tools applied in industrial diagnosis of those devices. The present article describes the most popular diagnostic systems used in the maintenance of internal transport conveyor systems. Also a new method of computer-aided maintenance of such systems is presented.

***Keywords**: Computer-aided Maintenance Management Systems, belt conveyors, maintenance.*

Niezawodność eksploatacyjna systemów transportu przenośnikowego, to zagadnienie dotyczące zapewnienia odpowiedniego poziomu gotowości systemu transportu ciągłego, który w przypadku przenośników taśmowych zależny jest nie tylko od ich walorów użytkowych, determinowanych zastosowanymi rozwiązaniami konstrukcyjnymi i wykonawstwem, ale też zależy w istotnym stopniu od właściwego poziomu ich użytkowania, rozumianego między innymi jako stopień zaawansowania i skuteczności zastosowanych metod i narzędzi diagnostyki przemysłowej. W artykule scharakteryzowano najpopularniejsze systemy diagnostyczne stosowane w eksploatacji przenośnikowych systemów transportu wewnątrzzakładowego. Zaprezentowano również nową metodę komputerowego wspomagania w utrzymaniu ruchu tego typu systemów.

***Słowa kluczowe**: Computer Aided Maintenance Management Systems, przenośniki taśmowe, eksploatacja.*

1. Introduction

Modern transport systems based on belt conveyors are often very complex, in terms of both technology and design, which translates directly into their costliness. At the same time, the performance of a conveyor-based system determines not only the efficiency of a company's entire transportation system, but also the efficient and effective operation of other, cooperating subunits of a company's production system. In many cases, belt conveyor transport and cooperating subsystems have a serial structure. This means that a failure of a belt conveyor, which is one of the basic components of a transport system of this type, brings all other machinery and equipment to a stall. Therefore, the problem of ensuring an adequate level of availability of a continuous transport system is extremely important. Availability, in the case of belt conveyors, depends not only on their usability, as determined by their design and manufacture, but also on the proper level of utilization, understood, among others, as the degree of complexity and effectiveness of the methods and tools employed for their industrial diagnosis.

2. Diagnostic systems in the maintenance of conveyor systems for internal transport

The application of methods of technical diagnosis in the maintenance of belt conveyors is becoming an important issue from the point of view of effective utilization of those machines. However, methods of technical diagnosis are still underused in the case of belt conveyors and internal transport systems based on them. This is particularly important as modern means of information transfer make it possible to effectively monitor the state of machinery and equipment, and other computer-aided methods and techniques enable processing of the data collected using inference and prediction methods.

Typical monitoring of the condition of a machine involves measurement of the physical parameters of the work process that guarantee correct and safe operation and user safety. In the case of belt conveyors, monitoring typically encompasses the following parameters or characteristics: belt speed, belt run-off, belt wear, temperature of the elements that cooperate with the belt, the filling status of the transfer section, detection of metal components in the material conveyed, rotational speed of the drum, and conveyor performance. In general terms, then, monitoring of belt conveyors relates to diagnostic symptoms and technological, operating and safety parameters [2, 5]. The current condition of a belt conveyor, in the context of effective management of this means of transport, and especially its belting, can be determined using a variety of methods, tools and systems. The most effective and efficient among those is diagnosis of the elements of a belt transport system. Therefore, in modern conveyor systems, the parameters that characterize the operation of the whole system, the conveyor, or the belt itself are usually monitored continuously using mobile or stationary measuring systems. In the case of mobile systems, the results of measurements can indicate that the conveyor operates properly, but they may also point to errors in design assumptions, calculations and selection of conveyor components, which result in decreased durability of the conveyor and its components as well as a lower reliability of the whole structure [7]. Another group of diagnostic methods make use of computer techniques to control a single parameter or component whose condition has a significant impact on the reliability of the whole device and the entire belt conveyor system. A good example is a method used for monitoring the core of a steel cord conveyor belt [15] that utilizes an apparatus for magnetic monitoring of damage. This system is used to monitor the condition of the belt core directly on the conveyor belt, and it works by measuring and recording changes in the magnetic field resulting from discontinuities in the belting, i.e. locations at which the steel cords of the core have been damaged. Another system that works in a similar manner is a system for monitoring and detection of damage of reinforced conveyor belts [8]. The system generates a static magnetic field with a circuit that is closed by the reinforcing cords at the measuring sections along the belt and in the plane perpendicular to its surface. The system works by measuring magnetic field strength, processing the measurement signal, and analyzing the damage by means of an electronic data processing technique. Other diagnostic methods include radiographic examinations of fabric and steel cord belting for breaks and corrosion and ultrasonic rip detection.

One of the few methods that have so far been developed to prevent the occurrence of some of the types of belt failures is one in which special sensors are installed in the belting which signal belt damage or wear; these sensors work on the premise of transponders – electronic devices used in active radar systems. A transponder receives radio signals from one (transmitting) system and then collects, processes and amplifies them to transmit to another (receiving) system. A system that works in a similar way is Sensor Guard [13]. However, it only serves to protect belting against extensive rip damage and signals lateral movement and excessive slippage of the belt. In modern transport systems, it is a common trend to install complete computer-assisted monitoring systems. One such system is ZEFIR NT that is used to monitor a system for transporting rock materials [6]. Its task is to collect, analyze and visualize data related to the operation time of the individual devices, the number of start-ups, the length of outage periods, etc. Unfortunately, this system does not perform any diagnostic functions, and the data it generates can only be analyzed with a considerable delay.

The systems described above and other, similar ones, are used to control certain parameters related to the failure rate of conveyor belts, but they do not protect belts against sudden breaking in the area of the splice. They do not represent comprehensive monitoring and control systems, and the data collected with their help are often analyzed manually. For these reasons, it is necessary to develop appropriate, fully automated systems for controlling internal transport, including conveyor-based transport. A number of types of damage that may be incurred by a conveyor belt and its joints necessitate the use of advanced and maximally versatile diagnostic systems. Efficient and effective diagnosis of belt conveyors is also essential for continuous assessment of the state of belting aimed at preventing major failures and thus prolonging the conveyor’s service life.

This means that it is necessary to implement into industrial practice a fully automated system for diagnosing internal conveyor transport, especially that the costs incurred as a result of losses due to belt damage are still very high. The use of simplified, manual analyses is time consuming and inefficient, especially in the era of the dynamic development of computer maintenance management systems (CMMs). Inefficient diagnosis may lead to reduced availability of a transport system, extend repair and maintenance time and increase maintenance costs. This is especially important when a conveyor belt breaks at the splice. Research on the design of belt conveyors and the structure of their load-bearing elements pays insufficient attention to conveyor belt splices, whose strength significantly affects the durability and reliability of conveyors, and even the entire systems that they are part of. For this reason, the present author has devoted the past several years to research and implementation work [9–13] aimed at analyzing the strength properties of bonded joints and developing and implementing a concept of a system for monitoring the condition of such joints during operation of a conveyor belt. As a result of this work, a monitoring system (Fig. 1) was designed and built, which consists, among others, of a magnetic field sensor and a pulse counter that make up a computer measuring system. A modem, RAM, and a microprocessor constitute its control system. The data transmission and processing system comprises a personal computer and phone lines or network cabling. Measurement points are marked with permanent magnets embedded in the conveyor belt on opposite sides of each splice. The location of the magnets during conveyor movement is identified by a semiconductor magnetic field sensor.

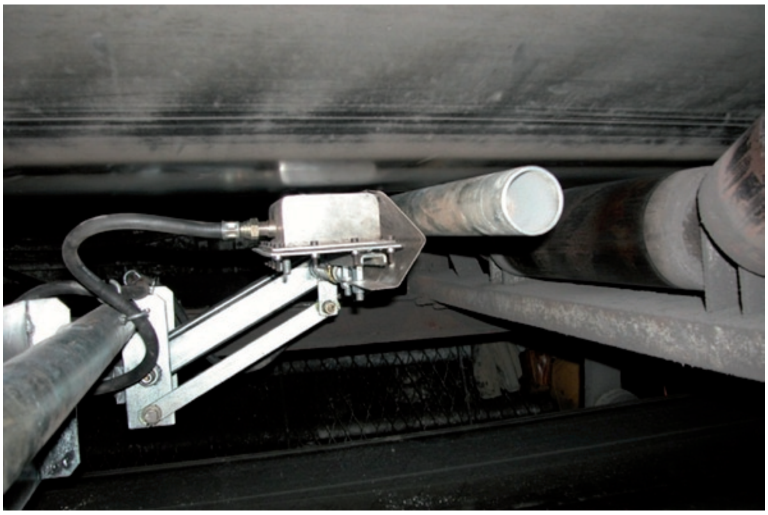


Fig. 1. Measuring–diagnostic system [13]

The principle of operation of the system is based on the appropriate use of the magnetic sensor located under the moving conveyor belt. Magnetic field strength increases as the magnet approaches the sensor; and when the magnet passes the sensor, the sense of magnetic field strength changes. At this point, the strength of the magnetic field changes the most in the shortest time. A pulse registration method that works in this way makes it possible to measure the time between two measuring points at a high accuracy. The measurement data obtained are sent to a computer equipped with suitable software for acquisition and analysis of data and visualization of results. The original monitoring system designed on this premise can signal in advance the occurrence of conditions that normally accompany an impending break in a belt splice. The monitoring system is designed in a way that endows it with powerful data analysis capabilities and offers additional benefits, allowing a specialist to:

- analyze those maintenance parameters of the belt conveyor that contribute to frequent failures due to belt breaks and to develop an appropriate intelligent system for monitoring the operating conditions of the device and for its automatic control aimed at eliminating critical situations;
- analyze, on the basis of the results of measurements conducted under actual conditions of use, the durability of different types of bonded joints used in conveyor belts to enhance the quality of the joints and develop or modify the technology of their manufacture;
- carry out analyses of the causes of defects in the manufacture of belt splices;
- and make a practical comparison of the peel strength and shear strength properties of adhesive materials used for splicing belts.

Measurement of the length of a conveyor belt splice begins when the magnetic field sensor sends a message that it has detected a reference point. This point marks the beginning of the measurement section and it resets the pulse counter. Splice length is measured indirectly. The moving belt turns a roller which is located under the upper section of the conveyor belt in such a way that there is permanent pressure between the roller and the moving belt, which prevents slippage between these two elements. Attached to the roller is a disk with evenly spaced indentations around its circumference. The pulse counter system uses a slotted optocoupler, which is a system that consists of two photoelements. One of them emits light beams (this is usually an LED), and the other is a photodetector, which converts the incoming light beams into electric voltage. The photodetector is a photodiode or a phototransistor. One of the disk slots is positioned between the two elements of the optocoupler. If the element that finds itself in the slot of the disk is impervious to light beams, a change in the voltage at the photodetector occurs (the voltage may rise or fall, depending on the electric system in which the photodetector operates). The moving belt turns the roller, which has a defined, constant diameter. The disk mounted on the roller has 200 holes (slots) around its circumference. This allows one to determine the distance travelled by the conveyor belt over the time when the disk has rotated by an angle equal to the rotation of the disk by one slot.

The optocoupler is used to count the passing holes (slots). The impulses start to be counted when a signal is received from the magnetic field sensor that the reference point has been detected. The counter starts counting the impulses coming from the optocoupler. Measurement ends when the magnetic field sensor sends a signal that it has detected a second reference point, which marks the end of the measuring section, i.e. the joint or section of the belt measured at a given time. The value of the counter is stored in the RAM of the control system and the counter is reset to zero. A microprocessor located in the pulse counter circuit controls the flow of data between the optocoupler, and the control system.

An important element of the discussed computer-assisted measurement system is its control system. The control system of the measurement system for monitoring instantaneous changes in the length of conveyor belt splices consists of three components – a microprocessor, RAM and a modem. The microprocessor controls data flow between the measurement system and the control system. Data incoming from the pulse counter circuit are transferred to RAM. Each measurement is stored according to a specified pattern. Readout of the pulse counter is saved along with the hour, minute, second and the date of receipt of the data by the control system. The control system has been equipped with a large storage capacity because of the substantial amount of incoming data and the high rate of transfer of these data by the user to a PC. The large number of measurement data recorded is a consequence of the high-speed at which the conveyor belt moves and the large number of splices, whose passage through the measuring system is recorded many times a day.

For measurements of this type, a separate analysis of static and dynamic errors has to be conducted. Static error is independent of the time-varying nature of the value measured and it is not associated with transient processes in the device. Since ideal static conditions do not exist in practice, it is assumed that changes in the value of the parameter measured are at the level of resolution of the measuring instrument. Static error of digital measurement is comprised of quantization error and analog error. Quantization error is a consequence of mapping an infinite set of continuous (analog) values onto a finite set of quantized digital values. Analog error is more complex. It consists of the threshold error of the system which is unable to respond to infinitely small changes in the parameter measured. Reference source error is usually the main component of analog error and it results from the imperfect nature of the reference. The last component of analog error is the error associated with noise and other types of interference in the measurement circuit [1, 3, 14].

Dynamic error arises when the value measured changes over time. Dynamic errors are the difference between the value measured and the actual value. The main components of this error are quantization errors associated with sampling, and transient states of the measuring instrument. Dynamic error is produced when a continuous value is averaged to discrete time intervals, when the sampling frequency is too low, i.e. lower than the frequency of changes of the analog signal, and when the instrument is in a transient state, for example, as a consequence of a change in the direction of the magnetic field in the magnetic field sensor.

Digital filters used in the proposed measurement system are designed to smooth out the signal, separate the relevant components of the frequency spectrum, and determine parameters in different frequency ranges. When designing a digital filter, care should be taken that the sampling frequency of the filter is higher than any component frequency of the signal at the input of the ADC. Another assumption is that the sampling period must be shorter than any time constant in the filter. This is to prevent a situation in which the analog signal ceases to be approximated by the digital signal. A digital filter must be sufficiently precise, i.e. it must have a low degree of quantization. If the degree of quantization of a filter is greater than the input of the ADC upstream of the filter, some information will be lost due to slower data-processing upstream of the filter. An unwanted effect in digital filters is clipping of the signal. It can lead to oscillation in the filter, which is a cause of errors [1].

Another problem related to measurement are the distortions arising from the nature of the environment in which the measurement system works. They give rise to signals in the measurement circuit that are not the result of the measurement process. The distortions are formed as a result of electromagnetic fields acting on the measuring system. Systems that are most liable to disturbances are those located in the vicinity of large electrical machines and those that are surrounded by a dense network of electric wires, as is usually the case with measurements made in industrial environments. Particularly sensitive to disturbance are analog systems. The continuous output signal can easily be deformed, especially when the voltage in the analog circuit is low, e.g., several millivolts. In this case, a distortion of the order of one millivolt can lead to an error between 10% and 20%. Measurement systems can be protected against this type of interference by shielding.

3. Computer-assisted analysis of measurement data and computer-aided maintenance

As noted, among others, by Kaźmierczak [4], the decision-making process for maintenance activities can and should be effectively aided by using appropriate computer tools. One type of such tools are Computer-aided Maintenance Management Systems (CMMS), which are developed to perform functions that go beyond typical database maintenance. An optimal CMM system should also encompass elements of the decision making process, including, among others [4]:

- effective collecting, processing and sharing of data stored in database systems using diagnostic methods for evaluating the state of an object,
- development of a knowledge base to complement the diagnostic data,
- designing an advisory system tool to support decisions on the state of the object using the information stored.

Identical assumptions were adopted for the design of software modules for the measuring system described in this article. The main functions of this software are recording and processing of process data, control of measuring equipment and communication with the user. This is why in this program the tool changing column has been placed beside the data window. The tool buttons are used to activate various functions of the program, such as setting the compensation of the magnetic sensor or importing data to the user's computer. The program that supports the measurement system described in this article offers, among others, the possibility of visualizing changes in the length of a particular splice during operation of a conveyor, visualizing changes in the length of each section of the belt between the joints and collective presentation of these values for all (or selected) splices of the conveyor and all sections of the belt between the splices. It also enables statistical analysis of the values recorded and export of measurement data, graphics, etc.

The measurement data and their analysis using the proposed computer program allow to visualize and comprehensively analyze measurement results, and the data obtained in this way allow to continuously monitor the status of all splices and sections of the belt between them. The program can signal an alarm condition and helps evaluate the operating conditions of the conveyor, identify a single splice any time during the operation of the conveyor belt, and assess the effectiveness of maintenance work – for example, strengthening of a joint with mechanical components. This means that the measurement system can signal in advance the occurrence of conditions that normally accompany an impending break in a belt splice. The appropriate design of the system and the dedicated program for data analysis also afford additional benefits arising from the application of the system, which include, among others:

- continuous measurement of changes in the length of all splices and belt sections between them to within 1 mm;
- about 1500 measurements per day for each conveyor, depending on the number of its splices;
- real-time visualization of changes in splice length,
- identification of splices and pinpointing of their exact location on the conveyor any time during its operation;
- analysis of historical data – the possibility of collecting information on each splice and belt section (date and place of manufacture, manufacturer's identity, operation time, load, place of operation, etc.);
- the possibility of integrating the system as a module into other diagnostic–visualization systems, especially in self-management systems (Fig. 2);
- availability of accurate information on the operating time of the conveyor, its instantaneous speed values, downtime periods, number of stops and times of restart, and the precise length of the conveyor belt at any time of its operation;
- data processing enabling predictive control of the operation of the conveyor belt.

It is due to these features that the system can be classified as a diagnostic system of the second level of security of the transport process, which is a higher level compared to the previously described conventional systems of locks and guards that perform the tasks of a lower level of security. Not without significance is also the possibility the system gives of long-term analysis of historical data for every single joint that forms part of the conveyor belt from its manufacture until the end of its useful life.

According to the concept of CMM, the monitoring device can also be transformed into an intelligent machine that has the ability to independently respond to the changing operating conditions and to eliminate conditions that cause belt breaks by anticipating future operating parameters and the consequences of their occurrence.

As it follows from previous analyses, both quality- and quantity-oriented knowledge of changes (a decrease) in the value of strength parameters of belt sections and splices during the entire maintenance cycle is indispensable when it comes to choosing a belt for a given conveyor and predicting the optimum time point for performing maintenance, based on the precisely specified operation time and load distributions. It may also be useful for the assessment and verification of newly developed mathematical models of operation time distribution.

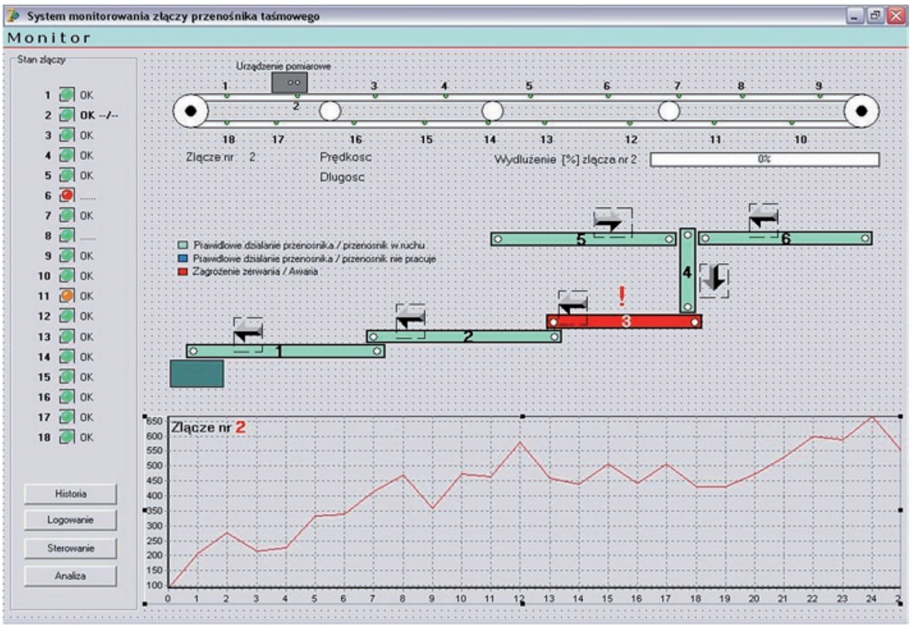


Fig. 2. A conception of a method for visualizing data from a system for monitoring conveyor belt joints

Therefore, bearing in mind that conveyor belts are commonly used in internal transport systems, and that there are no effective monitoring and control systems for preventing failures of these devices caused by splice breaks, it can be concluded that the results of the present study will not only increase the knowledge and experience of industrial use of diagnostic and monitoring systems, but will also reduce losses associated with downtime due to unexpected failures.

It can thus be stated that the objective of the present study is to design not only a comprehensive dispatching monitoring system, but also a smart actuator, which, when integrated into currently used transport monitoring and control systems, will allow to effectively expand them and make fuller use of diagnostics in self-management systems. The measurable benefits of implementing this concept of monitoring and controlling of belt conveyors include, among others:

- the possibility of constant surveillance of all conveyor belt splices during operation, using a non-invasive method;
- identification of joints and assessment of their durability and reliability throughout their operating life, regardless of their current location;
- real-time data mining of monitoring data along with a comprehensive analysis of historical data, which are especially useful in belt management;
- simple design and easy operation that does not require outsourcing;
- the possibility of assessing the effectiveness of maintenance activities.

The conception of a system for monitoring conveyor belt joints developed in this study corresponds, in accordance with the theory of diagnosis, to real time diagnosis of an object with simultaneous signalling and visualization of its state. Implemented in this way, the monitoring method yields a real-time diagnostic system for conveyor transport. This system collects during its operation a considerable amount of different types of data, which provide information on the object being diagnosed. This excess of information, which operators have to analyze using visualization of the current parameters characterizing the analyzed object, may eventually lead to an information overload, often resulting in ignoring the relevant information or a failure to use the potential that measurement data collected by the system over a long time have. The monitoring system itself, then, despite a number of undeniable advantages, is not, in the long run, a sufficient solution to the optimal use of information it generates.

That is why it has to be expanded to become a fully automated computerized system for supporting the diagnosis of belt conveyors. Its primary role should be to process collected data and analyze them to support the decisions made by supervisory personnel or to take such decisions automatically, as far as possible, and to utilize them for current or temporary control of the transport system.

In this way, the system for monitoring conveyor belt splices and the belting between them would become a system of continuous surveillance (supervision and safeguarding) of the state of the analyzed object. Depending on its application, it could be used to support decision-making to ensure proper operation of the object (Fig. 3) or to control the object to prevent its failure and the consequences thereof (Fig. 4).

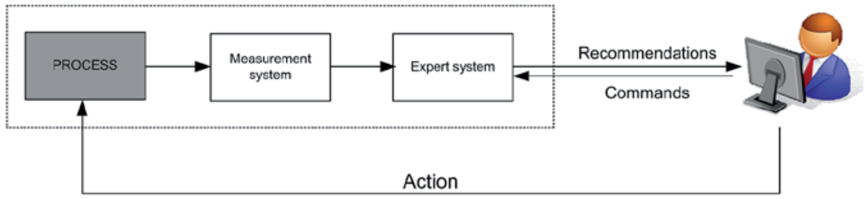


Fig. 3. The concept of an expert decision support system for conveyor belt transport

This is in line with diagnostic terminology, in which supervision stands for monitoring of an object and taking steps to maintain its proper operation, and safeguarding means the actions and technical measures taken to eliminate a potentially dangerous course of a process or to prevent the effects of such a course.

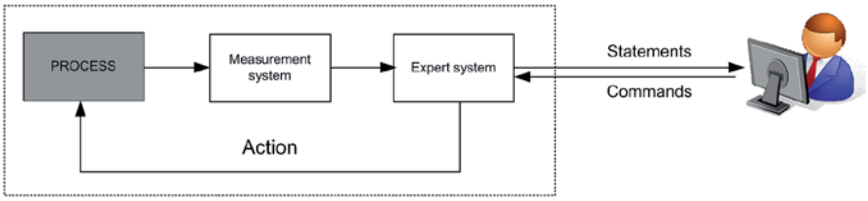
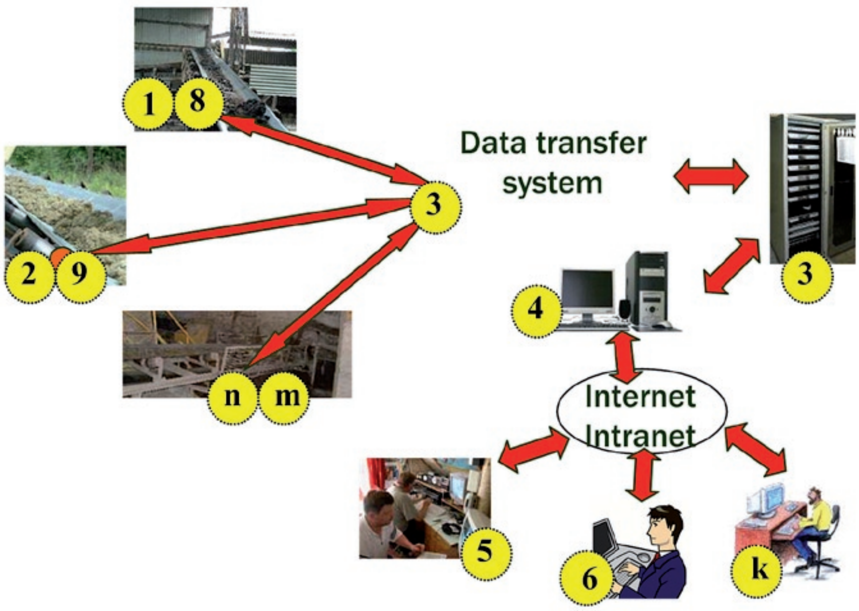


Fig. 4. The concept of a preventive control system for conveyor belt transport



Therefore, based on the results of the work carried out so far, it is also possible to build a comprehensive system of continuous surveillance of conveyor belts that could be used in company transport systems (Fig. 5). Such a system would combine diagnostic equipment for analyzing the reliability-relevant operation parameters of the individual conveyors with elements of a computer-based advisory and control system.

The primary objective of the system will be to automatically monitor the status of all bonded joints to prevent their rupture, and to track the operation history of all belts and their individual sections. A system of this kind is versatile in that it can be used in different types of companies, in accordance with their specific branch-related characteristics and requirements. Thanks to its versatility and the flexibility of its software modules, the system can be combined with other dispatching diagnostic and monitoring systems.

One problem related to operating this type of system is that it requires collection and processing of various types of data, which are often difficult to describe or analyze in real-time. It would therefore be important to expand the proposed splice monitoring system to a diagnostic advisory system forming part of a comprehensive diagnostic system based on intelligent tools.

4. Conclusions

The substantial amounts of data and accompanying information generated by a measuring system, particularly in the case of a large number of monitored bonded joints located on several, even up to 20, conveyors of a complex transport system, lead to numerous complications associated with the interpretation of collected data, making it difficult for supervisory personnel to make effective decisions. In such cases, the operator is faced with the necessity for constant decision making, having to analyze a large data set, observe changes in the data or predict their future values (the trend of changes on a safety scale of safe / emergency / above alarm level, which corresponds to the potential states of the analyzed object). In other words, the decisions have to be made under uncertainty [235, 237]. A conveyor transport system is an example of a system that is affected by many factors which have not been clearly defined or which are characterized by significant randomness and unpredictability. Objects of this type are difficult to describe.

Maintenance diagnosis of a single belt conveyor or an entire transport system conducted with the help of the described monitoring system enables effective assessment of the current state of bonded joints. When first symptoms indicative of reduced strength, and, hence, potential rupture of a joint, are registered, however, a monitoring system should additionally be able to predict for how long the joints will retain the characteristics that ensure proper operation of the conveyor. An ideal, newest-generation

Fig. 5. A schematic diagram of the concept of a comprehensive system for monitoring belt conveyor transport: 1, 2, ..., n – measuring devices; 3 – a long-distance data transmission system; 4 – software for real time analysis and collecting of measurement data, processing of historical data and controlling the system; 5, 6, ..., k – end-user interfaces compatible with the company's dispatching system; 8, 9, ..., m – signalling devices.

system should not only be able to collect and process current and historical measurement data and generate on their basis justified alarm signals, advisory information and guidelines for personnel, but also take corrective actions on its own. In other words, a distinguishing feature of such a monitoring system would be the ability to carry out an inference process with the use of expert systems, that would automatically generate and verify diagnoses to be used by the operator, and that would also have the ability to take action independently of the operator. To create a system of this type, it is important to develop diagnostic methods based on decision support systems and the latest developments in the field of intelligent techniques, which is also a significant challenge for the diagnostic system described in this paper.

References

1. Bilgili E. Modelling mechanical behaviour of continuously graded vulcanized rubbers. Plastics, Rubbers and Composites 2004; 33 (4): 163-169.
2. Błażej R, Hardygóra M, Komander H. Wpływ wybranych parametrów na trwałość zmęczeniową połączeń wieloprzekładkowych taśm przenośnikowych. Transport Przemysłowy 2002; 3 (9): 5-9.
3. Curley AJ, Hadavinia H, Kinloch AJ, Taylor AC. Predicting the service-life of adhesively-bonded joint. International Journal of Fracture 2000; 103: 41-69.
4. Kaźmierczak J. Eksploatacja systemów technicznych. Gliwice: Wydawnictwo Politechniki Śląskiej, 2000.
5. Kuzmenko VI, Kiriaz'ev PM. Developing high-strength joints for conveyor belts". Technical, Technological and Economic Aspects of Thin-Seams Coal Mining. International Mining Forum 2007. Edited by E. J. Sobczyk and J. Kicki. London: Taylor & Francis Group, 2007.
6. Kwon Y, Kwon H, Kim W, Leo S. Estimation of rubber material property by successive zooming genetic algorithm. Journal of Solid Mechanics and Materials Engineering 2007; 1 (6): 815-826.
7. Larose DT. Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych. Warszawa: Wydawnictwo Naukowe PWN, 2006.
8. Lesiak P, Świsulski D. Komputerowa technika pomiarowa na przykładach. Warszawa: Agencja Wydawnicza PAK, 2002.
9. Mazurkiewicz D. Problems of numerical simulation of stress and strain in the area of the adhesive-bonded joint of a conveyor belt. Archives of Civil and Mechanical Engineering 2009; 2: 75-91.
10. Mazurkiewicz D. Analysis of the ageing impact on the strength of the adhesive sealed joints of conveyor belts. Journal of Materials Processing Technology 2008; 208: 477-485.
11. Mazurkiewicz D. Computer system for monitoring conveyor belt joints. Canadian Mining Journal 2007; 5: 24.
12. Mazurkiewicz D. Tests of Extendability and Strength of Adhesive-Sealed Joints in the Context of Developing a Computer System for Monitoring the Condition of Belt Joints during Conveyor Operation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2010; 3: 33-38.
13. Mazurkiewicz D. A study of selected aspects of operational diagnosis of belt transport. Lublin: Wydawnictwo Politechniki Lubelskiej, 2011.
14. Packham D. E. Handbook of adhesion. Chichester: John Wiley & Sons, 2005.
15. Standard PN-EN ISO 7622-2:2004. Taśmy przenośnikowe z linkami stalowymi. Badanie właściwości trakcyjnych. Część 2: Pomiar wytrzymałości na rozciąganie.

Dariusz MAZURKIEWICZ
Lublin University of Technology
Nadbystrzycka 36, 20-618 Lublin, Poland
E-mail: D.Mazurkiewicz@pollub.pl

Article citation info:
MYSTKOWSKI A, KARBAY VK, MYSTKOWSKA J. A PLC based robust monitoring model for the labelling machine automation process. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 383–390.

Arkadiusz MYSTKOWSKI
Vedat Koray KARBAY
Joanna MYSTKOWSKA

A PLC BASED ROBUST MONITORING MODEL FOR THE LABELLING MACHINE AUTOMATION PROCESS

MODEL ODPORNY SYSTEMU MONITOROWANIA W AUTOMATYZACJI PROCESU ETYKIETOWANIA Z WYKORZYSTANIEM STEROWNIKÓW PLC

This paper presents a method for improving the labelling process and a robust monitoring model for the labelling machine with the purpose of reducing waste of labels and bottles.The proposed monitoring method is based on a combination of Matlab®-designed and hardware-in-the-loop (HIL) simulationas well as Arena Simulation. The method solves problems with the application of labels during the labelling stage and provides a robust monitoring algorithm that recognizes defective labels before they are stuck onto bottles.The Grafcet optimal algorithm for recognizing defective labels is executed. The Matlab®Stateflow model for monitoring and recognizing defective labels is applied. The proposed algorithms are complete, and optimized solutions are ready for implementation in the existing PLC supervisory control system. Based on HIL simulations, the proposed method ensures an increase of the total production quantity. Statistical data was collected directly from the field, classified using Statfit software, and used in Arena Simulation software to present the difference and benefits before and after using the PLC-based robust monitoring model for the labelling machine automation process.

Keywords*: label defects detection, robust monitoring Stateflow® model, labelling, PLC, HIL simulation.*

W pracy przedstawiono metodę poprawy procesu etykietowania oraz model odporny monitorowania uszkodzeń etykiet w celu zmniejszenia ilości odpadów etykiet i butelek. Opracowanie proponowanej metody monitorowania i wykrywania wad etykiet opiera się na wykorzystaniu kombinacji funkcji środowiska Matlab® oraz symulacji sprzętowej (ang. hardware-in-the-loop, HIL). Nowa metoda rozwiązuje problemy związane z wykrywaniem uszkodzeń przyklejanych etykiet do butelek w przemysłowej linii produkcyjnej oraz zawiera model odporny detekcji wad etykiet. Algorytm systemu monitorowania w procesie etykietowania został przedstawiony za pomocą sieci Grafcet, a następnie zrealizowany w środowisku Matlab Stateflow®. Proponowane algorytmy monitorowania/detekcji zostały zoptymalizowane pod kątem ich realizacji w istniejącym systemie sterowania opartym o programowalne sterowniki logiczne (ang. programmable logic controllers, PLCs). Przeprowadzone symulacje sprzętowe HIL pomyślnie weryfikują opracowane rozwiązania podnoszące efektywność produkcji. Zaproponowany odporny model detekcji uszkodzeń etykiet został zaimplementowany w układzie sterowania linii produkcyjnej i zweryfikowany eksperymentalnie. Zebrane dane statystyczne bezpośrednio z obiektu sterowania zostały opracowane w programie Statfit. Oprogramowanie Arena Simulation zostało wykorzystane do porównania wyników pracy linii produkcyjnej przed i po wprowadzeniu modelu wykrywania uszkodzeń etykiet.

Słowa kluczowe*: wykrywanie wad etykiety, model odporny monitorowania Stateflow®, etykietowanie, sterownik PLC, symulacja sprzętowa HIL.*

1. Introduction

Nowadays, many industrial applications require a low-cost solution for improving their productivity and the quantity of products. Monitoring systems are especially important at the production stage, where waste of material and products should be limited to a minimum. Monitoring systems based on programmable logic controllers (PLC), I/O-devices and distributed sensors/actuators are economic solutions [12]. These systems can also be coupled and cooperate with an existing network of digital control devices. Moreover, hardware-in-the-loop simulation using PLC-open functions, and, for example, Matlab® software models, can be used to design and test a suitable monitoring system in a cheap and fast way [10].

Many corporations have on-going research projects aimed towards reducing waste in manufacturing systems. In the literature, a number of papers are focused on using PLC technology in various industrial processes. Diagnostic techniques for PLC-controlled flexible manufacturing systems (FMS's) are proposed by Hu et al. (1999) [5]. Maria G. Ioannides (2004) describes the implementation of a monitoring and control system for an induction motor based on a programmable logic controller and provides the implementation of the hardware and software for speed control and protection with the results obtained from induction motor performance tests [6]. Georges, B. and Aubin, J. (2001) examined GE Syprotec Inc. of Canada's design of a PLC-based transformer monitoring and control system (TMCS). The TMCS aims to manage the operational flow and to enhance the performance of a transformer [2]. A. Ramirez-Serrano, S. C. Zhu, S. K. H. Chan, S. S. W. Chan, M. Ficocelli and B. Benhabib (2002) presented a new PC/PLC-based software/hardware architecture for the control of flexible manufacturing work cells [7]. Hairui, W. and Yong, Z. (2009) focus on a process control system for management of carbon dioxide content in food by using a distributed control system based on configuration software and PLC [4]. Theiss, S., et al. (2006) present an additional software entity ("monitoring agent") which provides process data acquisition and improves sampling resolution and flexibility of real-time PLC (programmable logic controller) devices [8]. Arrofiq, M. and Saad, N. (2007) designed and implemented a PLC-based fuzzy logic controller for induction motor speed control with a constant V/Hz ratio [1]. M.F. Zaeha, C. Poernbachera, J. Milberg (2005) discuss the difficulty of developing PLC software for modern machine tools due to their increasing complexity and functionality and present a model-based development and simulation-aided verification of control software [9].

The most modern and high-tech lubricants plant of the multinational oil, gas and energy company, British Petroleum, is located in Gemlik, Turkey. The plant produces motor oils and lubricants for the domestic and foreign market. The plant produced 70,000 tonnes of lubricants in 2011. The plant is mainly divided into three departments: logistics, production and filling. The logistics department is divided into two sections. The first one is responsible for delivering raw materials to the production facility. Section two receives produced lubricants from conveyors and organizes freight traffic within the plant complex. The production department receives raw materials and additives from both the logistics department and the dock via transfer pipes. Finally, the filling process is carried outin a separate building in the production complex, where aluminium cans and plastic bottles are filled with final products – the lubricants.

This paper focuses on a problem that occurred at the filling facility. Especially during the summer months, the glue of the labels melts, and the labelling machine cannot apply the labels properly. The labelling machine cannot detect defective labels, it can only detect them after sticking them onto a bottle via its sensors. However, the application of a defective label also causes the bottle to be disposed of, because the surface of the bottle becomes gluey and wet, which is inconvenient for sticking on another label.The workflow at the filling facility is presented in Fig. 1.

The work-flow begins with the loading of bottles onto the conveyor line by field workers, and bottles follow the path of conveyors presented in Fig. 1. Before the start of the work-flow, a field worker loads label rolls into the machine, and when a bottle reaches the labelling station, the machine sticks the label onto the bottle. After application, the machine's quality control system, which uses a programmable logic controller (PLC), performs quality control via its sensors and decides whether or not the applied label is compliant with quality standards. If it is, the bottle is delivered to the next station, the bottling station. If not, the machine disposes of the bottle, meaning that both the label and the bottle are lost. This complication was causing deviations from production plans and waste of materials used in production. Therefore, more bottles and labels had to be put into the system in order to obtain the desired total production quantity, since a lot of bottles, along with their labels, are disposed of because of the problem.

Finally, this paper presents a PLC-based robust monitoring model for the labelling machine, which reduces waste of labels and bottles and optimizesthe algorithm for recognizing defective labels.

2. Process

The detailed work-flow at the filling facility is presented in Fig. 2. In this figure, the PLC-based monitoring system is only introduced for the labelling machine. It is important to note that bottling, capping and packaging systems also have similar quality control systems, however they are not presented, since this paper is only focused on the problem occurring in the labelling sub-process.

A conveyor carries out the entire process at the filling facility, meaning that human intervention is only necessary for inspecting the process and loading supplementary production products (such as grease – labels, etc.). A brief description of the whole flow is presented below:

- after filling orders from the planning department are received, bottles are loaded into the conveyor system,
- at the labelling station, labels are stuck onto bottles. The labels show the features of the products. They consist of the corporate logo, product name and product-specific information. Rolls of labels are put into the machine by the responsible operator, and the machine automatically applies them. If the PLC detects a bottle with a defective label, the labelling machine disposes of both the label and the bottle. The bottle is also disposed of, be-

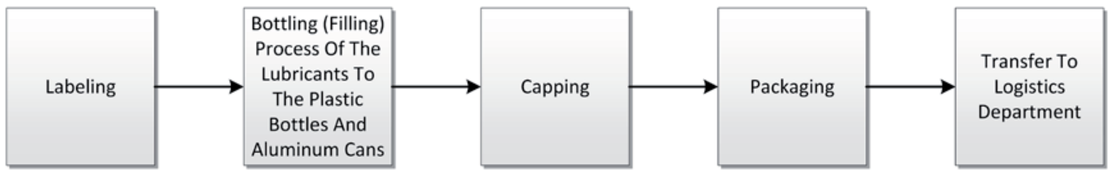


Fig. 1. Workflow of the filling process

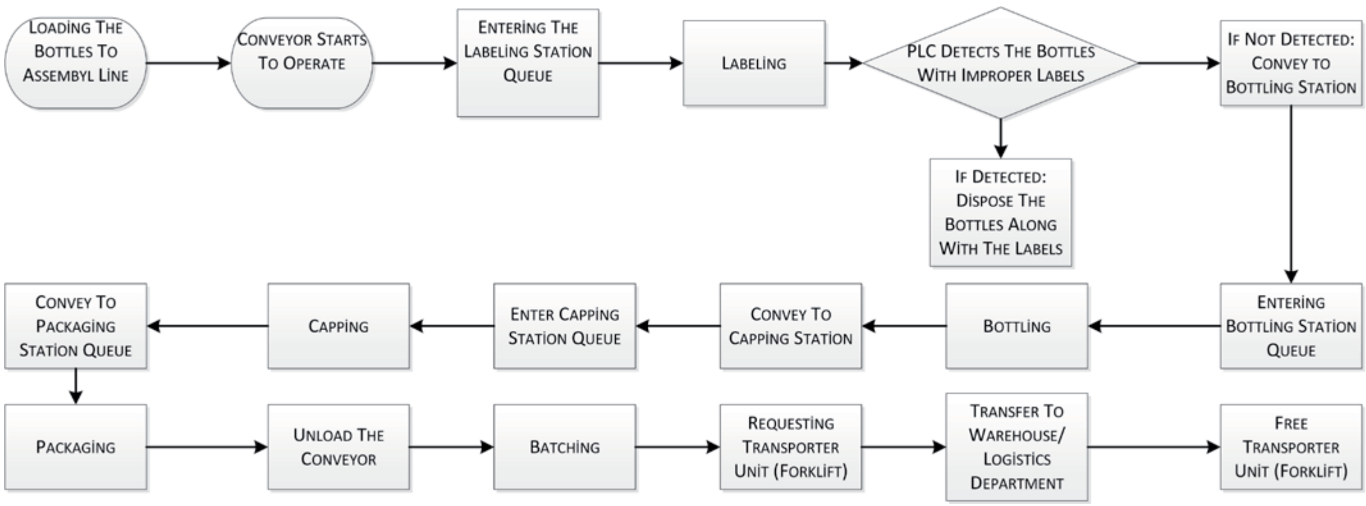


Fig. 2. Detailed process flow at filling facility

cause its surface also gets covered with glue, making repeated label application impossible,

- during the bottling stage, aluminium cans and/or plastic bottles are filled with lubricants, which are delivered from the production facility through tubes,
- during the capping stage, the machine seals bottles with the appropriate bottle caps,
- at the packing and batching station, after the completion of a group of products, (the number varies from product to product), the conveyor delivers products to the packaging machine, packing them in cardboard first, and then wrapping them with plastic material,
- during the last stage, transferring, the completed products are transported to the logistics department or warehouse to await freight transport to the final destination.

3. Proposed method for labelling process improvement

The aim of the proposed method consists of the following aspects:

- solving problems with label application at the labelling stage, which will reduce waste of labels and bottles, e.g. before the application of a previously recognized defective label, a bottle will be frozen to improve the quality of label sticking,
- providing a robust monitoring algorithm, which recognizes defective labels before they are stuck onto bottles. Thus, bottles will not be wasted. In this case, robust means that this monitoring system is accurate and is not sensitive to external disturbances of the process.

The method described above will increase the total economic production quantity (EPQ). Here, it must be emphasized that the algorithms within this method are complete and optimized solutions that are ready for implementation in the existing PLC supervisory control system. Moreover, these proposed solutions are cheaper than others, e.g. installation of a climate control system in the filling facility or of an automatic storage machine, which would be more costly. However, the technical specifications and parameter values of these solutions are not official, and they are not given in this paper.

3.1. Robust method for detecting defective labels

The method for detecting a defective label is comprised of three steps:

- first, the label's glue temperature is measured by a non-contact temperature sensor (e.g. the FLIR A310 – forward looking infrared – thermal imaging camera), if the temperature is higher than desired, the label is marked and information about the label's increased temperature is sent to the freezing system, after which the algorithm goes to the second step,
- during the second step, the label's glue density is measured by laser sensors, and if the glue density is below the lower limit, the label is removed, and information about the removed label is sent to the supervisory computer,
- finally, an image of the bottle with the applied label is generated using a CCD camera, and after that, a binary image containing a striped pattern of the label surface area is recognized. Some image recognition algorithms are used here, and if the label surface quality is lower than desired, the bottle is removed.

For ideal assessment of a label's glue, a liquid density sensor should be applied. However, this must be a non-contact sensor. Moreover, the liquid density is very sensitive to changes in temperature. Thus, the measurement of density and temperature should be closely coupled. All of the sensors (temperature and density) described above should operate online and in real-time without delay. They must provide stable and reliable measurement, even in a system with a high degree of agitation or disturbances.

Image processing of labels on bottles is the last and most complex step of label inspection. Here, the dynamic compensation algorithm is used to filter out shadows, tints and reflections on the background area of the label surface. The dynamic compensation algorithm is es-

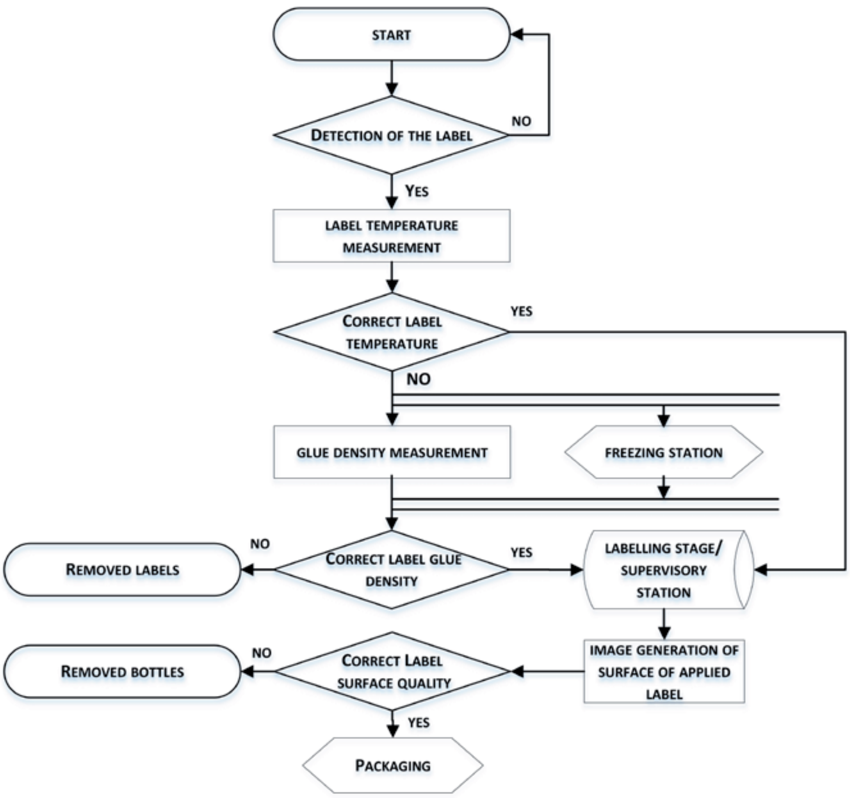


Fig. 3. Scheme of label defects detection

pecially useful in the case of highly reflective labels. Thus, the image processed by this algorithm contains less shadow area, making the image easier to analyse for defects. The Matlab® Image Processing toolbox can be used in order to design the dynamic compensation filter and obtain the image recognition algorithm [10]. The recognition algorithm is extended by a tool for setting vertical and horizontal inspection of the label surface. Here, the label quality is verified by detection of wrinkle defects on the leading edges of the label.

The Grafcet algorithm proposed for detection of label defects is presented in Fig. 3.

4. Stateflow monitoring model for the labelling process

The monitoring system for the labelling machine can be described by states and events for further recognition of defective labels. This system is known as the discrete event drive system (DED). The algorithm for classification of labels in the manufacturing process is provided in the Matlab® software by a tool for modelling and control of this system called the Stateflow® toolbox [10]. After that, the Simulink PLC Coder software is used to generate hardware-independent structured text from Stateflow® charts. As a result, it is possible to compile and deploy the proposed algorithm to programmable logic controller (PLC) and programmable automation controller (PAC) devices by using PLC-oriented tools [11]. Moreover, the Simulink PLC Coder provides optimizations that reduce the memory used by the PLC controller and increase the execution speed of the generated structured text. Finally, the Simulink PLC Coder can simulate the algorithm prior to structured text generation and package the results into a test harness that is generated by the algorithm code. The workflow of the Simulink PLC coder is presented in Fig. 4. The main goal is to propose an optimization algorithm for recognition of defective labels.

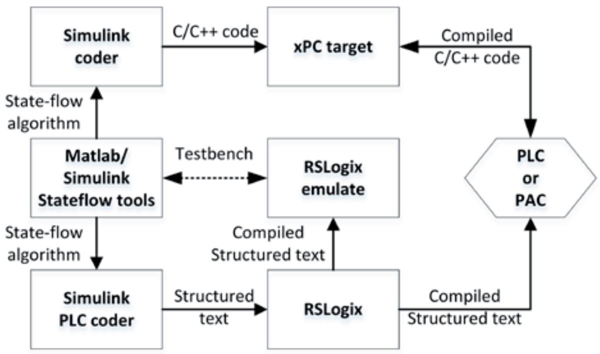


Fig. 4. Workflow of the model-based design [10]

The Simulink PLC coder (Fig. 4) generates structured text for the target PLC controller and provides verification support including test benches. The structured text of the control algorithm meets the standards of IEC 61131-3 e.g. for Rockwell Automation [11]. The Structured Text is generated in PLCopen XML, which is supported by widely used integrated development environments (IDEs) [10]. The Simulink PLC Coder also generates test benches in order to verify the structured text using PLC and PAC IDEs and simulation tools [10]. IDE support includes 3S-Smart Software Solutions CoDeSys, Rockwell Automation® RSlogix™ 5000, Siemens® SIMATIC® STEP® 7, Omron Sysmac Studio, and PLCopen XML [11].

4.1. The Stateflow label defects detection model

Stateflow® is a compressive tool for representation of event-driven (reactive) systems and includes designing and modelling of digital control algorithms [10]. In anevent-driven system, finite states are used to represent machine operations. The machine performing these operations is called a finite state machine (FSM). In a FSM, the relationships between inputs, outputs, and states are represented by truth tables, and the behaviour of an FSM is described by the conditions of transitions between states [3].

Label defects detection and the decision-making algorithm are presented by agraphical chart which represents the FSM. States (e.g. *labelling, freezing, packaging, surface_quality, temperature_sensor,* and *density_sensor*) and transitions (e.g. *label, temp, density*, and *surface*) form the blocks of the algorithm in the FSM (see Fig. 5).

Label validation according to glue temperature and density is implemented in the event algorithm by the Simulink models. The *freezing* station is designed by using Matlab functions. The history junction (H) records the activity of substates (e.g. total amount of removed bottles and labels) within the main finite states. Label surface validation is realized by using a complex image processing function. All transitions are conditioned by internal Boolean conditions according to the limits of main process parameters (e.g. [*temp<temp_limit*]). The solution of this algorithm concerns switching between states in the desired order and according to defined logical conditions. The process's modes of operation are modelled as states and represent the logic for switching between modes using transitions and junctions. The algorithm starts from the default state (*label detection*), then *temperature_sensor* state is activated. After that, if the glue temperatureand density are below the limit, the *labelling* state is activated, and if not, the *parallel_processes*state is activated. This state provides two states: *freezing* and *density_sensor* in a parallel configuration. Thus, the *labelling* state is activated again if the glue parameters meet the logical conditions for desired temperature and density. All defective labels are removed by activating the *removed_labels* state. During the

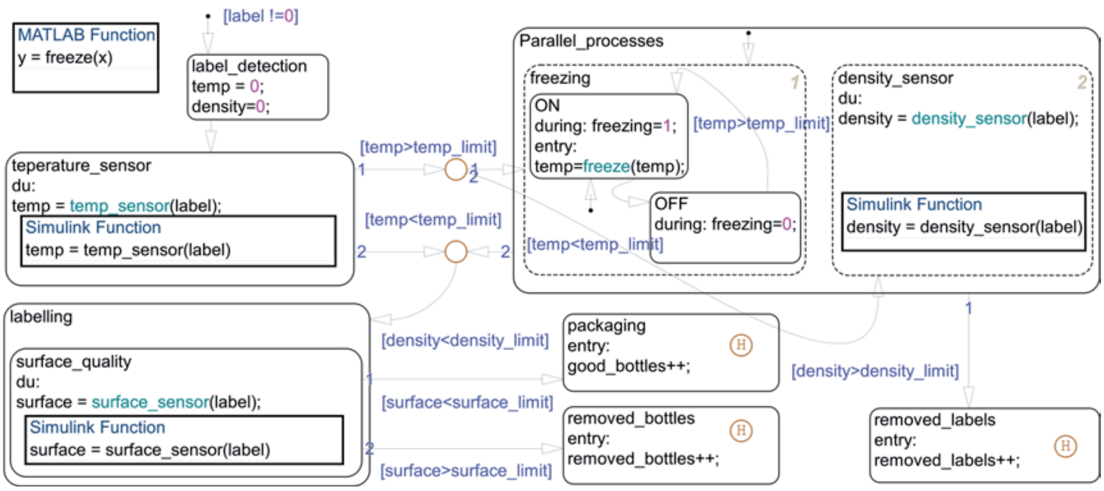


Fig. 5. Chart of the process algorithm

labelling state, label surface quality is validated, and the*surface_quality* state is activated. The outputs from the *surface_quality*state provide and activate the *packaging* or *removed_bottles* state depending on defined logical conditions.

5. Testing of the algorithm using the Simulink C Coder and PLC

In order to implement the Stateflow® algorithm in PLC or PAC memory, the flow graph is transformed to C/C++ code. The Embedded Coder™ software enables generation of real-time code directly from the flow graph. The C source code generated for the label defects detection model is needed to perform a simulation and tests. The generated code is optimized before the testing process. For example, the parallel states of freezing and density validation, with conditional transitions, are presented within the following C code:

```
/* Outputs for Function Call SubSystem: '<S1>/Parallel_processes.density_sensor.density_sensor' */
  /* Gain: '<S2>/Kd' incorporates:
   *  Sum: '<S5>/Diff'
   *  UnitDelay: '<S5>/UD'
   */
  /* During 'density_sensor': '<S1>:26' */
  /* Simulink Function 'density_sensor': '<S1>:29' */
  rtb_Kd_e = (rtb_randomlabels -
          label_defects_detection2_DWork.UD_DSTATE_d) *
    label_defects_detection2_P.Kd_Gain;

  /* Update for UnitDelay: '<S5>/UD' */
  label_defects_detection2_DWork.UD_DSTATE_d = rtb_randomlabels;
  label_defects_detection2_B.density = rtb_randomlabels + rtb_Kd_e;

  /* End of Outputs for SubSystem: '<S1>/Parallel_processes.density_sensor.density_sensor' */
}
```

The main advantage of automatic code generation is the elimination of errors that may accrue during manual writing.

The last stage of code validation is to test the fully implemented code in the PLC hardware and, after that, to compare the results with the original simulation results. Hardware-in-the-loop (HIL) simulation makes it possible to run simulations of all stages of process automation control in real-time using the original C code and a real PLC controller. For this purpose, real-time communication between MATLAB/Simulink and the industrial controller (PLC) is established. Then, input/output signals are scaled due to analog-digital/digital-analog conversion(A/D, D/A). The entire system of the HIL test is presented in Fig. 6.

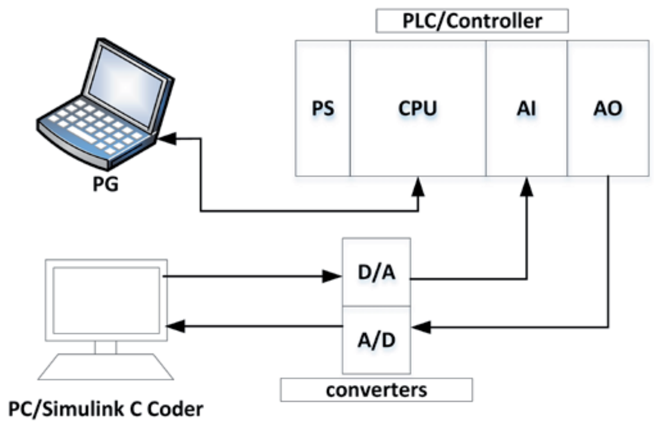


Fig. 6. Connection between the PLC and PC/Simulink C coder during HIL simulation

The experimental setup consists of: a PLC with PS, a CPU, and analog input/output modules (AI and AO), A/D and D/A converters, a PC with Matlab, and a PG station for programming the PLC. The results of the HIL tests will be used to validate the simulation model in another work.

6. Forecasting the benefits of the proposed model by using ARENA Simulation Software

By presenting a method for improvement of the labelling process and a robust monitoring model for the labelling machine, we aim to optimize three different outputs. First, the number of defective labels and bottles will be reduced. Second, costs resulting from the disposal of defective labels and bottles will be reduced. Third, the total production time will be optimized. These three outputs will help to assess the benefits of the Simulink/Stateflow model introduced in section 4. The statistical data of the labelling/bottling/capping and packaging processes before the application of the robust monitoring model are presented in this section.

6.1. Data Collection, Assumptions and Data Analysis

A detailed analysis was carried with the primarygoal of characterizing the performance of the system in different scenarios at the plant. To create a model for the simulation software, a chronometer was used to measure the time performance of the machines. Each step was observed 30 times in order to achieve reliable and descriptive statistical results. Statfit software was used to determine "the goodness of fit". Detailed information regarding machine operation times before implementation of the Simulink/Stateflow robust monitoring model is provided in Table 1.

A number of assumptions were made. First of all, production starts at 8 AM. Secondly, the times of transportation by forklift from the filling department to the logistics department were observed. The results of these observations were between 4.46 and 5.05 minutes, therefore, the transport time was assumed to be 5 minutes. Thirdly, the system doesn't stop until it reaches the desired production quantity, meaning that breakdowns, breaks, or any other disturbances are ignored and not included in the simulation model. Fourthly, detection of defects with the help of the PLC system is only defined for the labelling machine. The defectiveness rate on other machines is ignored since the rate is lower than 3%. Fifthly, the simulation is used for 2 different types of products (which have different batch sizes), and each of these simulations has an ending condition, which is the generation of, respectively, 500 (Product A) and 1000 (Product B) completed final products. Finally, the production (model) stops when the desired production quantity has been achieved, which means thatthe model doesn't requirewarm-up time.

In such a state-of-art lubes plant, production flow is carried out via tubes and conveyors. The personnel mainly inspect the flow of the system. Therefore, the following data are considered to be deterministic. Firstly, after observation of the conveyor's speed, it was noted that a bottle was loaded onto the conveyor system of the filling facility every 1.5 minutes. Secondly, the first in first out rule, FIFO (first input; first output), is applied for all queues on the conveyor line. Thirdly, the cost for disposing of a bottle with a label stuck onto it is 1.1€. Machine operation times were considered to be stochastic after chronometric observations were made.

Table 1. Descriptive statistics of the operation times of machines before the implementation of the robust monitoring model

	Sample Size	Range	Mean	Variance	Excess Kurtosis	Std. Deviation	Coef. Variation	Skewness
Labeling	30	19	69.96	35.96	−11.784	59.971	0.08612	0.1291
Bottling	30	4	30.3	1.61	−12.124	12.689	0.04188	−0.08615
Capping	30	21	51.1	46.82	−12.221	68.428	0.13391	0.05072
Packaging	30	3	61.83	0.73	−0.38431	0.85959	0.0139	0.95634

Table 2. Detailed results of 10 replications after the implementation of the robust monitoring model

		Number Of Replication									
Unit	**Description**	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
	Simulation Outputs Of Product A Under 30% Defection Rate										
Number	Number Of Defected Labels	228	197	236	237	214	184	223	244	221	217
Euros	Cost Of Disposed Labels and Bottles	125.95	108.9	130.35	130.9	118.25	101.75	123.2	134.75	122.1	119.9
Minutes	Total Production Time	558.441736	542.957	558.7365	551.0397	539.9819	518.728	571.0567	566.7704	552.2564	558.9632
	Simulation Outputs Of Product A Under 25% Defection Rate										
Number	Number Of Defected Labels	162	149	155	166	171	147	177	147	144	206
Euros	Cost Of Disposed Labels and Bottles	89.65	82.5	85.8	91.85	94.6	81.4	97.9	81.4	79.75	113.85
Minutes	Total Production Time	500.288966	503.198	497.2145	503.6616	512.5845	491.7889	531.3535	498.6743	489.7158	551.6037
	Simulation Outputs Of Product A Under 2% Defection Rate										
Number	Number Of Defected Labels	16	9	9	7	12	8	14	9	15	14
Euros	Cost Of Disposed Labels and Bottles	9.35	5.5	5.5	4.4	7.15	4.95	8.25	5.5	8.8	8.25
Minutes	Total Production Time	398.080675	394.413	395.2622	392.726	395.2118	394.7517	398.2378	395.2784	398.2744	396.0148
	Simulation Outputs Of Product B Under 30% Defection Rate										
Number	Number Of Defected Labels	424	403	435	468	418	393	443	449	426	422
Euros	Cost Of Disposed Labels and Bottles	233.75	222.2	239.8	257.95	230.45	216.7	244.2	247.5	234.85	232.65
Minutes	Total Production Time	1093.880785	1063.92	1103.216	1104.154	1074.024	1047.443	1108.876	1108.52	1089.014	1076.287
	Simulation Outputs Of Product B Under 25% Defection Rate										
Number	Number Of Defected Labels	346	328	304	323	325	311	336	295	296	362
Euros	Cost Of Disposed Labels and Bottles	190.85	180.95	167.75	178.2	179.3	171.6	185.35	162.8	163.35	199.65
Minutes	Total Production Time	1007.281788	995.135	990.6528	996.2137	1012.036	989.4882	1022.921	985.4722	981.9551	1055.244
	Simulation Outputs Of Product B Under 2% Defection Rate										
Number	Number Of Defected Labels	27	15	22	20	21	16	22	19	24	24
Euros	Cost Of Disposed Labels and Bottles	15.4	8.8	12.65	11.55	12.1	9.35	12.65	11	13.75	13.75
Minutes	Total Production Time	784.022205	775.099	778.2257	774.8996	778.6259	775.2485	781.2348	776.8666	782.7499	779.8694

Table 3. The average results of 10 replications after the implementation of the robust monitoring model

Product A			
Defectiveness Rate	Average Total Product Time	Average Total Number of Defective Labels	Average Cost Of
Defective Labels			
30% Defection	550.51	220.10	121.61
25% Defection	507.53	162.40	89.87
2% Defection	395.38	11.30	6.76

Product B			
Defectiveness Rate	Average Total Product Time	Average Total Number of Defective Labels	Average Cost Of
Defective Labels			
30% Defection	1086.64	428.10	236.01
25% Defection	1004.20	322.60	177.98
2% Defection	778.37	21.00	12.10

7. Experimental set-up

The Simulink/Stateflow simulation model was implemented and built using ARENA Simulation Student Edition Software Version 14.00.00000 from Rockwell Automation with the help of the conceptual model described in Section 2 . This section explains the outputs of the simulation model outputs and contains a discussion. Two different types of products were chosen from among the product assortment of the company. However it is important to note that all names have been manipulated for proprietary reasons. These two products will be called Product A and Product B. The features of these products and experimental results are given in this section.

7.1. Experimental results

Over 10 replications, the Simulink/Stateflow simulation model (robust monitoring model) observed the results of three outputs:

- average total cost of disposed products (Euros),
- average total production time for stated production quantity (minutes),
- total number of defective labels (number).

The simulation model's reaction toward these three outcomes was observed under three different defectiveness rates for labels: 30%, 25% and 2%. The defectiveness rate for labels causes instability in production schedules, therefore the simulation model experimented with these three values. The defectiveness rate for labels is known to be between 25% and 30%, therefore, these two values wereused. 2%, on the other hand, waschosen to see how much the plant can benefit on the outputs, if the problem were to be fixed. A 2% problem rate (possible risk-defect-danger) is used at the plant for operations which are considered to be almost risk-free. The experimental results of the system with the implemented simulation model are collected in table 2, and average results are presented in table 3.

Based on the results in Table 2, the robust monitoring model works properly and detects defective labels for three different defectiveness rates. The greatest accuracy is noted for the label defectiveness rate of 30%. However, the number of defective labels is different in each simulation performed for Product A and Product B. Each replication of the simulation gives a similar number of defective labels. The costs of disposed labels and bottles increase due to the amount of defective labels.

Finally, by using the robust monitoring model created using ARENA software, we can observe the difference before and after implementation of the proposed Simulink/Stateflow model in the production process. To describe the situation in detail, between 25÷30% of labels and bottles were being disposed of before, and the proposed simulation model shows that our improvement fixes the problem, and now the disposal rate is reduced to 2%. Thus, it forecasts the benefit of our model by simulating the process.

8. Summary

In order to survive in today's highly competitive and global market, all businesses, regardless of their size and scale, must have brisk and responsive problem-solving mechanisms, otherwise even small complications may result in catastrophic outcomes for companies. This paper's field of observation, British Petroleum's lubricant production plant, is a good example of how constant observation of subprocesses, measurement of both qualitative and quantitative results, and their benchmarking with the expected outcomes help companies to construct successful problem-solving frameworks.

Although the literature is full of research on manufacturing improvements, the use of programmable logic controllers in combination with data provided by chemical analysis is considerably rare. With the help of the proposed method for improvement of the labelling process, which recognizes defective labels before they are stuck onto bottles, the number of defective materials was reduced, costs resulting from the disposal were reduced, and the total production time was optimized. Furthermore, the solution proposed in this paper is much cheaper compared to alternative options of solving the problem, which include obtaining an automated storage and retrieval machine with a climate controlling feature.

It is important to note that the subject of this paper is based on a real-life scenario, however some data have been manipulated due to company anonymity. But still, the framework and the outcomes of this study can be helpful for future research. The model proposed in this paper provides three important outcomes:

a) total production time was optimized: because of the proposal of a quicker and more reliable monitoring system, the total production cycle time will be faster compared to the old one,
b) reduction of the number of disposed products: deviations from production plans will be considerably reduced due to the elimination of the source of instability,
c) reduction of costs: because of the improvement, costs resulting from the problem are noticeably reduced.

Acknowledgment:
The work has been supported with Statutory Work of Department of Automatic Control and Robotics, Faculty of Mechanical Engineering, Bialystok University of Technology, No. S/WM/1/2012.

References

1. Arrofiq M, Saad N. PLC-based fuzzy logic controller for induction-motor drive with constant V/Hz ratio. Intelligent and Advanced Systems, ICIAS International Conference, 25-28 Nov. 2007, 93-98.
2. Georges B, Aubin J. Application of PLC for on-line monitoring of power transformers. Power Engineering Society Winter Meeting, IEEE, 2001, 2, 483-486.
3. Harel D. Statecharts. A visual formalism for complex systems. Science of Computer Programming,1987,8, 231-274.
4. Hairui W, Yong Z. Configuration software and PLC based process control system of carbon dioxide for food. Computers and Applied Chemistry, 2009, 26(2), 175-178.
5. Hu W et al. Two diagnostic models for PLC controlled flexible manufacturing systems. International Journal of Machine Tools & Manufacture,1999, 39, 1979-1991.
6. Ioannides M. Design and Implementation of PLC-Based Monitoring Control System for Induction Motor. IEEE transactions on energy conversion, 2004, 19(3), 128-135.
7. Ramirez-Serrano A, Zhu SC, et al. A hybrid PC/PLC architecture for manufacturing-system control-theory and implementation. Journal of Intelligent Manufacturing, 2002, 13(4), 261-281.
8. Theiss S, Naake J, Dibowski H, Kabitzsch K. PLC-integrated Process Monitoring and Prediction of the Resulting Real-Time Load. Industrial Informatics, IEEE International Conference,16-18 Aug. 2006, 880-885.
9. Zaeha MF, Poernbachera C, Milbergb J. A Model-Based Method to Develop PLC Software for Machine Tools. CIRP Annals - Manufacturing Technology, 2005, 54(1),371-374.
10. www.mathworks.com.
11. www: plcdev.com, codesys.com, automation.siemens.com, rockwellautomation.com.
12. IEC 61131-3 PLC programming language standards.

Arkadiusz MYSTKOWSKI
Department of Automatic Control and Robotics
Bialystok University of Technology, Wiejska 45C, 15-351 Bialystok, Poland
E-mail: a.mystkowski@pb.edu.pl

Vedat Koray KARBAY
Department of Industrial Engineering
Beykent University, Sisli-Ayazaga Campus, 34396, Istanbul, Turkey
E-mail: koraykarbay@gmail.com

Joanna MYSTKOWSKA
Department of Materialsand Biomedical Engineering
Bialystok University of Technology, Wiejska 45C, 15-351 Bialystok, Poland
E-mail: j.mystkowska@pb.edu.pl

Article citation info:
KATUNIN A, PRZYSTAŁKA P. Detection and localization of delaminations in composite beams using fractional B-spline wavelets with optimized parameters. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 391–399.

Andrzej KATUNIN
Piotr PRZYSTAŁKA

DETECTION AND LOCALIZATION OF DELAMINATIONS IN COMPOSITE BEAMS USING FRACTIONAL B-SPLINE WAVELETS WITH OPTIMIZED PARAMETERS

DETEKCJA I LOKALIZACJA ROZWARSTWIEŃ W KOMPOZYTOWYCH BELKACH Z WYKORZYSTANIEM B-SPLAJNOWYCH FALEK UŁAMKOWYCH Z OPTYMALIZOWANYMI PARAMETRAMI*

In this paper the method of detection and localization of delaminations in composite layered beams using an algorithm based on the wavelet transform of bending modal shapes of beams was presented. For the basis functions the fractional B-spline wavelets with single- and multi-objective optimization of the values of their parameters were selected. The analysis was carried out basing on the results of numerical simulations. Several cases of the delaminations occurrence with respect to their location on the thickness of the beams, different sizes and geometrical features, were analyzed. Results of the conducted analyzes show the high effectiveness of a method in the task of detection of delaminations and a possibility of its application in industrial conditions.

Keywords: *detection and localization of delaminations, layered composites, fractional B-spline wavelets, optimization.*

W pracy przedstawiono metodę detekcji i lokalizacji rozwarstwień w kompozytowych belkach warstwowych z wykorzystaniem algorytmu opartego na transformacji falkowej giętnych postaci własnych drgań belek. Jako funkcje bazowe zastosowano ułamkowe falki B-splajnowe z jedno- i wielokryterialną optymalizacją wartości ich parametrów. Analizę przeprowadzono na wynikach obliczeń numerycznych. Przeanalizowano przypadki występowania rozwarstwień w różnej lokalizacji na grubości płyt oraz przypadki z rozwarstwieniami o różnych rozmiarach i postaciach geometrycznych. Wyniki przeprowadzonych analiz wykazały wysoką skuteczność metody w wykrywaniu rozwarstwień i możliwość jej zastosowania w warunkach przemysłowych.

Słowa kluczowe: *detekcja i lokalizacja rozwarstwień, kompozyty warstwowe, ułamkowe falki B-splajnowe, optymalizacja.*

1. Introduction

Application of layered composites as a constructional material in the elements of engineering constructions became currently common and popular thanks to a possibility of significant reduction of the mass of elements with simultaneous retaining of strength properties, great elasticity in design of such constructions, the possibility of integration of control elements and actuators into the structure, etc. For this reason there occurred a necessity of development of diagnostic methods dedicated for these materials. Some of the methods and techniques of structural diagnostics dedicated for the homogeneous materials also found an application in the case of composite materials, however for the composites diagnosis these methods need to be developed considering the possibility of diagnosing of new types of damages, which have not occurred in the homogeneous materials, e.g. delaminations, cracks of armed fibers, interphase decohesion, etc. Furthermore, these methods should be non-invasive, they should be highly sensitive to the damages, resistant to the external influences, low-cost of carrying out the research and easiness of their application in the industrial conditions.

Among the methods applied in the structural diagnostics of polymeric composites the next groups should be mentioned: interferometric, radiologic, thermographic methods and others, which require however the usage of advanced measurement devices for carrying out the research, which often limits their applicability to the laboratory conditions. Ones of the dynamically developed methods are the methods based on the analysis of modal shapes of vibration or deflection profiles with using of advanced signal processing techniques. One of such techniques, actively developed in the problems of structural diagnostics, is the wavelet transform. This technique gains a great popularity thanks to the high sensitivity to abrupt changes in the signal, and furthermore the possibility of selection of basic and scaling functions depending on the type of detected changes in the analyzed signal.

The methods of cracks detection in beams using the wavelet transform have been analyzed by many researchers. An attention should be paid to the studies [3, 14], which describe an algorithm of cracks detection in beams based on modal shapes with use of continuous wavelet transform (CWT) and symlets applied as the basis functions. The authors of [24] also used CWT and symlets of order 4 for detection of cracks in simply-supported beams. In the case of application of the wavelet transform in the problems of structural diagnostics the crucial importance has the selection of wavelet, which is used for the transform. The authors of [19] presented the problem of an identification of cracks using CWT and carried out a comparative analysis of wavelets, which shows that the most efficient wavelets in the problems of structural diagnostics are the Gabor wavelets. Another approach was presented in [16, 20], where the wavelet analysis was carried out basing on the static deflection profiles of beams. An alternative method developed by the first author of this paper [8-10] is the method based on the discrete wavelet transform (DWT), which allows to signifi-

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

cantly reduce the computation time and is characterized by the highest accuracy with respect to other types of wavelet transforms, which was conducted during the comparative study presented in [12]. The application of DWT imposes several limitations on wavelets, which could be applied during the analysis. These wavelets should consist a compact support and fulfill the conditions of orthogonality, which eliminates a possibility of application of Gabor wavelets described as the most effective in the considered problems [19]. However, an application of B-spline wavelets gives a possibility of obtaining of good results as well due to their convergence to the Gabor wavelets, especially in the case of higher order B-spline wavelets [22].

Detection and localization of delaminations in composite beams constitute the more difficult problem in comparison with cracks detection and localization. It is caused by the occurrence of non-monotonicities only on the boundaries of delaminations, which makes difficult the process of identification of these damages. Only the few studies could be found in the literature, which describe research concerned with detection and localization of delaminations in composite structures. In [18] the authors described a method of delaminations localization basing on analysis of vibration signals. The authors of a study [7] presented results of the application of the Haar wavelet for detection of delaminations in composite beams. However, the vibration-based methods are characterized by poor detectability of such damages, which confirmed the authors of [25].

Another important aspect, on which the attention should be paid, is that the diagnostic methods based on the wavelet transform require an appropriate selection of the parameters of applied wavelets. Several interesting studies, which describe this problem should be highlighted. The authors of [4] proposed a method of multi-objective optimization with application of the evolutionary algorithm for searching the parameters of wavelets applied in the analysis of ECG signals in order to increase an accuracy of the signal description. Rafiee et al. [17] show the method of application of artificial neural networks and genetic algorithms for the selection of the best wavelet functions in order to improve the results of diagnostics of the gearbox faults. This problem also was a topic of interesting studies in the area of selection of wavelets parameters applied for filtering of signals [6] and image processing [5]. From the other hand, the lack of research studies in the area of application of optimization methods for selection of parameters of fractional B-spline wavelets applied in the problem of detection and localization of delaminations in composite beams is noticeable.

In the presented paper the authors tried to solve a problem concerned with detectability of delamination due to the application of an algorithm based on DWT and the application of fractional B-spline wavelets developed by the authors of [23]. Previous results presented in [11] demonstrate the increase of the effectiveness of the method during detection of damages using fractional B-spline wavelets with respect to the B-spline wavelets of an integer order. For increasing the effectiveness of delaminations detection the algorithm was extended by adding the single- and multi-objective optimization of the parameters of fractional B-spline wavelets, which allows for automatic selection of an optimal wavelet for the considered case of a damage. Four groups of cases were analyzed in this study, which differ with respect to their shapes and location on the thickness of a beam. In particular, the influence of the shape and size of delaminations on their detectability and the variability of parameters of optimally selected wavelet was analyzed. The results of presented studies pointed out to the fact that the proposed method allows for detection and precise localization of delamination areas in composite beams.

2. Description of the method

In this section the mathematical fundamentals of the B-spline wavelets of fractional order as well as the method of detection and localization of delaminations with their application and a method of selection of parameters of the B-spline wavelets of fractional order with use of single- and multi-objective optimization will be presented.

2.1. B-spline scaling and wavelet functions of fractional order

The fundamental approach of DWT is the application of the multiresolution analysis proposed by Mallat in [15], where the B-spline scaling functions $\beta(x)$ constitute the space of square-integrable functions $L^2(\mathbf{R})$ and create the sequence of functional spaces in the form:

$$\{0\} \subset \ldots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset L^2(\mathbf{R}). \tag{1}$$

The general form of B-spline scaling function of fractional order could be presented by the following equation [1]:

$$\beta_\tau^\alpha(x) = \sum_{k=0}^{\infty} (-1)^k \left|\begin{matrix}\alpha+1\\ k-\tau\end{matrix}\right| \rho_\tau^\alpha(x-k), \tag{2}$$

where $\alpha \in \mathbf{R}$ is an order of a scaling function, $\tau \in \mathbf{R}$ is a shift parameter and ρ_τ^α is a function in the form:

$$\rho_\tau^\alpha(x) = -\frac{\cos\pi\tau}{2\Gamma(\alpha+1)\sin(\pi\alpha/2)}|x|^\alpha - \frac{\sin\pi\tau}{2\Gamma(\alpha+1)\cos(\pi\alpha/2)}|x|^\alpha \operatorname{sgn}(x) \tag{3}$$

where $\Gamma(\alpha+1)$ is the Euler's gamma function, which allows for fractional factorization. It should be noticed that when $\alpha \in \mathbf{Z}$ and $\tau = (\alpha+1)/2$ from the equations (2) and (3) the classic B-spline scaling functions (i.e. with an integer order) could be obtained.

Basing on the Mallat's algorithm the scaling function (2) fulfills the two-scale relation:

$$\beta_\tau^\alpha(x) = 2^{-\alpha} \sum_{k\in\mathbf{Z}} \left|\begin{matrix}\alpha+1\\ k-\tau\end{matrix}\right| \beta_\tau^\alpha(2x-k). \tag{4}$$

The general form of the B-spline wavelets of fractional order was defined by Unser and Blu in [23]:

$$\psi_\tau^\alpha\left(\frac{x}{2}\right) = \sum_{k\in\mathbf{Z}} \frac{(-1)^k}{2^\alpha} \sum_{l\in\mathbf{Z}} \binom{\alpha+1}{l} \beta_0^{2\alpha+1}(l+k-1)\beta_\tau^\alpha(x-k). \tag{5}$$

B-spline scaling and wavelet functions of fractional order fulfill the most of properties of their analogs of an integer order. The only property, which is not fulfilled for the purposes of DWT is the existence of the non-compact support of these scaling and wavelet functions for $\alpha \notin \mathbf{Z}$. However, the algorithm of fractional discrete wavelet transform (FrDWT) proposed in [23] is based on the Fourier series and allows to avoid this problem.

2.2. Method of detection and localization of delaminations

As it is known, the process of wavelet transform could be presented as filtering with use of a set of high-pass and low-pass filters. The presented method is based on single-level decomposition (filtering) of the signals of displacements of modal shapes of beams with use of a set of filters with fractional order. Considering the two-scale relation of the B-spline scaling function of fractional order (4) the high-pass filter (scaling filter) could be defined following its pulse response in the form [1]:

$$H_\tau^\alpha\left(e^{j\omega}\right)=2^{-\alpha}\left(1+e^{j\omega}\right)^{\frac{1}{2}(\alpha+1)-\tau}\left(1+e^{-j\omega}\right)^{\frac{1}{2}(\alpha+1)+\tau}, \tag{6}$$

the low-pass filter (wavelet filter) could be presented in the similar manner as:

$$G_\tau^\alpha\left(e^{j\omega}\right)=-e^{-j\omega}H_\tau^\alpha\left(-e^{-j\omega}\right)A^\alpha\left(-e^{j\omega}\right), \tag{7}$$

where:

$$A^\alpha\left(e^{j\omega}\right)=\sum_{k\in Z}\beta_\tau^{2\alpha+1}(k)e^{-j\omega k}. \tag{8}$$

The proposed algorithm is based on single-level decomposition of displacements signal of modal shapes s_n, by using (6) and (7), and then the downsampling procedure. During these operations the sets of approximation coefficients a_n and details coefficients d_n could be obtained, the length of realization of both sets is reduced twice with respect to s_n due to the downsampling operation. The information about eventual damages (non-monotonicities of a signal) is stored in the set of details coefficients. In the case of occurrence of the damages the coefficients in the location of damages gain much greater or much lower values with respect to other coefficients in the given set.

During the previous analyzes [8–10, 12] it was noted that the values of coefficients d_n are in the great dependence to the displacements values in modal shapes, i.e. in the case, when the damage is located in the node of a given modal shape its detection becomes impossible. In order to avoid of such a situation the multiple modal shapes should be considered during the analysis. Additionally, in the case when $\alpha < 2$ the details coefficients consisted of a trend, with is connected with insufficient filtering of the modal shape from a signal. In order to obtain correct detection and localization of delaminations it is necessary to apply an approximation of each of the obtained sets of details coefficients for the considered modal shapes. For increasing the detectability of the damages it is suitable to add up the absolute values of details coefficients, which eliminates the differences in signs and expose the extrema of these coefficients in the locations of damages. The scheme of the method was presented in Fig. 1.

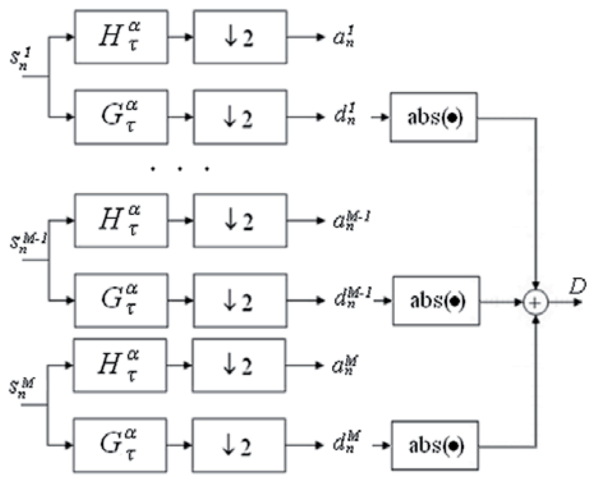


Fig.1. Scheme of the method of delaminations detection and localization

2.3. Selection of optimal parameters of wavelets

Results obtained in [13] show that the detection and localization of delaminations boundaries could be detected basing on non-monotonicities of deflection functions. As it was mentioned before, the symptoms of such a damage are the extrema of details coefficients, which occurred on the boundaries of the damaged areas. An accuracy of the proposed method of detection and localization of delaminations is strongly dependent on the appropriate selection of the B-spline scaling (2) and wavelet (5) functions of fractional order. Due to this reason it is necessary to determine optimal values of parameters of these functions. The optimization problem in this case could be reduced to the searching problem of such values of parameters α, τ, for which the multi-objective fitness function **F** in the general case reaches the minimum value. Assuming that the criteria of the fitness function are not in conflict each other the optimization problem could be formulated as follows:

$$\min_{\alpha,\tau}\mathbf{F}(\alpha,\tau)=\left[f_1(\alpha,\tau)\quad f_2(\alpha,\tau)\right]^T, \tag{9}$$

where $\alpha, \tau \in$ R, $0<\alpha<\alpha_c$, $1<\tau<\tau_c$, α_c and τ_c are the upper boundaries of the parameters. It should be noticed that the lower boundary value of a parameter α should be greater than -0.5 [23]. However, as the introductory studies [11] show the range of $-0.5<\alpha\le 0$ is not suitable for the investigated problem.

Considering the above-presented assumptions, the first optimization criterion could be defined as a proportional value to the inverse of sum of maximal values of the details coefficients:

$$f_1=\left[1+\sum_{k=1}^{M}\sum_{i=1}^{2H}\max\left(D_i^k\right)\right]^{-1}, \tag{10}$$

where H is the number of delaminations, M is the number of considered modal shapes, the details coefficients in a set $D_1^k=D^k$ and $D_i^k=D_{i-1}^k\setminus\hat{d}_{i-1}^k, i\neq 1$ for $\hat{d}_{i-1}^k$ is a point in the set of details values D_{i-1}^k, for which the maximal value of the details coefficient was obtained. The second criterion could be defined basing on a measure defined for the rest of values of details coefficients in a set D^k. Such a criterion could be presented as follows:

$$f_2=\sum_{k=1}^{M}\sum_{i=1}^{\left|\hat{D}^k\right|}d_i^k, \tag{11}$$

where the following relation occurs: $\hat{D}^k=D^k\setminus\hat{d}_1^k\setminus\hat{d}_2^k\cdots\setminus\hat{d}_{2H}^k$.

In the general case the problems of multi-objective optimization do not reach a single global solution. Therefore, it is justified to consider a set of solutions, which fulfill the boundary conditions and optimization criteria. In the following study the problem of multi-objective optimization was solved by searching an optimum of the fitness function in the Pareto sense. The solution is Pareto-optimal, if there is no another solution, which could improve at least one of the criteria and simultaneously will not make worst the other criteria. Such a solution is considered as non-dominated one.

Often in the practical applications the problem of multi-objective optimization is reduced to the one-dimensional problem. In this study such a cases was also considered. For this purpose the scalar fitness function in the form of metacriterion created basing on weighted product of the elementary criteria:

$$U = \left[1 + \sum_{k=1}^{M} \sum_{i=1}^{2H} \max\left(D_i^k\right)\right]^{-w_1} \cdot \left(\sum_{k=1}^{M} \sum_{i=1}^{\left|\hat{D}^k\right|} d_i^k\right)^{w_2}, \tag{12}$$

where $w_{\{1,2\}} \in \langle 0,1 \rangle$ denote the weights of the particular criteria, which values could be defined basing on experts' knowledge or by application of systematic search method.

During this study the scalar fitness function, defined based on the criterion presented in [11], was also used:

$$U_\mu = \left[1 + \sum_{k=1}^{M} \sum_{i=1}^{2H} \max\left(D_i^k\right)\right]^{-w_1} \cdot \sum_{k=1}^{M} \mu_{1/2}^{w_2}\left(\hat{D}^k\right), \tag{13}$$

where $\mu_{1/2}\left(\hat{D}^k\right)$ denotes median achieved for a set of the rest of details coefficients. Moreover, for the fitness function in the form (12) an additional decision variable was introduced, which represents an information about the modal shapes considered during the determination of its value.

In the following study the searching for the optimal solution (minimization of the fitness function) was realized by application of evolutionary algorithms [2, 21]. The classic optimization methods (e.g. gradient-based methods) could not be adapted in this context mainly because of the form of the criterion (10), which leads to the discontinuity of the fitness function. For this reason the multi-objective optimization process was carried out using evolutionary algorithm with sorting of non-dominated solutions [2]. Whereas in the case of the scalar fitness function the classic evolutionary algorithm was applied [21].

3. Procedure and results of detection and localization of delaminations

The analysis of detection and localization of delaminations in composite beams was carried out on the simulation data obtained from numerical models prepared with use of the finite element method. Four groups of cases were considered: delamination on the large area along the length and through-the-width of a beam (symbol 'l'), delamination on the large area along the length and on the limited area along the width (symbol 'il') and the same cases for the delaminations on the small area along the length (symbols 'sl' and 'isl'). Each of the considered groups of delaminations consisted of 11 cases: the delamination was modeled between the layers of 12-layered laminate. The schemes of considered groups were presented in Fig. 2.

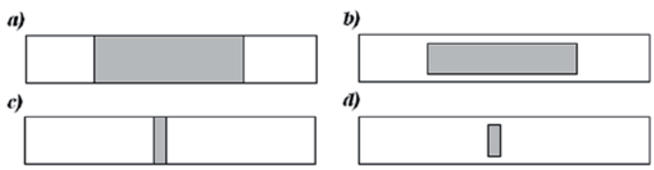


Fig. 2. Considered groups of delamination cases

3.1. Simulation data preparation

The numerical models of composite cantilever beams were prepared in the commercial software MSC Marc/Mentat. In order to simulate delaminations correctly, the beams were modeled as a three-dimensional ones with the following dimensions: length x – 200 mm, width w – 10 mm and a thickness h – 2.4 mm. The laminate consisted of 12 orthotropic layers with an equal thickness made of epoxy resin reinforced by a carbon cloth with the following material properties: Young moduli – $E_{11} = 82$ GPa, $E_{22} = 82$ GPa, $E_{33} = 8.5$ GPa; Kirchhoff moduli – $G_{12} = 5.2$ GPa, $G_{23} = 3.05$ GPa, $G_{31} = 3.47$ GPa and Poisson ratios – $\nu_{12} = 0.312$, $\nu_{23} = 0.29$, $\nu_{31} = 0.27$. The applied lay-up of the laminate is defined by the following structural formula: $[0/60/-60]_{2S}$. For the geometric models of beams the mesh of finite elements was defined. There were used hexagonal 8-node elements with the following dimensions in particular directions: 127 elements along the length – in order to fulfill the dyadic criterion of FrWT (see [23]), 5 elements along the width and 12 elements in the thickness direction. The ideal contact constraints were defined between the layers.

The delaminations were modeled by deactivation of the contact constraints in the selected regions. The locations of delaminations in the considered cases are as follows: for the groups 'l' and 'il': $x_1^0 = 85$ mm, $x_2^0 = 137$ mm and for thee groups 'sl' and 'isl': $x_1^0 = 95$ mm, $x_2^0 = 105$ mm. The numerical analysis was carried out for determination of the modal shapes of the beams. First five bending modal shapes were selected for further analysis. The displacements of these modal shapes were quantified along the length and in the half-width of the beams.

3.2. Selection of parameters of optimization algorithms

The application of an evolutionary algorithm for searching the optimal parameters of wavelets is connected with a necessity of definition of its fundamental properties. In the following study the process of single- and multi-objective optimization was carried out using the MATLAB® environment and the Genetic Algorithm Toolbox. The selection of algorithm parameters was realized following the requirements suggested in the literature [21]. The fitness function was declared basing on the criteria (10) and (11) for the case of multi-objective optimization and basing on metacriterion (12) or (13) for the one-dimensional case. Considering the results obtained in the previous study [10], the upper boundaries of the variability of the wavelet's parameters were defined as $\alpha_c = \tau_c = 18$. For the both cases the real-number coding of the individuals in a population was assumed, where the genes in particular chromosomes represented the wavelets parameters. The initial population was selected randomly (with the uniform distribution) with taking into consideration the assumed boundaries. The ranking method was applied for scaling the fitness function whereas the uniform stochastic selection method was employed for selection of the parents, which create the new individuals for the next initial population. The reproduction method was carried out basing on the elite succession (the number of individuals was 2) and the operators of crossover and mutation. The crossover function was realized using the heuristic method, where the descendant individual is created as a linear combination of genes of the parental individuals (the multiplication factor of the individual of the better fitness value was $\lambda = 1.2$). It was assumed that the crossover operation will be realized with the probability p_c. Remained parental individuals were processed using the adaptive mutation method, where the diminishing of a probability of genes mutation in chromosomes is dependent on the event, if in the last epoch the improvement of the fitness function took place.

In order to determine appropriate values of the crossover probability p_c and the number of individuals in a population N the results of research in the range of convergence of the evolutionary algorithm presented in [13] were used. The study was based on the systematic search of the combinations of values of these parameters (p_c = {0.4, 0.5, …, 1} and N = {5, 10, 20, 30}). Finally, the following values were assumed: p_c = 0.6 i N = 30, for which the lowest averaged values of

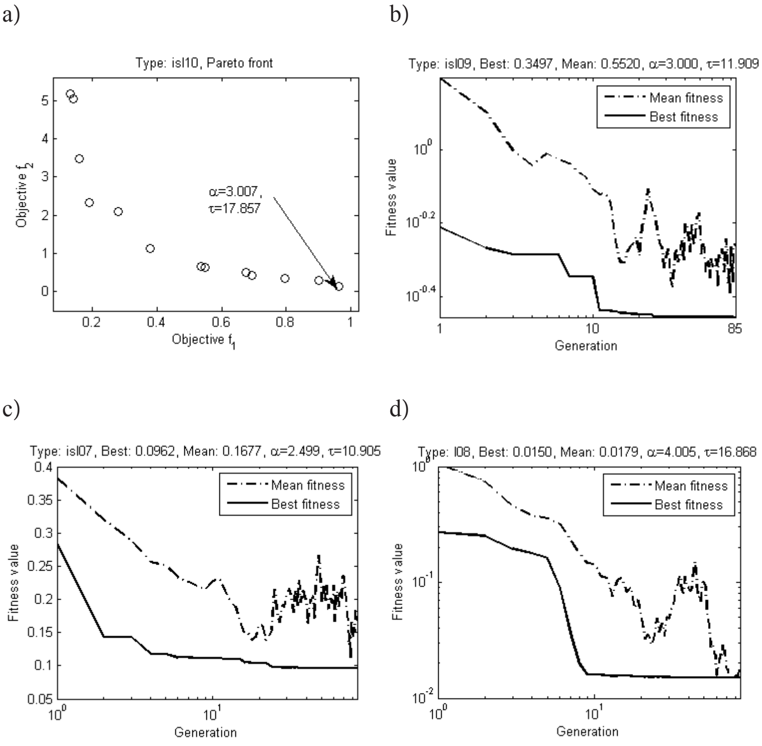


Fig.3. Exemplary results of multi-objective (a) and single-objective (b-d) optimization

the fitness function with the minimal standard deviation of the results were obtained.

Fig. 3 presents the selected results of optimization, which was performed following various strategies. The first case (Fig. 3a) shows the optimal solution in the Pareto sense. The optimal parameters of a wavelet ($\alpha = 3.007$, $\tau = 17.857$) were selected for a case, for which the minimal value of the noise and simultaneously the maximal possible value of the magnitude of peaks (coefficients) on the boundaries of the delamination were obtained. Fig. 3b-d present the results of single-objective optimization which was carried out for the fitness function: b) in the form of metacriterion (12) with the weights of $w_1 = w_2 = 0.5$; c) similarly as in the previous case, but only for the second modal shape; d) in the form of metacriterion (13). In the further part of the paper the detailed comments concerned these cases will be presented.

3.3. Analysis of results of localization of delaminations

The analysis of effectiveness of the delamination localization was carried out with use of the decomposition algorithm described in the section 2.2 and using the algorithm of optimal selection of parameters of B-spline wavelets of fractional order presented in the section 2.3. During the analysis the strategies of multi-objective optimization following the criteria (10) and (12) and single-objective optimization following the metacriterion (12) were considered. Moreover, the comparative studies for the metacriterion (13), which also was a topic of research in [11], were carried out. The results of detection and localization of delaminations for the cases of the single- and multi-objective optimization were comparable. Considering a great capability of single-objective optimization in the practical application further the results obtained for this case were described in detail.

The evaluation of results of detection and localization of delaminations was realized basing on the two measures described as follows:

$$m_{diff} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{abs}\left(Diff_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{abs}\left[\left(x_2^0 - x_1^0\right) - \left(x_2^{opt} - x_1^{opt}\right)\right]_i, \quad (9)$$

$$m_{cd} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{abs}\left(Err_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{abs}\left[\frac{1}{2}\left(x_2^0 - x_1^0\right) - \frac{1}{2}\left(x_2^{opt} - x_1^{opt}\right)\right]_i, \quad (10)$$

where m_{diff} describes the mean absolute value from the difference of real positions of the boundaries of the delamination x_1^0 i x_2^0 and the locations of the boundaries of a delamination detected by the proposed algorithm x_1^{opt} i x_2^{opt}; m_{cd} describes in the same way the mean value of the deviation of geometric center of the real and detected delamination. Obtained results for the four considered groups of damages were stored in Table 1. In this table the data for the locations of delaminations between various layers with enumeration from the bottom of beams was presented, i.e. the case 1 denotes the delamination between the layers 1 and 2, the case 2 – between the layers 2 and 3. In order to visualize the results from each group two representative cases were selected. They were presented in Fig. 4.

As it could be noticed, the results of delaminations localization presented in Fig. 4 in many cases are close to the real locations of the delaminations. In many considered cases, especially in the groups 'sl' and 'isl', the order of a wavelet is integer, however there is no characteristic relation between α and τ for the B-spline wavelets of the integer order: $\tau = (\alpha + 1)/2$. The tabulation of the values of wavelets parameters for the considered cases was presented in Table 2.

Basing on the obtained results one could conclude that the optimal order of the wavelet is consisted in the range of $\alpha \in [2,4]$. This fact is confirmed by the results of previous studies presented in [11]. The occurrence of the α value in this range is justified by the lowest number of moments of wavelets and their shortest support, which have an influence on the quicker decay of energy of a wavelet from the center to the boundaries of the support. This allows for minimization

Table 1. Averaged measures of sensitivity obtained during optimization following the metactriterion (12) for the considered locations of delaminations

Symbol of a group	'l'		'il'		'sl'		'isl'	
No. of a case	m_{diff}, mm	m_{cd}, mm	m_{diff}, mm	m_{cd}, mm	m_{diff}, mm	m_{cd}, mm	m_{diff}, mm	m_{cd}, mm
1	20.0	10.0	38.6	19.3	25.2	12.6	4.5	2.2
2	21.5	10.7	23.5	11.7	25.7	12.9	20.0	10.0
3	6.2	3.1	18.9	9.5	26.2	13.1	30.6	15.3
4	23.1	11.5	18.1	9.0	24.9	12.4	30.5	15.2
5	27.7	11.8	29.7	14.9	23.2	11.6	32.2	16.1
6	19.3	9.7	31.5	15.8	27.4	13.7	33.4	16.7
7	18.8	9.4	24.7	12.3	26.1	13.0	31.4	15.7
8	17.5	8.7	28.8	14.4	24.1	12.0	35.4	17.7
9	18.9	9.4	12.4	6.2	25.7	12.9	36.1	18.1
10	29.5	14.8	16.3	8.1	26.6	13.3	38.1	19.0
11	19.1	9.5	13.6	6.8	30.1	15.0	31.1	15.5

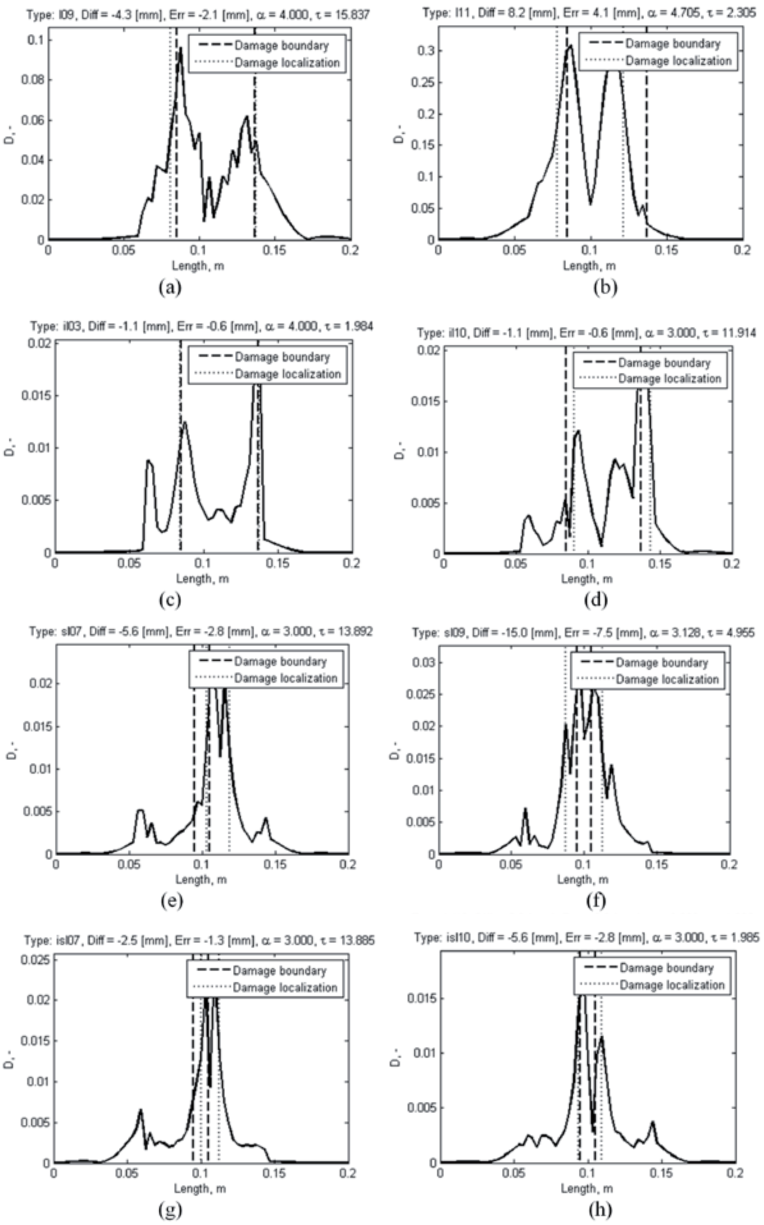


Fig. 4. Visualization of localization of delaminations for the considered groups of the cases

of disturbances near the detected boundaries of the delamination (cf. e.g. Fig. 4a and Fig. 4h).

For some of the considered cases of delaminations the incorrectly detected boundaries of the delaminations were observed. Exemplary results for these cases were presented in Fig. 5. There are two reasons which have an influence on the incorrect detection of the delaminations boundaries. As it could be observed, for the cases presented in Fig. 5a and Fig. 5b the peak values for the tight band of delamination are consisted in the range of [0.12, 0.145] of the length of a beam, however the width of this range was not sufficient for defining both of the boundaries on it and the optimization algorithm was tended to finding the additional peak value in order to fulfill the defined criterion. In the rest of the considered cases (Fig. 5c, d) in the sets of details coefficients the peak values were occurred, which exceeded the values of details coefficients of real locations of the delaminations boundaries. It was caused by the generation of the random initial population in the evolutionary algorithm, which resulted in incorrect selection of optimal parameters of the scaling and wavelet functions applied to the problem of detection and localization of delamination boundaries. It should be mentioned that the values τ for these cases achieve the highest values from the considered cases (see Tab. 2). These values cause the braking of the symmetry of a wavelet, which influences unfavorably on the obtained details coefficients [10]. In order to illustrate this phenomenon two extreme values of τ from the presented results were selected: the wavelet, whose parameters are close to the B-spline wavelet of an integer order: $\alpha = 3$, $\tau = 1.985$ (Fig. 4h) (for the wavelet of an integer order the value of τ will be equal 2) and the wavelet used in the case presented in the Fig. 5c: $\alpha = 3$, $\tau = 17.851$. The comparison of these wavelets was presented in Fig. 6.

As it was mentioned, the selection of single objective optimization, which was preformed following the metacriterion (12) was conditioned by obtaining quantitatively and qualitatively better results of detection and localization of the delaminations boundaries. Furthermore, the selection of such a method for optimization of wavelets parameters was justified by the greatest potential of this method in the further practical applications with respect to the solutions based on the multi-objective optimization. In order to compare these approaches the studies with usage of the metacriterion (13) were carried out, which was the problem of interest in [11]. During this stage of research the great attention should be paid to the most difficult detectable group of delaminations described as 'isl'. The cases 'isl7' (Fig. 4g) and 'sil10' (Fig. 4h) were selected. Obtained results of detection and localization with application of optimization following the metacriterion (13) were presented in Fig. 7.

Results presented in Fig. 7 reveal that the metacriterion (13) is not adjusted to the considered problem and justified the selection of the metacriterion (12) in the performed analyzes.

Table 2. Tabulation of the values of parameters of B-spline wavelets obtained during the optimization following the metacriterion (12)

Symbol of a group	'l'		'il'		'sl'		'isl'	
No. of a case	α, -	τ, -	α, -	τ, -	α, -	τ, -	α, -	τ, -
1	6.345	17.704	4.000	16.871	3.000	9.918	3.000	3.965
2	6.408	17.787	5.130	14.829	3.000	17.851	3.000	14.878
3	4.863	10.907	4.000	1.984	3.000	7.940	3.000	17.854
4	4.893	5.404	2.161	14.873	3.000	11.906	3.000	15.870
5	5.461	2.472	4.293	16.855	3.000	3.965	3.000	8.924
6	3.188	15.879	3.000	2.983	3.000	17.850	3.000	15.870
7	4.962	11.781	4.647	15.867	3.000	13.892	3.000	13.885
8	5.123	9.805	3.073	1.991	3.000	1.981	3.000	9.917
9	4.000	15.837	2.000	16.921	3.128	4.955	3.000	11.909
10	4.186	2.299	3.000	11.914	3.000	15.874	3.000	1.985
11	4.705	2.305	2.000	16.920	3.000	17.857	3.000	8.930

3.4. Influence of the number of considered modal shapes on the effectiveness of detection and localization of delaminations

In the presented algorithm of detection and localization of delamination boundaries first five bending modal shapes were considered. However, the selection of the modal shapes, which contained the maximal quantity of the diagnostic information about the damages will allow for increasing of a sensitivity of the method for occurred damages by

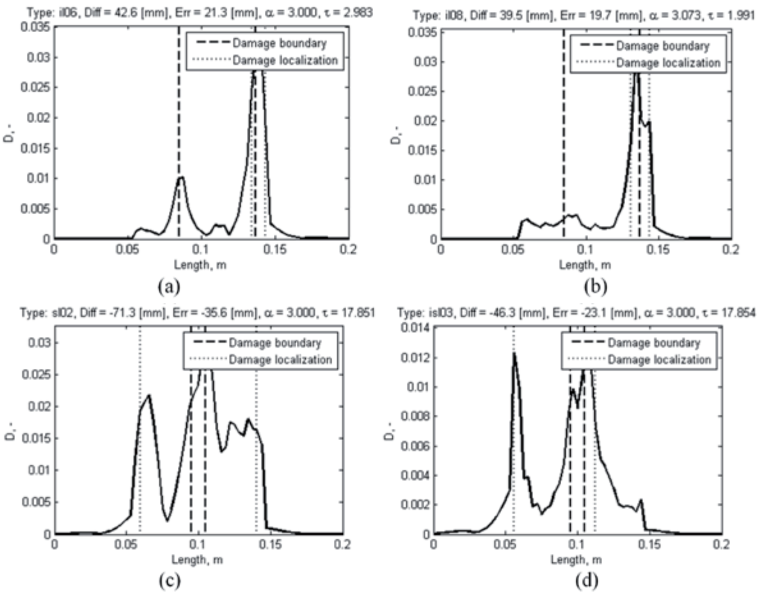


Fig. 5. Examples of incorrectly localized delaminations

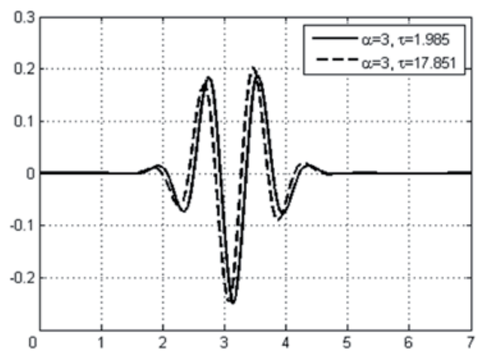


Fig. 6. Comparison of the B-spline wavelets of fractional order with different values of the shift parameter

omitting the modal shapes, which do not bring new diagnostic information and simultaneously to be the source of noise added to the resulted coefficients D. Additionally, it will allow for reduction of quantity of processed data in the case of considering of the lower number of modal shapes, which could accelerate the processing of the detection and localization algorithm.

In order to investigate the influence of considered modal shapes in the analysis the additional optimization problem following the metacriterion (12) was formulated with the additional decision variable, which represented the number and identifiers of the considered modal shapes. The weights of minimization of the noise level and searching for the maximal values of the details coefficients described as equal, i.e. $w_1 = w_2 = 0.5$. The results of analyzes for the selected cases (one for each considered group of damages – Fig. 2) were presented in Fig. 8.

Besides the defined metacriterion, which assumes the consideration of variance from all of the five modal shapes, each time only one of them was selected. The reason of such behaviour of an algorithm is the equal values of the weights of criteria concerned with minimization of the noise and searching for the maximal values of the details coefficients.

Basing on the results presented in Fig. 8 it could be observed that in the case of a delamination from the group 'l' and considering only the second modal shape the boundaries of a delamination were detected with high precision, whereas for the case from the group 'il' one of the boundaries of a delamination was detected incorrectly. In this case it was reasoned by low values of magnitude of displacements in the first modal shape selected by an optimization algorithm. In the other two cases (Fig. 8c,d) the delaminations were localized incorrectly, in each case the third modal shape was selected.

In order to obtain the cases, where the multiple modal shapes were considered it is suitable to change the ratio of weights w_1 i w_2 in a metacriterion (12), where w_1 is the weight responsible for maximization of the peak values of details coefficients on the boundaries of a delamination, while w_2 is responsible for the minimization of noise in the signal. Additionally six cases of weights ratios were considered: $w_1 = 0.8$, $w_2 = 0.2$; $w_1 = 0.2$, $w_2 = 0.8$; $w_1 = 0.6$, $w_2 = 0.4$; $w_1 = 0.55$, $w_2 = 0.45$ and two extreme cases $w_1 = 0.95$, $w_2 = 0.05$; $w_1 = 0.05$, $w_2 = 0.95$. In order to compare all of the considered cases the numbers of cases with correctly localized boundaries of delaminations m and the ranges of considered number of layers M for each of the considered groups were prepared (see Table 3). The cases, where the difference between the real and detected boundaries of delaminations was lower than 15 mm were assumed as cases with correct localization.

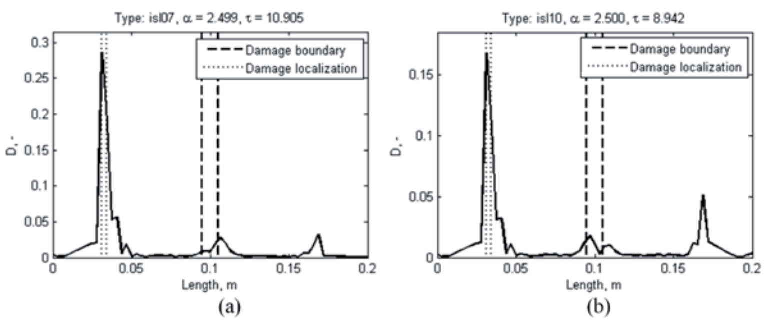


Fig. 7. Results of localization of delaminations boundaries obtained during optimization following the metacriterion (13)

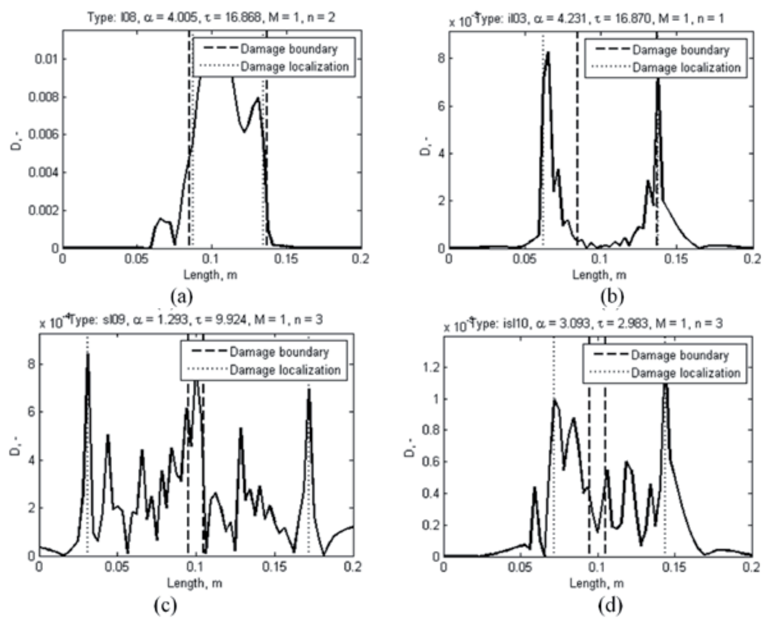


Fig. 8. Results of localization of delaminations boundaries with additional optimization of the number and enumeration of considered modal shapes

Table 3. Comparison of the effectiveness of delamination detection with respect to optimization of number of considered modal shapes and various combinations of weights in the metacriterion (12)

No. of a case	w_1, -	w_2, -	The number of cases with correct localization of delaminations and the number of considered modal shapes							
			'l'		'il'		'sl'		'isl'	
			m, -	*M*, -	*m*, -	*M*, -	*m*, -	*M*, -	*m*, -	*M*, -
1	0.5	0.5	5	1	4	1	0	1	2	1
2	0.8	0.2	4	1÷4	0	3÷5	0	3÷5	0	3÷5
3	0.2	0.8	7	1	1	1	0	1	2	1÷2
4	0.6	0.4	8	1	5	1÷2	2	1÷2	1	1÷2
5	0.55	0.45	8	1	6	1	2	1÷2	3	1÷2
6	0.95	0.05	2	3÷5	0	3÷5	0	3÷5	0	4÷5
7	0.05	0.95	9	1	6	1	1	1÷2	2	1÷2
Without optimization of the number of considered modal shapes			10	5	4	5	10	5	8	5

The results of performed analyzes show that the weight w_2, which was responsible for the noise minimization, has the much greater influence on the correct localization of the boundaries of delamination. During this process the lower number of modal shapes is considered, which was affected by an input of additional disturbances during adding each additional modal shapes considered in the analysis. Additionally, the results presented in Table 3 pointed to the difficulties of an algorithm in the detection of delaminations in the areas, which do not reach the boundaries of the specimens along the width (the groups of cases 'il' and 'isl'). The reason of such errors in the localization of the delamination boundaries for these cases is the much lower displacements in the areas of delaminations during resonant vibrations.

4. Conclusions

In the paper the new method of structural diagnostics of composite beams oriented to the detection and localization of delaminations boundaries was presented. The delaminations are the one of the most difficult damages to detect by applying the methods based on the modal analysis. The results of the studies reveal a potential of a method in the detection of such damages thanks to the application of evolutionary algorithms for single- and multi-objective optimization of parameters of B-spline wavelets of fractional order. The research results show that the method is characterized by the high effectiveness in the detection and localization of through-the-width delaminations. In the case of the internal delaminations the method reveals the lower effectiveness and the recognition of the delamination is about 50%. It was observed that the order of applied wavelets in the considered cases was relatively low and usually was located in the range of $\alpha \in [2,4]$. This is because that the optimization algorithm selected the wavelets with short supports, which in consequence generated the lower disturbances during the application of the wavelet transform. Moreover, the best results were achieved for such cases, where the value of a shift parameter τ was selected by an optimization algorithm in such a way that the wavelet tended to the symmetry with respect to the center of its support.

The discrepancies between the groups of damages in delaminations localization presented in the paper could be reduced by the development of separate optimization criteria for various shapes of delaminations, which is planned in the further studies. Furthermore, the increase of accuracy of damage detection with the use of the proposed algorithm could be achieved due to the creation of an initial population from the cases, for which the damages were detected and localized correctly in the presented analyzes (see Table 2). After performing analyzes on the simulation data it is planned to verify the method experimentally.

Acknowledgements*: The research project was financed by the National Science Centre (Poland) granted according the decision no. DEC-2011/03/N/ST8/06205.*

References

1. Blu T, Unser M. A complete family of scaling functions: the (α,τ)-fractional splines. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing 2003; 6: 421-424.
2. Deb K. Multi-objective optimization using evolutionary algorithms. Wiley, 2009.
3. Douka E, Loutridis S, Trochidis A. Crack identification in beams using wavelet analysis. International Journal of Solids and Structures 2003; 40: 3557-3569.
4. Dumont J, Hernandez A, Carrault G. Improving ECG beats delineation with an evolutionary optimization process. IEEE Transactions on Biomedical Engineering 2010; 57(3): 607-615.
5. Ghribi SF, Jammoussi AY, Masmoudi DS. A multi objective genetic algorithm based optimization of wavelet transform implementation for face recognition applications. World Academy of Science, Engineering & Technology 2011; 56: 1590-1593.
6. Ghribi SF, Masmoudi DS, Derbel N. A multi objective genetic algorithm based optimization of wavelet transform implementation. Proc. of 3rd International Design and Test Workshop, Monastir 2008; 87-91.
7. Hein H, Feklistova L. Computationally efficient delamination detection in composite beams using Haar wavelets. Mechanical Systems and Signal Processing 2011; 25: 2257-2270.
8. Katunin A. Identification of multiple cracks in composite beams using discrete wavelet transform. Scientific Problems of Machines Operation and Maintenance 2010; 45: 41-52.
9. Katunin A. The construction of high-order B-spline wavelets and their decomposition relations for fault detection and localisation in composite beams. Scientific Problems of Machines Operation and Maintenance 2011; 46: 43-59.

10. Katunin A. Damage identification in composite plates using two-dimensional B-spline wavelets. Mechanical Systems and Signal Processing 2011; 25: 3153-3167.
11. Katunin A. Crack identification in composite beam using causal B-spline wavelets of fractional order. Modelowanie Inżynierskie 2013; 15: 57-63.
12. Katunin A, Holewik F. Crack identification in composite elements with non-linear geometry using spatial wavelet transform. Archives of Civil and Mechanical Engineering 2013; 13: 287-296.
13. Katunin A, Przystałka P. Structural diagnostics of composite beams using optimally selected fractional B-spline wavelets. Intelligent Systems in Technical and Medical Diagnostics, Korbicz J., Kowal M., Eds., Advances in Intelligent Systems and Computing 2014; 230: 475-486.
14. Loutridis S, Douka E, Trochidis A. Crack identification in double-cracked beams using wavelet analysis. Journal of Sound and Vibration 2004; 277: 1025-1039.
15. Mallat S. A theory of multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 1989; 11: 674-693.
16. Pai PF, Young LG. Damage detection of beams using operational deflection shapes. International Journal of Solids and Structures 2001; 38: 3161-3192.
17. Rafiee J, Tse PW, Harifi A, Sadeghi MH. A novel technique for selecting mother wavelet function using an intelligent fault diagnosis system. Expert Systems with Applications 2009; 36: 4862-4875.
18. Ratcliffe CP, Bagaria WJ. Vibration technique for locating delamination in a composite beam. American Institute for Aeronautics and Astronautics Journal 1998; 36: 1074-1077.
19. Rucka M, Wilde K. Application of continuous wavelet transform in vibration based damage detection method for beams and plates. Journal of Sound and Vibration 2005; 27: 1327-1338.
20. Rucka M, Wilde K. Crack identification using wavelets on experimental static deflection profiles. Engineering Structures 2006; 28: 279-288.
21. Rutkowski L. Methods and techniques of artificial intelligence (in Polish). PWN, Warsaw, 2005.
22. Unser M, Aldroubi A, Eden M. On the asymptotic convergence of B-spline wavelets to Gabor functions. IEEE Transactions on Information Theory 1992; 38: 864-872.
23. Unser M, Blu T, Fractional splines and wavelets. SIAM Review 2000; 42: 43-67.
24. Zhong S, Oyadiji SO. Detection of cracks in simply-supported beams by continuous wavelet transform of reconstructed modal data. Computers and Structures 2011; 89: 127-148.
25. Zou Y, Tong L, Steven GP. Vibration based model-dependent damage (delamination) identification and health monitoring for composite structures – a review. Journal of Sound and Vibration 2000; 230: 357-378.

Andrzej KATUNIN
Piotr PRZYSTAŁKA
Institute of Fundamentals of Machinery Design
Silesian University of Technology
ul. Konarskiego 18A, 44-100 Gliwice, Poland
E-mails: andrzej.katunin@polsl.pl, piotr.przystalka@polsl.pl

Article citation info:
MUTINGI M. System reliability optimization: A fuzzy multi-objective genetic algorithm approach. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 400–406.

Michael MUTINGI

SYSTEM RELIABILITY OPTIMIZATION: A FUZZY MULTI-OBJECTIVE GENETIC ALGORITHM APPROACH

OPTYMALIZACJA NIEZAWODNOŚCI SYSTEMU: METODA ROZMYTEGO ALGORYTMU GENETYCZNEGO DO OPTYMALIZACJI WIELOKRYTERIALNEJ

System reliability optimization is often faced with imprecise and conflicting goals such as reducing the cost of the system and improving the reliability of the system. The decision making process becomes fuzzy and multi-objective. In this paper, we formulate the problem as a fuzzy multi-objective nonlinear program. A fuzzy multi-objective genetic algorithm approach (FMGA) is proposed for solving the multi-objective decision problem in order to handle the fuzzy goals and constraints. The approach is able flexible and adaptable, allowing for intermediate solutions, leading to high quality solutions. Thus, the approach incorporates the preferences of the decision maker concerning the cost and reliability goals through the use of fuzzy numbers. The utility of the approach is demonstrated on benchmark problems in the literature. Computational results show that the FMGA approach is promising.

***Keywords**: System reliability optimization, multi-objective optimization, genetic algorithm, fuzzy optimization, redundancy.*

Często spotykanym problemem w optymalizacji niezawodności systemu są niedokładnie określone i sprzeczne cele, takie jak zmniejszenie kosztów systemu przy jednoczesnej poprawie jego niezawodności. Proces podejmowania decyzji staje się wtedy rozmyty i wielokryterialny. W niniejszej pracy, sformułowaliśmy ten problem jako rozmyty wielokryterialny program nieliniowy (FMOOP). Zaproponowaliśmy metodę rozmytego wielokryterialnego algorytmu genetycznego (FMGA), która pozwala rozwiązać wielokryterialny problem decyzyjny z uwzględnieniem rozmytych celów i ograniczeń. Podejście to jest uniwersalne, co pozwala na rozwiązania pośrednie, prowadzące do rozwiązań wysokiej jakości. Metoda uwzględnia preferencje decydenta w zakresie celów związanych z kosztami i niezawodnością poprzez wykorzystanie liczb rozmytych. Użyteczność FMGA wykazano na przykładzie wzorcowych problemów z literatury. Wyniki obliczeń wskazują, że podejście FMGA jest obiecujące.

***Słowa kluczowe**: Optymalizacja niezawodności systemu, optymalizacja wielokryterialna, algorytm genetyczny, optymalizacja rozmyta, nadmiarowość.*

1. Introduction

System reliability optimization is a very important subject matter in industry. Reliable systems are essential for sustainable productivity and competitiveness in modern industry [22, 24–25, 31]. To maximize productivity, industrial systems, such as manufacturing systems, must be available and operational as much as possible. Nevertheless, since industrial systems consist of a number of components, the ultimate probability of system survival directly depends on the characteristics of the constituent components. Hence, system failure is inevitable. As such, it is essential to enhance system reliability through suitable reliability optimization methods, so as to improve the overall system productivity. Developing effective methods for system reliability enhancement is imperative.

The ever-increasing need for highly reliable systems necessitates the search for improved methods for system reliability optimization. In system reliability design, two typical approaches can be used to enhance system reliability: (i) adding redundant components in the subsystems of the system, and (ii) increasing the reliability of the components that constitute the system.

Industrial systems are designed under several restrictions, including cost, weight, and volume of the resources. With limited resources, the major aim is to find a trade-off between reliability and other resource constraints [22]. One of the feasible ways is to maximize system reliability via redundancy and component reliability choices, a problem called reliability-redundancy allocation problem [24]. However, in designing a highly reliability system, the main problem is to find a trade-off between reliability enhancement and resource consumption. This calls for an application of a suitable multi-criteria approach. Various multi-criteria programming approaches and multi-criteria solution approaches have been applied on different problems in the literature [1–3, 23].

In the real world, system reliability optimization problems are inundated with a number of uncertainties and difficulties. This is due to the reasons that: (i) the management goals and the constraints are often characterised with some imprecision or vagueness; (ii) the coefficients or parameters as understood by the decision maker may be characterized with some vagueness; and, (iii) the available historical data, collected under specific conditions, are often imprecise and vague. In addition, variability and changes in the manufacturing processes that produce the components of the systems lead to uncertainties in component reliability. Probabilistic approaches, which essentially deal with uncertainty arising from randomness, cannot adequately address inherent uncertainties in the data. While probabilistic approaches deal with uncertainties arising from randomness, fuzzy approaches seek to address uncertainties that arise from vagueness of human judgment and imprecision due to system complexity [4–6, 13–15, 27]. As a result, the concept of fuzzy reliability is more promising [7–9, 28].

Bellman and Zadeh [5] introduced the fuzzy optimization approach that utilizes aggregation operators for combining fuzzy goals and fuzzy decision space. Since the inception of the fuzzy optimization approach, a number of methods and applications have been proposed to solve optimization problems that involve vagueness and ambiguity [12, 20, 21, 30]. These approaches treat parameters (coefficients) as fuzzy numerical data. Apart from the fuzziness of the system reliability problem, the presence of conflicting, nonlinear and ambiguous objectives further complicates the problem. In such a fuzzy environment, with multiple objectives, simultaneous reliability maximization and cost minimization calls for a cautious trade-off approach. Thus, finding the optimal solution is almost impossible. Metaheuristic and other intelligent methods are a potential application method for such complex problems [11, 10, 26]. Therefore, the most appropriate procedure is to cautiously find a set of solutions that satisfy the decision maker's expectations to the highest possible degree. Clearly, this calls for an interactive fuzzy multi-objective optimization approach which incorporates the preferences and expectations of the decision maker, allowing for human (expert) judgment. Iteratively, it becomes possible to obtain the most satisfactory solution in a fuzzy environment.

In view of the above issues, the purpose of this paper is to address the problem of system reliability optimization in a fuzzy environment characterized with multiple conflicting objectives. Therefore, our specific objectives are as follows:

(1) to develop a fuzzy multiple-objective nonlinear programming model for the reliability optimization problem;
(2) to use an aggregation method to transform the fuzzy model to a single-objective optimization problem; and,
(3) to use a global metaheuristic optimization method to obtain a set of acceptable solutions.

In our current study, we develop a fuzzy multi-objective genetic algorithm (FMGA) which utilizes a fuzzy theory based method to evaluate the objective functions represented as membership functions. We use the max-min operator to aggregate the membership functions of the objective functions while incorporating the decision maker's judgment. In this respect, we define our notations and assumptions as follows.

Nomenclature:

m	the number of subsystems in the system
n_i	the number of components in subsystem i, $1 \le i \le m$
n	$\equiv(n_1, n_2, \ldots, n_m)$, the vector of the redundancy (number of redundant components) allocation for the system
r_i	the reliability of each component in subsystem i, $1 \le i \le m$
r	$\equiv(r_1, r_2, \ldots, r_m)$, the vector of the component reliabilities for the system
q_i	$=1 - r_i$, the failure probability of each component in subsystem i, $1 \le i \le m$
$R_i(n_i)$	$=1-q_i^{n_i}$, the reliability of subsystem i, $1 \le i \le m$
R_s	the system reliability
g_i	the i^{th} constraint function
w_i	the weight of each component in subsystem i, $1 \le i \le m$
v_i	the volume of each component in subsystem i, $1 \le i \le m$
c_i	the cost of each component in subsystem i, $1 \le i \le m$
V	the upper limit on the sum of the subsystems' products of volume and weight
C	the upper limit on the cost of the system
W	the upper limit on the weight of the system
b	the upper limit on the resource

Assumptions:

(1) The availability of the components is unlimited;
(2) The weight and product of weight and square of the volume of the components are deterministic;
(3) The redundant components within the individual subsystems are identical;
(4) Failures of individual components are independent;
(5) All failed components will not damage the system and are not repaired.

2. System reliability optimization

The system reliability optimization problem is a maximization problem subject to multiple non-linear constraints. In this connection, the problem can be expressed as a mixed integer nonlinear programming problem. In this study, we present a reliability redundancy problems commonly found in the literature, with a particular emphasis on the series system [22, 24]. The series system reliability problem consists of five subsystems as reported in the literature as shown in Fig. 1.

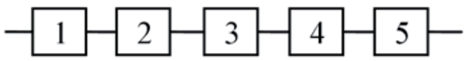


Fig. 1. The series system

Following our notation proposed in section 1, the system reliability optimization problem can be formulated as a nonlinear mixed integer program:

$$\text{(P1)} \quad \text{Max}\, f(\mathrm{r},\mathrm{n}) = \prod_{i=1}^{m} R_i(n_i)$$

Subject to:

$$g_1(\mathrm{r},\mathrm{n}) = \sum_{i=1}^{m} w_i v_i^2 n_i^2 \le V$$

$$g_2(\mathrm{r},\mathrm{n}) = \sum_{i=1}^{m} \alpha_i \left(-1000/\ln r_i\right)^{\beta_i} \left(n_i + \exp(n_i/4)\right) \le C$$

$$g_3(\mathrm{r},\mathrm{n}) = \sum_{i=1}^{m} w_i n_i \exp(n_i/4) \le W$$

$$0 \le r_i \le 1,\ \ n_i \in Z^+,\ \ 1 \le i \le m$$

where, r_i, and n_i, are the reliability and the number of components in the i^{th} subsystem respectively; $f(\cdot)$ is the objective function for the overall system reliability; $g(\cdot)$ is the constraint function; m si the number of subsystems. The primary goal is to determine the number of components and their reliability in each subsystem so that the overall system reliability is maximized. Thus, the problem falls in the category of constrained non-linear mixed integer optimization problems. The next section presence the proposed fuzzy multi-objective optimization approach, based on genetic algorithm.

3. Fuzzy multi-objective optimization approach

In a fuzzy environment, the objective goal, the constraints and the consequences of the decision taken are inherently imprecise. Thus, in practice, the decision maker seeks to consider a trade-off between reliability, cost, weight and volume. For instance, a common approach may be to simultaneously maximize reliability and minimize cost. In this connection, the multi-objective formulation is obtained by transforming constraints to objective functions, such that reliability and other costs functions can be optimized jointly. This is achieved through the use of membership functions for the objective functions. This makes the approach more applicable and adaptable to the real life human decision process. Therefore, the fuzzy multi-objective optimization problem (FMOOP) can generally be represented as follows;

$$\text{(P2) Min } \tilde{f}(x)$$
$$\text{Subject to:}$$
$$g_z(x) \lesssim \text{ or } \equiv \text{ or } \gtrsim 0 \quad z = 1,2\ldots,p$$
$$x_q^l \le x_q \le x_q^u \qquad q = 1,2,\ldots,Q$$

where, $x = (x_1, x_2, \ldots, x_Q)^T$ is a vector of decision variables that optimize a vector of fuzzy objective functions, $\tilde{f}(x) = \{\tilde{f}_1(x), \tilde{f}_2(x), \ldots, \tilde{f}_d(x)\}$ over the decision space X; $\tilde{f}_1(x), \tilde{f}_2(x), \ldots, \tilde{f}_d(x)$ are d individual objective functions; x_q^l and x_q^u are lower and upper bounds on decision variable x_q, respectively. Here, we use the symbol "~" to denote a fuzzy function or operator.

3.1. Membership functions

The notion of fuzzy set theory permits gradual assessment of membership, defined in terms of a suitable membership function that maps to the unit interval [0,1]. To date, several membership functions, such as Generalized Bell, Gaussian, Triangular and Trapezoidal have been used to represent fuzzy membership in a several applications. Though various functions can be used, it has been shown that linear membership functions can provide equally good quality solutions with much ease [29]. The triangular and trapezoidal membership functions have widely been recommended [8, 12, 29]. In this study, therefore, we use linear functions to define the fuzzy membership functions of the objective functions.

In this study, we assume that an expert user has a range of acceptable feasible values of each objective functions. However, we further assume that there is a lower and upper limit to that range of acceptable objective function values, as specified by the expert user. Let m_t and M_t denote the minimum and maximum acceptable values of each objective function $\tilde{f}_t(x)$, $t = 1,2,\ldots,h$, where h is the number of objective functions. Further, let μ_{f_t} denote the membership function corresponding to the objective function f_t. Then, the membership function corresponding to minimization and maximization of specific objective functions can be defined in terms of degree of satisfaction. Fig. 2 illustrates the linear membership functions, both for minimization as well as for maximization problems. We define the membership functions for both situations.

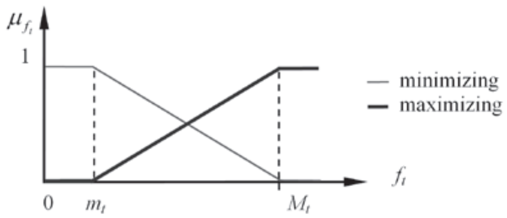


Fig. 2. Fuzzy membership function for $f_t(x)$

When the objective is concerned with minimization, the linear membership function can be formulated as in the following expression:

$$\mu_{f_t}(x) = \begin{cases} 1 & f_t(x) \le M_t \\ \dfrac{M_t - f_t(x)}{M_t - m_t} & m_t \le f_t(x) \le M_t \\ 0 & f_t(x) \ge M_t \end{cases} \tag{1}$$

where, m_t and M_t denote the minimum and maximum acceptable feasible values of each objective function. Clearly, the function $\mu_{f_t}(x)$ is monotonically decreasing in $f_t(x)$. On the other hand, when the objective is about maximization, the membership function can be defined as follows:

$$\mu_{f_t}(x) = \begin{cases} 1 & f_t(x) \ge M_t \\ \dfrac{f_t(x) - m_t}{M_t - m_t} & m_t \le f_t(x) \le M_t \\ 0 & f_t(x) \ge m_t \end{cases} \tag{2}$$

It can be seen from this analysis that $\mu_{f_t}(x)$ is a monotonically increasing function of $f_t(x)$. The next step is to formulate the corresponding crisp model. The use of fuzzy evaluation in FMGA allows the algorithm to accept inferior which would otherwise be infeasible when using conventional crisp formulation. The advantage of this approach is that it makes the algorithm robust enough to cope with any infeasibility. Allowing the FMGA to pass through inferior solutions gives the algorithm speed and flexibility, which ultimately improves the search power of the approach.

3.2. Corresponding crisp model

In practice, it is desirable to consider the imprecise management or decision maker's preferences in our formulation. Therefore, to incorporate the decision maker's preferences and to enhance the interactive flexibility of the model, a set of user-defined weights $w = \{w_1, w_2, \ldots, w_h\}$ are introduced. We convert the multi-objective system reliability optimization problem into a single objective optimization problem [14]:

$$\text{(P3) Max } \left(\frac{\lambda_1(x)}{w_1} \wedge 1\right) \wedge \left(\frac{\lambda_2(x)}{w_2} \wedge 1\right) \wedge \ldots \wedge \left(\frac{\lambda_h(x)}{w_h} \wedge 1\right)$$
$$\text{Subject to:}$$
$$\lambda_t(x) = \mu_{f_t}(x) \qquad w_t \in [1,0)\ t = 1,\ldots,h$$
$$x_q^l \le x_q \le x_q^u \qquad q = 1,\ldots,Q$$

where, $\mu_{f_t}(x) = \{\mu_{f_1}(x), \mu_{f_2}(x), \ldots, \mu_{f_h}(x)\}$ is a set of fuzzy regions that satisfy the objective functions; x is a vector of decision variables, λ_t denotes the degree of satisfaction of the t^{th} objective, w_t denotes the weight of the t^{th} objective function as suggested by the expert judgment of the user or decision maker, and the symbol "∧" is the aggregate min operator or the intersection operator. For instance, the expression $(\lambda_1(x)/w_1)\wedge 1$ gives the minimum between 1 and $\lambda_1(x)/w_1$. Though the values of $\lambda_1(x)$ are in the range [0,1], the value of $\lambda_1(x)/w_1$ may exceed 1, howbeit, by the min operator the final value of $(\lambda_1(x)/w_1)\wedge 1$ will always lie in [0,1]. We use a metaheuristic approach to solve problem P3.

3.3. Genetic algorithm approach

Genetic Algorithm (GA) is a stochastic global optimization technique that attempts to evolve a population of candidate solutions by giving preference of survival to quality solutions, whilst allowing some low quality solutions to survive in order to maintain diversity in the population [16, 18]. Each candidate solution is coded into a string of digits, called chromosomes. New offspring are obtained from probabilistic genetic operators, such as selection, crossover, mutation, and inversion [16]. A comparison of new and old (parent) candidates is done based on a given fitness function, retaining the best performing candidates into the next population. Thus, characteristics of candidate solutions are passed from generation to generation through probabilistic selection, crossover, and mutation. The general flow of the GA approach is presented in Fig 3. The metaheuristic is represented as an iterative procedure consisting of sub-procedures: initialization, evaluation, selection, crossover, and mutation.

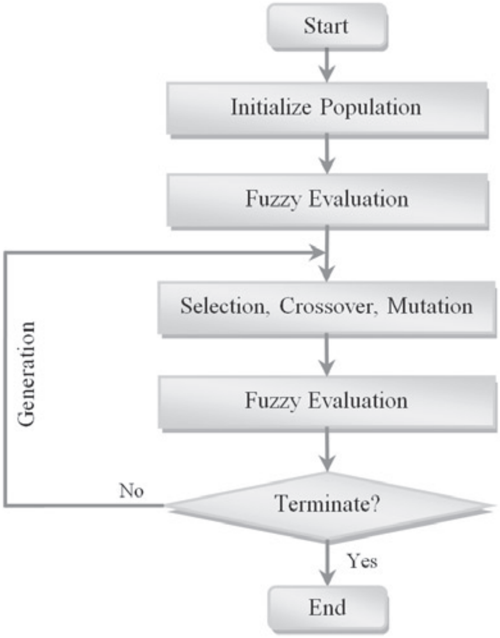


Fig. 3. Fuzzy multi-objective genetic algorithm approach

3.4. Genetic encoding scheme

In our FMGA implementation for the system reliability problem, the genetic chromosome uses the variable vectors n and r. Thus, we use a real-coded genetic encoding scheme, where the integer variable n_i is coded as a real variable and transformed to the nearest integer value upon evaluating the objective function.

3.5. Initialization and evaluation

In the initialization procedure, an initial population of the desired size, *pop*, is generated randomly from the solution space. FMGA then computes the objective function for each string according to the objective function represented in model P3. The value of the objective function is always in the range [0,1].

3.6. Selection and recombination

A number of selection strategies exist in literature [13]. In this study, we adopted the remainder stochastic sampling without replacement. By this strategy, each chromosome j is selected and stored in the mating pool according to the expected count e_j, represented by the expression;

$$e_j = \frac{f_j}{\sum_{j=1}^{pop} f_j / pop} \tag{3}$$

where, f_j is the objective function value of the j^{th} chromosome. Each chromosome receives copies equal to the integer part of e_i, that is, $[e_i]$, while the fractional part is treated as success probability of obtaining additional copies of the same chromosome into the mating pool. The crossover operator is then applied to selected parent chromosomes for the purpose of exchanging genetic information between the selected chromosomes, thereby producing new offspring. Here, we use the arithmetic crossover operator as in [26] to define a linear combination of two chromosomes. A crossover probability of 0.42 was assumed in this application. For instance, let p_1 and p_2 denote the selected parents, and α represent a random value in the range [0,1], then the resulting offspring, q_1 and q_2, are given by the following expression:

$$\begin{aligned} q_1 &= \alpha p_1 + (1-\alpha) p_2 \\ q_2 &= (1-\alpha) p_1 + \alpha p_2 \end{aligned} \tag{4}$$

3.7. Mutation operator

As generations proceed, the population converges to a common solution, which may lead to result in pre-mature convergence. To curb premature convergence, and to maintain population diversity, a mutation operator is applied to every new chromosome, at a very low probability. In our application, we used a uniform mutation with a mutation probability of 0.032.

3.8. Replacement

In every generation, new offspring are created. The new offspring may be better or worse than the preceding generation. As such, the non-performing individuals are replaced with better ones using a replacement strategy. According to Goldberg (1979) [16], some of the replacement strategies found in the literature include probabilistic replacement, crowding strategy, and elitist strategy. In this application, we a combination of these strategies was implemented.

3.9. Termination criteria

Two termination conditions are used to stop the FMGA iteration, that is, when the number of generations exceeds the preset maximum iterations, or when the average improvement in the fitness of the best solution over specific generations is less than a small number, which is assumed to be 10^{-6} in this application. The maximum generations was set at 500.

3.10. Overall FMGA procedure

The overall structure of the FMGA for the system reliability problems consists of all the procedures discussed in the previous sections; that is, initialization, selection, evaluation, crossover, mutation, replacement, and termination. Fig. 4 presents the pseudo-code of the algorithm.

Algorithm 1: Pseudo code for FMGA

```
1: randomly generate initial population
Repeat
  2: evaluation of fitness, objective: f(x), x = (x1, x2,…,xh)
  3: selection strategy
  4: crossover and mutation
  5: replacement
  6: advance population; oldpop = newpop
Until (termination criteria is satisfied)
```

Fig. 4. Pseudo code for the overall FMGA procedure

The next section presents the comparative results of our FMGA computations based on the benchmark problems found in the literature [17, 19, 24, 31].

4. Numerical experiments

To evaluate the usefulness of our proposed FMGA for solving mixed integer reliability problems, the series reliability system illustrated in P1 will be solved using the approach. We use the parameter values in [19] and to define the specific instances of this problems as shown in Tables 1.

Table 1. Basic data used in series system

i	$10^5\alpha_i$	β_i	$w_i v_i^2$	w_i	V	C	W
1	2.330	1.5	1	7	110	175	200
2	1.450	1.5	2	8	110	175	200
3	0.541	1.5	3	8	110	175	200
4	8.050	1.5	4	6	110	175	200
5	1.950	1.5	2	9	110	175	200

The parameters of the FMGA were set as follows: The crossover and mutation were set at 0.45 and 0.035, respectively. A two-point crossover was used in this application. The population size was set to 20. The maximum number of generations or iterations was set at 150. This implies that the termination criterion is either limited to a maximum number of iterations or to the order of the relative error set at 10^{-6}, whichever comes earlier. Specifically, whenever the best fitness f^* at iteration t is such that $|f_t - f^*| < \varepsilon$ is satisfied, then three best solutions are selected; where ε is a small number equal to 10^{-6}. The FMGA was implemented in JAVA, and the program was run 25 times, while selecting the best 3 solutions out of the converged population.

The FMOOP provided by formulation (P3) is used to solve benchmark problems in [19]. A fuzzy region of satisfaction is constructed for each objective function, that is, objective functions corresponding to system reliability, cost, volume, and weight, which are denoted by λ_1, λ_2, λ_3, and λ_4, respectively. By using the constructed membership functions together with their corresponding weight vectors, we obtain an equivalent crisp optimization formulation for our problem:

$$\text{(P4) Max } \left(\frac{\lambda_1(x)}{\omega_1} \wedge 1\right) \wedge \left(\frac{\lambda_2(x)}{\omega_2} \wedge 1\right) \wedge \left(\frac{\lambda_3(x)}{\omega_3} \wedge 1\right) \wedge \left(\frac{\lambda_4(x)}{\omega_4} \wedge 1\right)$$

Subject to :

$$\lambda_t(x) = \mu_{f_t}(x) \qquad t = 1,\dots,4$$

$$0.5 \le r_i \le 1 - 10^{-6} \qquad r_i \in [0,1]$$

$$1 \le n_i \le 10 \qquad n_i \in Z^+$$

$$0.5 \le R_s \le 1 - 10^{-6} \qquad R_s \in [0,1]$$

The weight set $\omega = \{\omega_1, \omega_2, \omega_3, \text{and } \omega_4\}$ was selected in the range [0.2,1], where the values of the weights indicate the bias towards specific objectives as specified by the expert decision maker. In particular, the weight set $\omega = [1,1,1,1]$ implies that the expert user prefers that there should be no bias towards any objective goal, that is, there is no preference at all. Every other combination of weights implies that there is some bias towards one or more specific objectives, and the relative importance of objectives is ranked accordingly. For instance, with a weight set defined by $\omega = [1,0.5,0.5,0.5]$, the preference is biased towards the region that is closer to the objective corresponding to reliability than to the rest of the objectives that are equally ranked with weight value of 0.5. Therefore, the decision making process takes into account the decision maker's preferences and choices based on expert opinion. In addition, the FMGA approach is a useful decision support tool that can provide a set of good solutions in an interactive manner, rather than prescribe a single solution. Furthermore, the approach enables the decision maker to specify the minimum and maximum values of objective functions in terms of reliability, cost, volume, and weight, denoted by f_1, f_2, f_3, and f_4, respectively. Table 2 provides a list of the selected minimum and maximum values of the objective functions, for the series. This approach makes the FMGA algorithm more adaptable and flexible for addressing specific problem situations while accommodating the expert user's managerial preferences. Computational results and discussions are presented in the next section.

Table 2. Minimum and maximum feasible values of objective functions

	Series System			
	f_1	f_2	f_3	f_4
M_i	1	180	120	210
m_i	0.6	60	5	100

5. Results and Discussions

This section presents the comparative results of the numerical experiments. The best three FMGA solutions are compared with the results obtained by other algorithms in the literature, for the series, series-parallel and complex bridge systems. We compare our results with those in [19] and [31].

Table 3 shows the comparative numerical results in which the best three solutions of the problem are compared against solutions from the literature. The results indicate that the best three FMGA solutions are better than the solutions reported previously in [19, 31], particularly in terms of system reliability. In terms of cost, the solutions are slightly less than the previously reported solutions; the difference in cost is, however, not significant. Though there are a few exceptional instances where the cost of the FMGA are slightly higher with differences in the order of 10^{-6}, it can be seen that, overall, FMGA provides better solutions than the approaches reported previously. FMGA approach found high quality solutions, most of which are better than those previously recorded in the literature. In summary, the approach offers a number of practical advantages to the decision maker, including the following:

- FMGA addresses the imprecise and fuzzy characteristics of the system reliability optimization problem;
- The method address conflicting multiple objectives, giving a trade-off between the objectives;
- The approach accommodates the decision maker's preferences in its procedure;
- The method gives a population of alternative solutions for the decision maker, rather that prescribe a solution;
- The method is practical, flexible and easily adaptable to specific problem situations.

In view of the above advantages, FMGA is a potentially useful approach that can be ffurther developed into a decision support tool for optimizing practical industrial system reliability situations.

6. Conclusions

In the real world, decision makers concerned with system reliability optimization encounter problems of finding a judicious trade-off between maximizing reliability and minimizing cost to an acceptable degree of satisfaction. In such a fuzzy environment, the management goals and constraints are not known precisely. Moreover, the goals are often conflicting, which further complicates the reliability optimiza-

Table 3. Comparison of best-3 FMGA solutions with other algorithms

	Best 3 FMGA Solutions			Wu et al. [31]	Hsieh et al.[19]
No.	$(r_i: n_i)$	$(r_i: n_i)$	$(r_i: n_i)$		$(r_i: n_i)$
1	(0.779401321:3)	(0.77940279:3)	(0.77939597:3)	(0.78037307:3)	(0.779427:3)
2	(0.871839015:2)	(0.87181554:2)	(0.87183716:2)	(0.87178343:2)	(0.869482:2)
3	(0.902877370:2)	(0.90287257:2)	(0.90288515:2)	(0.90240890:2)	(0.902674:2)
4	(0.711415792:3)	(0.71141514:3)	(0.71140318:3)	(0.71147356:3	(0.714038:3)
5	(0.787779580:3)	(0.78783097:3)	(0.78780147:3)	(0.78738760:3)	(0.786896:3)
R_s	0.931682387	0.931682384	**0.931682388**	0.9316800	0.93157800
C_s	175.0000000	175.0000000	175.0000000	174.99899	174.878546
W_s	192.4810818	192.4810818	192.4810818	192.48108	192.481082
V_s	83.00000000	83.00000000	83.00000000	83.000000	83.0000000

Note: Bold indicates the best FMGA solution

tion problem. One most viable and useful option is to us a fuzzy satisficing approach that includes the preferences and expert judgments of the decision maker. We provided a multi-objective non-linear mixed integer program for addressing system reliability optimization problems. The fuzzy multi-objective model is transformed into a single-objective model which uses a fuzzy evaluation method. Genetic algorithm uses the fuzzy evaluation method to evaluate the fitness of individuals in each population at every generation. Numerical results demonstrate that the fuzzy multi-objective Genetic Algorithm approach is able to provide high quality solutions while accommodating the preferences of the user.

This study offers a useful contribution to decision makers in system reliability design. Contrary to single-objective approaches which seek to optimize system reliability only. FMGA provides a trade-off between management goals. At design stage, the information required for system reliability design is imprecise and incomplete. To that effect, the problem becomes ill-structured such that reliance on expert information is inevitable. Using FMGA, the vagueness and imprecision of the expert knowledge, at the design stage, can be addressed effectively while taking into account the multiple conflicting objectives. Furthermore, FMGA provides a population of good alternative solutions in an interactive manner, giving the decision maker a wide choice of practicable solutions and an opportunity to consider other practical factors that cannot be included in the formulation. Overall, FMGA is a useful platform for decision support for system reliability design when the parameters, the management goals, the design constraints, and the impact of the possible alternative actions are not precisely known.

Acknowledgment: *We would like to appreciate the constructive comments of the anonymous reviewers on the earlier version of this paper.*

References

1. Abbass HA. Speeding up backpropagation using multi-objective evolutionary algorithms. Neural Computation 2003; 15 (11): 2705-2726.
2. Anchor KP, Zydallis JB, Gunsch GH, Lamont GB. 2003. Different multi-objective evolutionary programming approaches for detecting computer network attacks. In Carlos M. Fonseca, Peter J. Fleming, Eckart Zitler, Kalyanmoy Deb and Lothar Thiele (editors), Evolutionary Multi-criterion Optimization. Second International Conference, EMO 2003. Springer. Lecturer Notes in Computer Sceince 2003; 2632: 707-721.
3. Anderson J. Applications of a multi-objective genetic algrorithm to engineering design problems. In Carlos M. Fonseca, Peter J. Fleming, Eckart Zitler, Kalyanmoy Deb and Lothar Thiele (editors), Evolutionary Multi-criterion Optimization. Second International Conference, EMO 2003. Springer. Lecturer Notes in Computer Sceince 2003; 2632: 737-751.
4. Bag S, Chakraborty D, Roy AR. A production inventory model with fuzzy demand and with flexibility and reliability considerations. Journal of Computers and Industrial Engineering 2009; 56: 411-416.
5. Bellman R, Zadeh L. Decision making in a fuzzy environment. Management Science 1970; 17: 141–164.
6. Bing L, Meilin Z, Kai X. A practical engineering method for fuzzy reliability analysis of mechanical structures. Reliability Engineering and System Safety 2000; 67 (3): 311–315.
7. Cai KY, Wen CY, Zhang ML. Fuzzy variables as a basis for a theory of fuzzy reliability in the possibility context. Fuzzy Sets and Systems 1991; 42: 145–172.
8. Chen L. Multiobjective design optimization based on satisfaction metrics. Engineering Optimization, 33: 601–617, 2001.
9. Chen SM, Fuzzy system reliability analysis using fuzzy number arithmetic operations Fuzzy Sets and Systems 1994; 64 (1): 31–38.
10. Chen TC, You PS. Immune algorithm based approach for redundant reliability problems. Computers in Industry 2005; 56: 195–205.
11. Coit DW, Smith AE. Reliability optimization of series-parallel systems using genetic algorithm. IEEE Transactions on Reliability 1996, R-45, (2): 254-260.
12. Delgado M, Herrera F, Verdegay JL, Vila MA. Post optimality analysis on the membership functions of a fuzzy linear problem. Fuzzy Sets and Systems 1993; 53: 289-297.
13. Duque O, Morifiigo D. A fuzzy Markov model including optimization techniques to reduce uncertainty. IEEE Melecon 2004; 3: 841–844.
14. Garg H, Sharma SP. Multi-objective reliability-redundancy allocation problem using particle swarm optimization. Computers & Industrial Engineering 2013; 64: 247-255.

15. Garg H, Sharma SP. Stochastic behavior analysis of industrial systems utilizing uncertain data.5 ISA Transactions, vol. 51, no. 6, pp.752–762, 2012.
16. Goldberg DE. Genetic Algorithms: In Search, Optimization & Machine Learning, Addison-Wesley, Inc., MA, 1989.
17. Hikita M. Nakagawa Y. and Harihisa H. Reliability optimization of systems by a surrogate constraints algorithm. IEEE Transactions on Reliability 1992, R-41 (3): 473–480.
18. Holland JH. Adaptation in Natural and Artificial System, University of Michigan Press, Ann Arbor, MI, 1975.
19. Hsieh YC. Chen TC, Bricker DL. Genetic algorithm for reliability design problems. Microelectronics and Reliability 1998; 38 (10): 1599-605.
20. Huang HZ, Fuzzy multi-objective optimization decision-making of reliability of series system. Microelectronics Reliability 1997; 37 (3): 447–449.
21. Huang HZ, Gu YK, Du X. An interactive fuzzy multi-objective optimization method for engineering design. Engineering Applications of Artificial Intelligence 2006; 19(5): 451–460.
22. Huang HZ, Tian ZG, Zuo MJ. Intelligent interactive multi-objective optimization method and its application to reliability optimization. IIE Transactions 2005, 37 (11): 983–993.
23. Jozefowiez N, Glover F, Laguna M. Multi-objective meta-heuristics for the traveling salesman problem with profits. Journal of Mathematical Modelling and Algorithms 2008; 7 (2): 177-195.
24. Kuo W, Prasad VR. An annotated overview of system-reliability optimization. IEEE Transaction on Reliability 2000; 49 (2): 176–187,.
25. Mahapatra GS, Roy TK. Fuzzy multi-objective mathematical programming on reliability optimization model. Applied Mathematics and Computation, vol. 174, pp.643–659, 2006.
26. Michalewicz Z. Genetic Algorithms + Data Structures = Evolutionary Programs. Springer, 1996.
27. Mohanta DK, Sadhu PK, Chakrabarti R. Fuzzy reliability evaluation of captive power plant maintenance scheduling incorporating uncertain forced outage rate and load representation. Electric Power Systems Research 2004; 72: 73–84.
28. Onisawa T. An application of fuzzy concepts to modeling of reliability analysis. Fuzzy Sets and Systems 1990; 37: 267–286.
29. Sakawa M, Fuzzy Sets and Interactive Multi-objective Optimization, Plenum Press, New York, 1993.
30. Slowinski R. Fuzzy sets in decision analysis. Operations research and statistics. Boston: Kluwer Academic Publishers 1998.
31. Wu P, Gao L, Zou D, Li S. An improved particle swarm optimization algorithm for reliability problems. ISA Transactions 2011; 50: 71-8.

Michael MUTINGI
Faculty of Engineering and the Built Environment
University of Johannesburg
P. O. Box 524, Bunting Road, Auckland Park 2006, South Africa
E-mail: mmutingi@gmail.com

Article citation info:
BRZOZOWSKI K, NOWAKOWSKI J. Model for calculating compression ignition engine performance. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 407–414.

Krzysztof BRZOZOWSKI
Jacek NOWAKOWSKI

MODEL FOR CALCULATING COMPRESSION IGNITION ENGINE PERFORMANCE

MODEL DO WYZNACZANIA PARAMETRÓW PRACY SILNIKA O ZAPŁONIE SAMOCZYNNYM

Optimising the performance of an internal combustion engine requires both empirical and theoretical work. The latter involves reasoning based on results yielded by mathematical models. This paper presents a computationally efficient model of the working cycle for a compression ignition engine. The model enables analysis of the working cycle of an engine with an electronically controlled common-rail type power supply and a controlled exhaust gas recirculation system. The model's parameters are chosen in a two-stage identification process based on the results of the experiments. The first stage of identifying the parameters requires formulating and solving an appropriate dynamic optimisation problem for multiple discrete points describing the engine's operation. To this end a genetic algorithm is used with an additional condition controlling the quality of the solution. Artificial neural networks are used for the second stage of identification. The paper shows an example of using the model to assess the influence of the kinetic combustion phase, which results from the way in which the injection process proceeds on the course of the working cycle. The accuracy of calculations with respect to basic parameters characterising the working cycle is also discussed.

Keywords: *compression ignition engine, identification, engine performance parameters.*

Doskonalenie parametrów pracy silnika spalinowego poprzez odpowiednie sterowanie cyklem roboczym wymaga stosowania zarówno prac o charakterze doświadczalnym jak i obliczeniowym. W tym drugim przypadku podstawą wnioskowania są wyniki uzyskiwane z modeli matematycznych. Artykuł przedstawia efektywny obliczeniowo model cyklu roboczego silnika o zapłonie samoczynnym. Model umożliwia analizę cyklu roboczego silnika z elektronicznie sterowanym układem zasilania typu common-rail oraz układem sterowanej recyrkulacji spalin. Parametry modelu dobrano w dwuetapowym procesie identyfikacji bazującym na wynikach badań stanowiskowych. Pierwszy etap identyfikacji parametrów wymagał sformułowania i rozwiązania odpowiedniego zadania optymalizacji dynamicznej dla wielu dyskretnych punktów pracy silnika. W tym celu zastosowano algorytm genetyczny z dodatkowym warunkiem kontroli jakości rozwiązania. W drugim etapie identyfikacji do uogólnienia wyników wykorzystano sztuczne sieci neuronowe. W pracy przedstawiono przykład zastosowania modelu w ocenie udziału fazy spalania kinetycznego wynikającej z realizacji przebiegu procesu wtrysku na przebieg cyklu roboczego oraz przedstawiono dokładność obliczeń w odniesieniu do podstawowych parametrów charakteryzujących cykl roboczy.

Słowa kluczowe: *silnik o zapłonie samoczynnym, identyfikacja, parametry pracy silnika.*

1. Introduction

Empirical models of real phenomena are commonly used in practical applications. Their usefulness depends on knowing a number of parameters, the so-called model parameters. Using models of this type results from the requirement that the time complexity of numeric computations performed must be acceptable. The following papers from recent years present models of this type [3, 12, 14–16], most of which involve the assumption that the subject of study are the mean values of pressure and temperature in the entire combustion chamber. Such an assumption leads to formulating zero-dimensional models characterised by adequate numeric efficiency due to the duration of computations being the key parameter in determining the potential usefulness of a model in control tasks. The main problem in zero-dimensional empirical models is choosing the function describing the dynamics of the combustion process of an injected dose of fuel. This function can be expressed by way of elaborating the results of direct measurements of pressure in the cylinder, as in [3, 12]. A more common approach, however, is to choose one of the traditionally used functions describing the combustion process, whose parameters are determined in such way that the results of the experiments are matched relatively as closely as possible. Thor et al. [18] is one of several papers to use this function. Also, the paper [15] employs a composition of two empirical functions.

The parameters of zero-dimensional models are functions of a given excitation, and choosing appropriate values for them is the key problem concerning the accuracy of calculation. The problem of determining parameters for empirical models of an engine's working cycle is discussed in [11]. The task may be formulated as a dynamic optimisation problem, as in the current paper which proposes and applies an evolutionary algorithm to choose the values of the model parameters for the working cycle of a compression ignition engine. Evolutionary algorithms are a group of the so-called artificial intelligence methods applicable to problems defying efficient algorithmisation, and they complement classical approaches to computation. They are notable for their applications to problems concerning working cycles and control of internal combustion engines, as exemplified by [1, 6, 7, 10, 13, 19]**.** Genetic algorithms, which form a popular subset of evolutionary algorithms, are used in those optimisation problems which lend themselves easily to specialised methods and whose search space is too large for classical algorithms. The current paper also uses a genetic algorithm to choose the values of model parameters. It is based on the results of actual measurements taken on a test bench.

2. Model of the working cycle and the scope of experimental measurements

For a zero-dimensional model of an engine's working cycle the phenomena occurring in the engine's cylinder are described by ordinary non-linear differential equations obtained from the principle of mass conservation and energy balance [4]. By putting "d" and "w" into the lower index for the values characterising the inlet and exhaust system, respectively, the considered mathematical model of the working cycle of a compression ignition engine may be presented in the following form [10]:

$$\frac{dm}{d\varphi} = \frac{dm_d}{d\varphi} - \frac{dm_w}{d\varphi} + \frac{dm_B}{d\varphi}, \tag{1}$$

$$\begin{aligned} B_0 W_0 \frac{dx_B}{d\varphi} + \frac{30}{\pi n} h_c A_c (T_{sc} - T) + c_{pd} T_d \frac{dm_d}{d\varphi} &= \\ c_v T \frac{dm}{d\varphi} + c_v m \frac{dT}{d\varphi} + p \frac{dV}{d\varphi} + c_{pw} T_w \frac{dm_w}{d\varphi} \end{aligned}, \tag{2}$$

$$\frac{dm_d}{d\varphi} = f\left(n, \mu_d p_d, p, T_d\right), \tag{3}$$

$$\frac{dm_w}{d\varphi} = f\left(n, \mu_w p_w, p, T\right), \tag{4}$$

$$\begin{aligned} \frac{dx_B}{d\varphi} = \beta \left\{ -e_2 \left[1 - \left(\frac{\varphi - \varphi_z}{\Delta\varphi_s} \right)^{e_1 \tau} \right]^{e_2 - 1} \cdot (-e_1 \tau) \left(\frac{\varphi - \varphi_z}{\Delta\varphi_s} \right)^{e_1 \tau - 1} \right\} + \\ + (1 - \beta) \left\{ -\exp \left[-e_3 \lambda \left(\frac{\varphi - \varphi_z}{\Delta\varphi_s} \right)^{e_4} \right] \cdot (-e_3 \lambda) e_4 \left(\frac{\varphi - \varphi_z}{\Delta\varphi_s} \right)^{e_4 - 1} \right\} \end{aligned}. \tag{5}$$

The above equations are complemented by algebraic relations resulting from the assumptions that the working medium is a semi-perfect gas and that the heat transfer coefficient is a function of state parameters and the engine's structural parameters [5]:

$$h_c = e_5 V^{e_6} p^{e_7} T^{e_8} \left(c_m + 1.4\right)^{0.8}, \tag{6}$$

where c_m is the mean piston speed.

A further discussion assumes that the vector of the model's independent input parameters is the following vector of the engine's control parameters:

$$\boldsymbol{X} = \left[n, B_0, \varphi_w, X_{EGR} \right]^T, \tag{7}$$

where X_{EGR} is the degree of exhaust gas recirculation and φ_w is the injection advance angle.

The input values, which are dependent on the control parameters vector $\boldsymbol{X}$, are the pressure and the temperature of the medium in the inlet system and the excess air coefficient. These values form the auxiliary parameters vector as follows:

$$\boldsymbol{G} = \left[p_d, T_d, p_w, T_w, \lambda \right]^T, \tag{8}$$

where λ is the relative air/fuel ratio.

Since the formulated model uses a few empirical relations and coefficients, a proper choice of their values depending on a given engine's design constitutes an important step. These coefficients are thus called the model parameters and they form the model parameters vector of the form:

$$\boldsymbol{E} = \left[\mu_d, \mu_w, \Delta\phi_s, \phi_z, \beta, e_1 \ldots e_8 \right]^T. \tag{9}$$

A practical application of the formulated model of the engine's working cycle requires knowing the functions describing the dependency of the model parameters on the values of the control parameters for all technically feasible states of engine operation. Determining those functions is an identification problem and proceeds in two stages.

The first stage consists in identifying the values of individual components of the model parameters vector for discrete states of engine operation. Suitable experiments on a test bench are therefore required. Planning the agenda for experimental measurements should involve choosing states of engine operation as discrete points so that performing registration of the engine's operation parameters and those of the inlet and exhaust system encompasses a possibly wide range of actual operation states. Following the choice of loads in the ESC test [2], partial values of the maximal moment of force for each chosen value of angular velocity were chosen as the values of moment constituting the engine's load. Based on previous work [10, 17], discrete points describing the engine's operation were determined for four fixed values of angular velocity. For each of the angular velocities, experiments were performed with variable load, whereby five different values of the braking moment were used in each case. The load was changed in an increasing manner and different degrees of exhaust gas recirculation, starting with none, were tested for a given point at which measurements were taken. The degree of exhaust gas recirculation was subsequently increased by 5% until reaching its maximum defined for a given point. The definition of these maxima set them as high as possible while the engine could sustain the assumed moment of inertia. The exhaust gas recirculation system was inactive for maximal loads. The dose supplied during a single experiment to the operating engine consisted of two parts: a constant pilot dose and a variable main dose. For each point measurements were taken with a standard injection advance angle (preset by the built-in microcontroller controlling the engine's operation) and with angles advanced and delayed by 2°CA and 4°CA from the standard value, respectively. In the entire range of intermediate loads engine operation was regulated so that the assumed constant angular speed and constant moment of inertia were maintained. Such regulation enabled analysis of parameters of engine operation at constant mean effective pressure. Increases (or decreases) of moment of inertia were compensated by changing the doses of fuel. Fig. 1 schematically presents the proposed way of choosing the discrete points describing the engine's operation for which to carry out the experimental measurements.

The varying pressure in the cylinder was registered along with other characteristic parameters of the working cycle for each point describing the engine's operation. Experimental measurements allowed to collect data describing the parameters of the inlet and exhaust system's operation and parameters of the working cycle for $n_i = 400$ different sets of control parameters, whereby a presently produced four-cylinder compression ignition engine suitable for propelling an automobile was used as the subject of study. Characteristic technical data of the engine used in measurements and subsequent identification of model parameters are presented in Table 1.

Subsequently, the second stage of the identification problem of model parameters involves the task of generalising the results of discrete identification in order to obtain a relationship of the form:

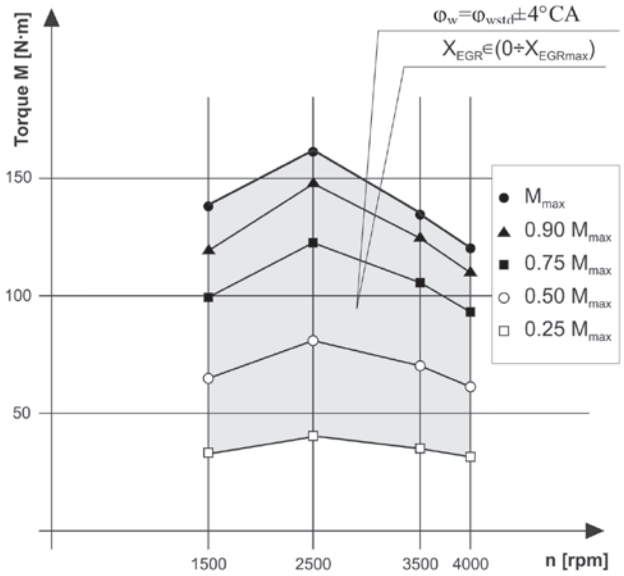


Fig. 1. Scheme of the scope of experimental measurements used in the first stage of model parameters identification

Table 1. The engine's technical data

Engine	Compression ignition engine supercharged by a turbo compressor with direct injection equipped with an electronically controlled Common Rail system
Layout of cylinders	4 in line
Number of valves per cylinder	4
Bore	69.6 mm
Stroke	82 mm
Total displacement	1248 cm^3
Compression ratio	16.8
Maximum power	55.2 kW / 4000 rpm
Maximum torque	190 N·m / 1500 rpm

$$\boldsymbol{E} = f_E(\boldsymbol{X}), \quad (10)$$

and

$$\boldsymbol{G} = f_G(\boldsymbol{X}). \quad (11)$$

The proposed method is discussed in the next chapter.

3. Determining the model parameters

The first stage of identifying the model parameters is performed for $i = 1,..,n_i$ discrete states of the engine's operation and consists in determining the values of the components of vector $\boldsymbol{E}$. The task may be considered as a minimisation problem for the following functional:

$$\Omega(\boldsymbol{X},\boldsymbol{E},\boldsymbol{G}) = c_1 \int_0^{4\pi} \left[p_E(\varphi) - p_F(\varphi) \right]^2 d\varphi + c_2 \left(\max p_E(\varphi) - \max p_F(\varphi) \right)^2 \to \min, \quad (12)$$

where: $p_{\mathbf{F}}(\varphi)$ – pressure calculated according to the model of the working cycle

$p_{\mathbf{E}}(\varphi)$ – smoothed pressure from bench testing:

$$p_{\mathbf{E}} = \frac{a_0}{2} + \sum_{j=1}^{m_j} a_j \cos\frac{j\varphi}{2} + b_j \sin\frac{j\varphi}{2},$$

$$m_j = 60,$$

$$a_j = \frac{1}{2\pi} \int_0^{4\pi} p(\varphi) \cos\frac{j\varphi}{2} d\varphi, \quad j = 0,1,..,n,$$

$$b_j = \frac{1}{2\pi} \int_0^{4\pi} p(\varphi) \sin\frac{j\varphi}{2} d\varphi, \quad j = 1,..,n,$$

c_1, c_2 – constant weight coefficients.

Solving a problem thus formulated requires knowledge of the actual pressure varying in the cylinder for each set of control parameters $\mathbf{X}^{(i)}$ and auxiliary parameters $\mathbf{G}^{(i)}$, where $i = 1,..,n_i$ denotes the discrete states of an engine's operation. Calculating $p_{\mathbf{F}}(\varphi)$, on the other hand, necessarily involves integration of the model's equations (1÷5) for every possible choice of the model parameters. The task is solved by using a genetic algorithm, thereby ensuring a concurrent search for the solution set by assuming the following fitness function:

$$\Phi(\boldsymbol{X},\boldsymbol{E},\boldsymbol{G}) = \frac{1}{\Omega(\boldsymbol{X},\boldsymbol{E},\boldsymbol{G})} \to \max. \quad (13)$$

A genetic algorithm with real encoding is used, thus making each individual's chromosome into:

$$\mathbf{Z} = [z_1,...,z_{13}]^T, \quad (14)$$

where $z_1 = \mu_d, z_2 = \mu_w, z_3 = \Delta\varphi_s, z_4 = \varphi_z, z_5 = \beta, z_{5+i} = e_i$ *for* $i = 1..8$.

For each of the genes z_i, the range of admissible values are defined, respectively, as $z_{i,\min}$ and $z_{i,\max}$, which are presented in Table 2.

The initial values are generated with a random strategy and the selection operation is done with the direct tournament selection method for pairs of chromosomes. The arithmetical crossover and the non-uniform mutation [8] were used as genetic operators. The set of chromosomes used in the next step of the algorithm is determined by ranking the fittest individuals. Since the number of individuals used ($n_p = 24$) is relatively small, the possibility of not obtaining a fully satisfying solution to an individual identification problem of the form (12) within the limited number of the algorithm's iterations ($k_{\max} = 15$) is assumed. This assumption leads to the introduction of the following additional criterion whose failure requires repeating the identification procedure:

Table 2. Range of admissible values for particular genes of chromosome ***z***

z_i	z_1	z_2	z_3	z_4	z_5	$z_6 \div z_8$	z_9	z_{10}	z_{11}	z_{12}	z_{13}
$z_{i,\min}$	0.1	0.1	15°	343°	0	0	0.5	10^{-3}	-0.1	0	-1
$z_{i,\max}$	0.4	0.95	36°	375°	0.3	5	5	5	0	1	0

$$\frac{\left|p_{i(r)E} - p_{i(r)F}\right|}{p_{i(r)E}} \leq 5\% . \tag{15}$$

This means that the relative difference between the mean indicated pressure registered experimentally $p_{i(r)E}$ and the computed value of the mean indicated pressure $p_{i(r)F}$ for the fittest chromosome in the expansion phase of the cycle must be below 5%.

It turns out that using the criterion for restarting the identification procedure enables to obtain satisfying solutions to the whole set of analysed states of the engine's operation. The average number of necessary repetitions of the identification procedure for $n_i = 400$ identification problems solved consecutively is 4.5.

In the second stage of identifying a suitable approximation problem for the results obtained for discrete identification is formulated and solved in order to determine the appropriate functions $f_E(\mathbf{X})$ and $f_G(\mathbf{X})$. In the case of parameters needed to compute the heat transfer coefficient $e_5 \div e_8$, as their values are very close in discrete identification, a simplification is made by taking their mean. Other model parameters may be approximated using artificial neural networks. A series of numeric experiments allowed to conclude that the problem of determining $f_E(\mathbf{X})$ and $f_G(\mathbf{X})$ requires using two separate feed-forward multilayer neural networks (Fig. 2):

network I:

$$\boldsymbol{X} = \left[n, B_0, \varphi_w, X_{EGR}\right]^T \rightarrow Out_1 = \left[p_d, T_d, p_w, T_w, \lambda, \mu_d, \mu_w, \Delta\varphi_s, \varphi_z\right]^T, \tag{16}$$

network II:

$$\boldsymbol{X} = \left[n, B_0, \varphi_w, X_{EGR}\right]^T \rightarrow Out_2 = \left[\beta, e_1, e_2, e_3, e_4\right]^T . \tag{17}$$

As a result of the learning process, the architectures of the networks are set to be 15:9 (n_1:n_2) for network I and 31:5 for the network II reflecting the parameters of the Watson function. Both networks use a unipolar neuron activation function.

4. Computing the cycle's characteristic parameters

The formulated model enables assessment of the characteristic parameters of the working cycle for arbitrary states of the engine, i.e. for an arbitrary technically feasible vector of the control parameters. It therefore lends itself to an analysis of the ability to shape the working cycle in a real engine with an electronically controlled common-rail type power supply system and a controlled exhaust gas recirculation system. One application of such a model is assessing the influence of contribution of the kinetic combustion phase, resulting from the way in which the injection process proceeds, on the engine's working cycle. The analysis encompasses the computed varying pressures and temperatures of the medium in the cylinder and the changes of the characteristic parameters of the working cycle, such as mean indicated pressure p_i, thermal efficiency η_c and maximal temperature of the medium $T_{\max}$.

The course of heat emission, and hence also the contribution of the kinetic and diffusion combustion phase, may be adjusted by dividing the dose injected into the cylinder into individual partial doses depending on the load and the engine's angular velocity. Of particular importance are the choice of the pilot dose's value, the time interval between the pilot dose and the main dose, and the injection advance angle of the main dose. Figure 3 illustrates the influence of reducing the contribution of the kinetic combustion phase on the parameters of the working cycle at a selected point describing the engine's operation for set values of the control parameters.

Analogously, the model enables an assessment of the influence of the degree of recirculation on the working cycle. The value of the maximal degree of exhaust gas recirculation depends on the engine's boost pressure and design of the recirculation systems. It also follows from a compromise between the ability to reduce emission of nitrogen oxides, emission of solid particles (thus also smoke opacity of the exhaust gases) and decline in the engine's capabilities. An illustration of the influence of the degree of recirculation on the parameters of the working cycle at a selected point describing the engine's operation for set values of control parameters is shown in Fig. 4.

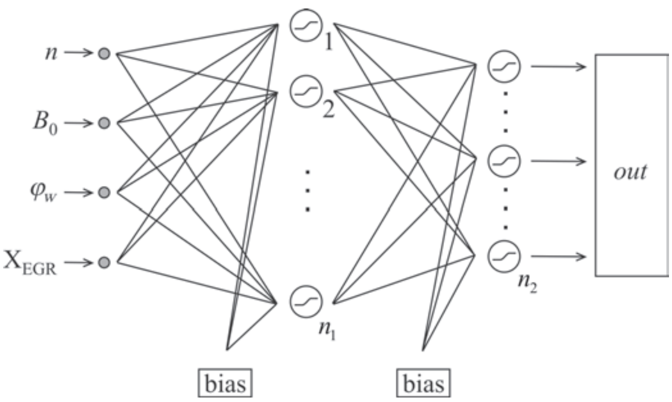


Fig. 2. Schematic layout of the artificial neural networks used in the work

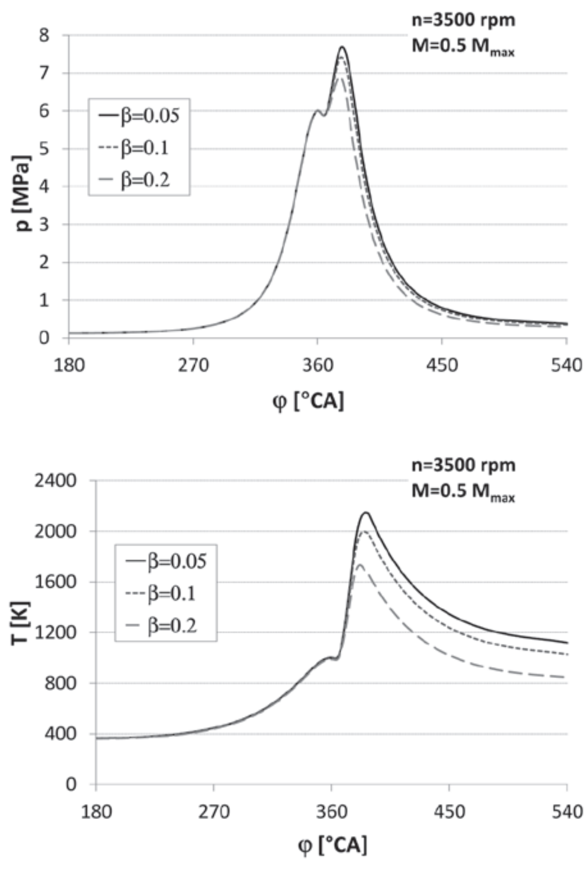


Fig. 3. Influence of the kinetic combustion phase's contribution on the working cycle for 3500 rpm

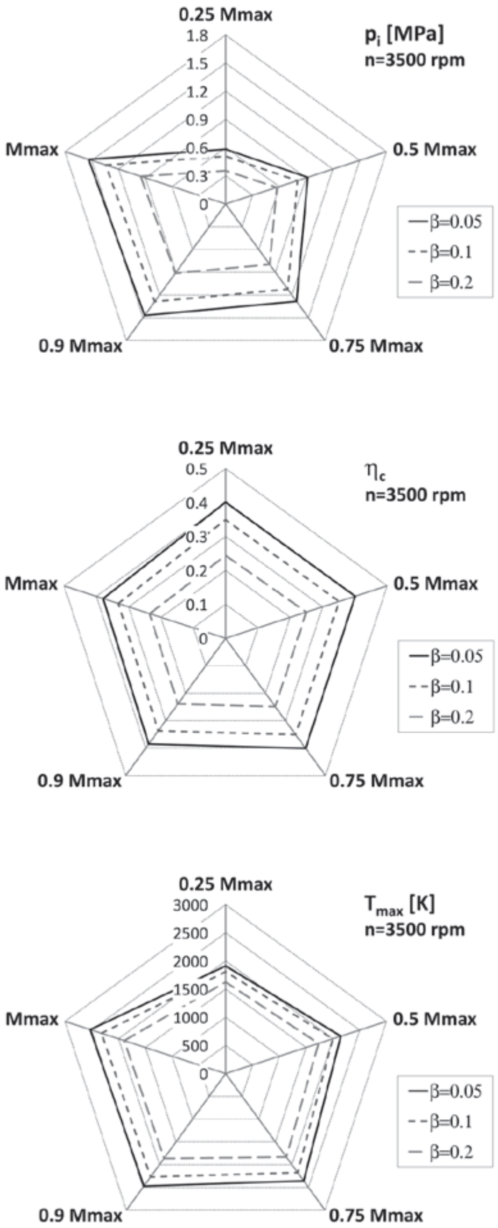


Fig. 3. (continued) Influence of the kinetic combustion phase's contribution on the working cycle for 3500 rpm

5. Conclusions

The model presented in the paper enables to perform a number of computations related to assessing the working cycle of an engine. Error of the calculation can be determined by comparing the computed and experimentally registered varying pressures in the cylinder for selected points describing the engine's operation. Sample comparisons of varying pressures in the cylinder at selected points describing the engine's operation (constituting elements of the set used to verify the quality of approx imation of model parameters by applying artificial neural networks) are shown in Fig. 5.

Accuracy of the model can be determined by comparing the computed and experimentally registered values of parameters such as: mean indicated pressure p_i, mean indicated pressure in the expansion phase of the cycle $p_{i(r)}$, maximal pressure in the cycle $p_{\max}$, and mass of the medium in the cylinder m. In Fig. 6 the values of the listed characteristic parameters of the working cycle obtained from the ex-

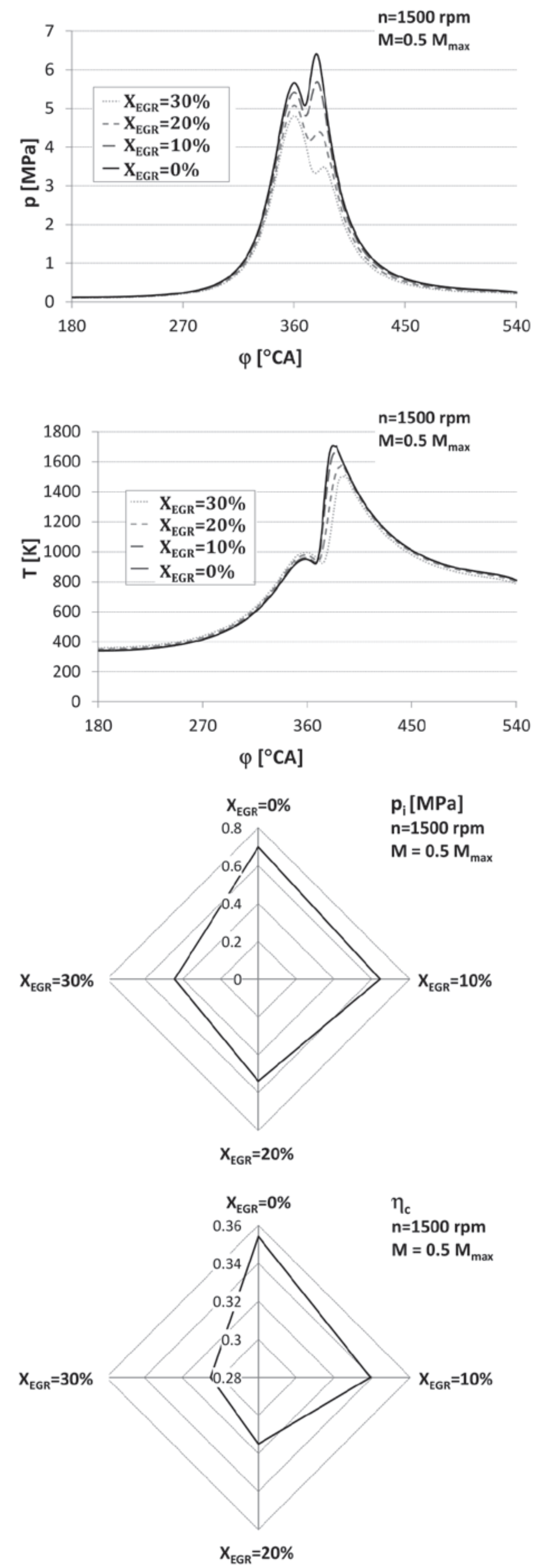


Fig. 4. Influence of the degree of recirculation on the working cycle for 1500 rpm and 50% of maximum load

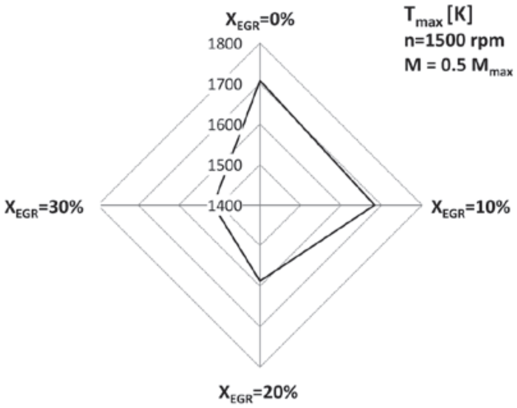


Fig. 4. (continued) Influence of the degree of recirculation on the working cycle for 1500 rpm and 50% of maximum load

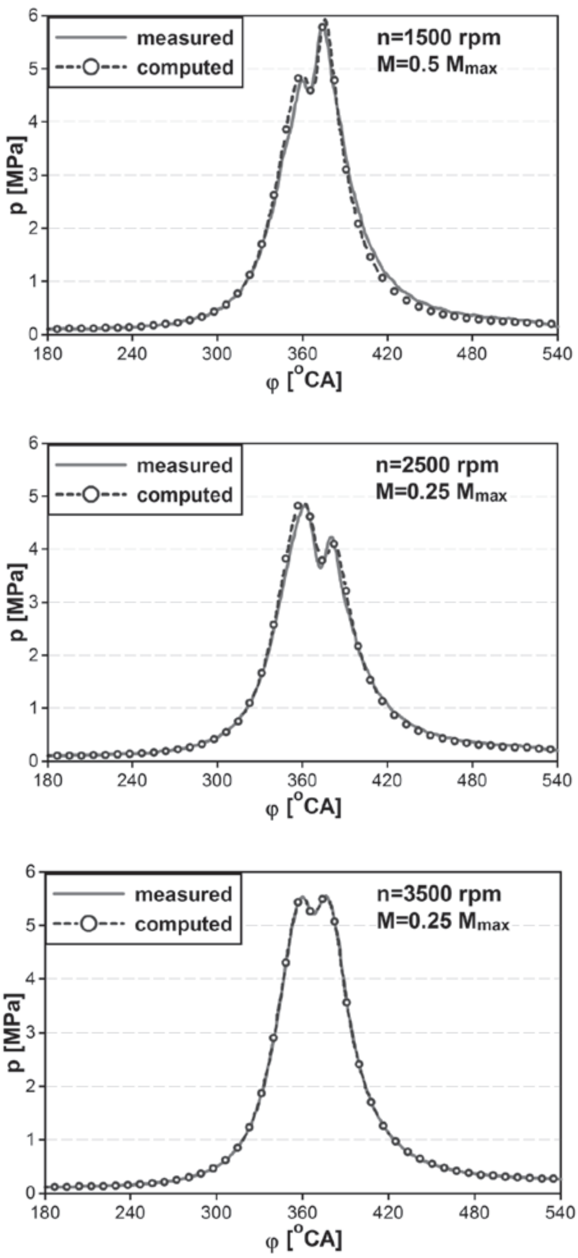


Fig. 5. Comparison of experimentally registered and computed varying pressures for selected states of engine operation

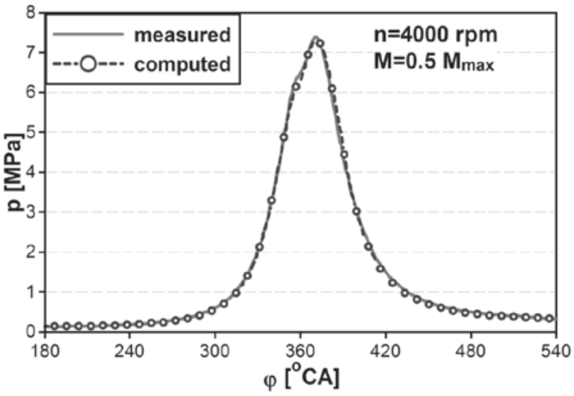


Fig. 5. (continued) Comparison of experimentally registered and computed varying pressures for selected states of engine operation

perimental measurements are compared to those computed according to the model. Values of mean relative error defined as the difference between the measured and computed value related to the measured value are collected for individual quantities in Table 3.

Based on the obtained values of mean relative error and on an interpretation of the comparisons of values of individual characteristic parameters of the working cycle, the conclusion follows that actual varying pressures are accurately represented by the functions computed by this model. This means that the model allows to forecast varying pressure and to determine quantities which characterise the working cycle for a given point describing the engine's operation with acceptable error. Maximal relative errors are in all cases below 15% for every characteristic parameter of the working cycle analysed, whereas the mean errors presented in Table 3 for the three of four characteristic parameters considered do not exceed 3%.

Table 3. Mean relative error of computation of individual characteristic parameters of the working cycle

Parameter	p_i	$p_{i(r)}$	$p_{\max}$	m
Value of mean relative error [%]	5.02	2.74	2.7	2.18
Coefficient of determination R^2	0.989	0.992	0.988	0.971

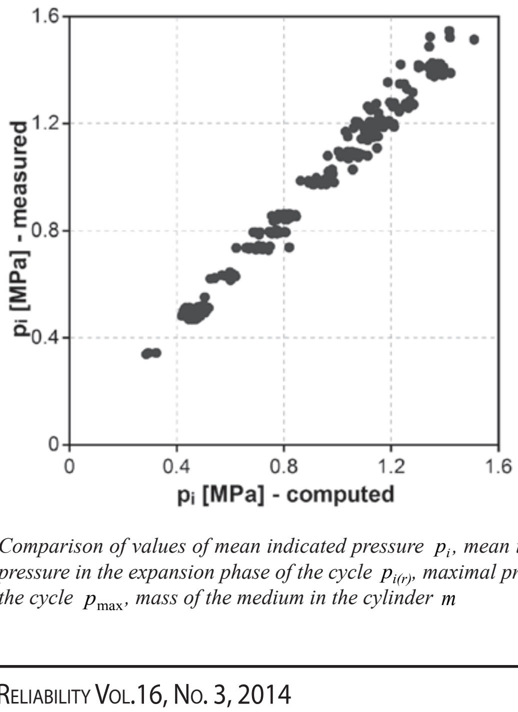


Fig. 6. Comparison of values of mean indicated pressure p_i, mean indicated pressure in the expansion phase of the cycle $p_{i(r)}$, maximal pressure in the cycle $p_{\max}$, mass of the medium in the cylinder m

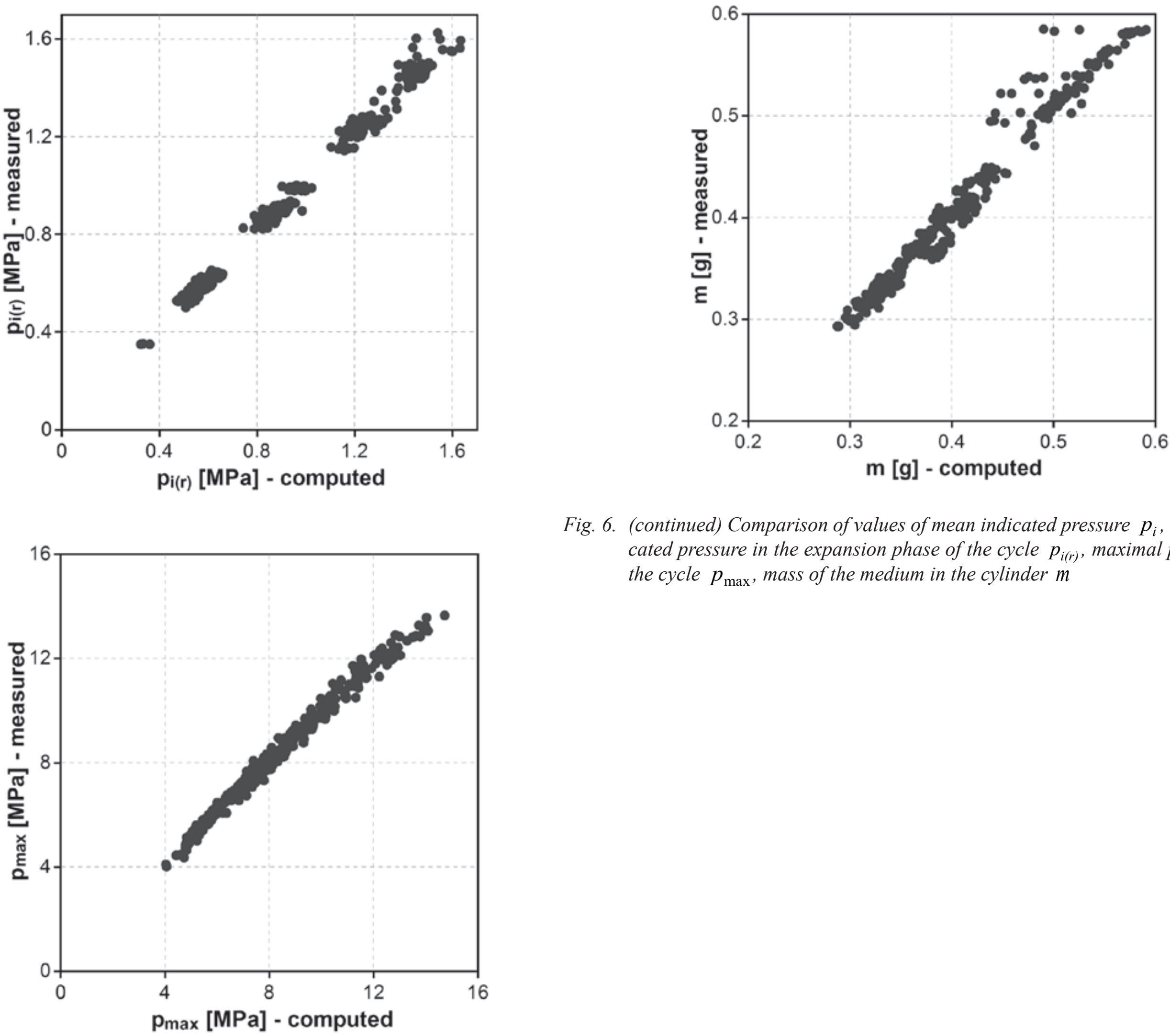


Fig. 6. (continued) Comparison of values of mean indicated pressure p_i, mean indicated pressure in the expansion phase of the cycle $p_{i(r)}$, maximal pressure in the cycle $p_{\max}$, mass of the medium in the cylinder m

References

1. Atashkari K, Nariman-Zadeh N, Gölcü M, Khalkhali A, Jamali A. Modelling and multi-objective optimization of a variable valve-timing spark-ignition engine using polynomial neural networks and evolutionary algorithms. Energy Conversion and Management 2007; 48: 1029-1041.
2. Directive 1999/96/EC of the European Parliament and of the Council of 13 December 1999. Official Journal of the European Communities 1999.
3. Gambarotta A, Lucchetti lng G. Control-oriented crank-angle based modeling of automotive engines. SAE Technical Paper 2011; 2011-24-0144.
4. Heywood JB. International combustion engine fundamentals. New York: Mc-Graw-Hill, 1988.
5. Hohenberg G. Advanced approaches for heat transfer calculations. SAE Technical Paper 1979; 790825.
6. Kesgin U. Genetic algorithm and artificial neural network for engine optimization of efficiency and NOx emission. Fuel 2004; 83: 885-895.
7. Ma F, Wang J, Wang Y, Wang Y, Zhong Z, Ding S. An investigation of optimum control of a spark ignition engine fueled by NG and hydrogen mixtures. International Journal of Hydrogen Energy 2008; 33: 7592-7606.
8. Michalewicz Z. Genetic Algorithms + Data Structures = Evolution Programs. 3rd ed. London: Springer, 1996.
9. Nowakowski J. Model of a CI engine's working cycle and its application to choice of control parameters. Bielsko-Biala: University Press (Dissertation in Polish), 2006.
10. Nowakowski J, Brzozowski K. Mathematical model for simulation of a working cycle of compression-ignition engine based on experimental measurements. In proceedings of FISITA 2010 World Automotive Congress 2010; 229: 1-8.
11. Pirker G, Chmela F, Wimmer A. Automated parameter determination for IC engine simulation models. SAE International Journal of Engines 2009; 2(1): 539-547.
12. Ponti F, Corti E, Serra G, De Cesare M. Common rail multi-jet diesel engine combustion model development for control purposes. SAE Technical Paper 2007; 2007-01-0383.
13. Radziszewski L, Kekez M. Application of a genetic-fuzzy system to diesel engine pressure modeling. The International Journal of Advanced Manufacturing Technology 2010; 46: 1-9.
14. Rezaei R, Eckert P, Seebode J, Behnk K. Zero-dimensional modeling of combustion and heat release rate in DI diesel engines. SAE International Journal of Engines 2012; 5(3): 874-885.

15. Saad C, Maroteaux F, Millet J, Aubertin F. Combustion modeling of a direct injection diesel engine using double Wiebe functions: application to HiL real-time simulations. SAE Technical Paper 2011; 2011-24-0143.
16. Savva N, Hountalas D. Detailed evaluation of a new semi-empirical multi-zone NOx model by application on various diesel engine configurations. SAE Technical Paper 2012; 2012-01-1156.
17. Sobieszczański M, Brzozowski K, Nowakowski J, Tekielak M. The impact of SI engine controlling parameters on the content of toxic components and smokiness of exhaust gases. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2008; 3: 53-61.
18. Thor M, Andersson I, McKelvey T. Parameterized diesel engine heat release modeling for combustion phasing analysis. SAE Technical Paper 2009; 2009-01-0368.
19. Togun NK, Baysec S. Genetic programming approach to predict torque and brake specific fuel consumption of a gasoline engine. Applied Energy 2010; 87: 3401-3408.

Nomenclature

A_c	heat transfer surface area [m^2],
B_0	injected fuel mass [kg],
c_p	specific heat of the medium at constant pressure [kJ/ (kg·K)],
c_v	specific heat of the medium at constant volume [kJ/ (kg·K)],
c_m	mean piston speed [m/s],
c_1 , c_2	constant weight coefficients,
e_i	model coefficients,
h_c	heat transfer coefficient [W/(m^2·K)],
m	mass of the medium [kg],
m_B	mass of fuel [kg],
n	crankshaft rotational speed [rpm],
p	pressure [Pa],
p_i	mean indicated pressure [Pa],
$p_{i(r)}$	mean indicated pressure in the expansion phase of the cycle [Pa],
T	temperature [K],
T_{sc}	wall temperature [K],
V	volume [m^3],
W_0	fuel caloric value [kJ/kg],
x_B	fuel mass burning rate,
X_{EGR}	degree of exhaust gas recirculation,
z_i	genes of chromosome,

Greek letters

β	proportionality factor,
η_c	thermal efficiency,
λ	relative air/fuel ratio,
μ	valve discharge coefficient,
τ	ignition delay time [s],
φ	crank angle [°CA],
φ_w	injection advance angle [°CA],
φ_z	start of combustion [°CA],
$\Delta\varphi_s$	total combustion duration [°CA],

Subscripts

d	inlet system,
w	exhaust system.

Krzysztof BRZOZOWSKI
University of Bielsko-Biala
Department of Transport and Computer Science
ul. Willowa 43-309 Bielsko-Biała, Poland

Jacek NOWAKOWSKI
University of Bielsko-Biala
Department of Combustion Engines and Vehicles
ul. Willowa 43-309 Bielsko-Biała, Poland
E-mails: kbrzozowski@ath.eu, jnowakow@ath.eu

Article citation info:
VAIČIŪNAS G, BUREIKA G. Approach modelling of constant interfailure process of renewal multi-unit fleet. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 415–421.

Gediminas VAIČIŪNAS
Gintautas BUREIKA

APPROACH MODELLING OF CONSTANT INTERFAILURE PROCESS OF RENEWAL MULTI-UNIT FLEET

PODEJŚCIE DO MODELOWANIA STAŁEGO PROCESU MIĘDZYAWARYJNEGO ODNOWY WIELOELEMENTOWEJ FLOTY PASAŻERSKIEJ

While railway companies operate rolling-stock, a substantial part of its expenses goes to maintenance and repair. However, the amount of repair works is directly proportional to the average age of a rolling-stock fleet or its reliability. When renewal an existing fleet of a few dozen rolling-stocks, the installation of the new vehicles reduces the overall failure amount of the fleet proportionally to the number of acquired vehicles. This article provides a concept for creating the model of a passenger rolling-stock's failure intensity according to the mileage. According to this model, a vehicle fleet renewal algorithm can be created and used in order to limit the fluctuation of the fleet's average failure intensity as much as possible and to achieve the most accurate correlation between the number of failures and the fleet's average mileage. Thus a railway company has an opportunity to avoid the unplanned expenses for repairing the vehicles during the unforeseen failure peaks. The SPLINE method is proposed in order to indicate the vehicle failure flow's dependency on the vehicle mileage. After using this method to indicate the variation of the fleet's constant interfailure according to the mileage, the fleet's failure intensity can be modelled according to the algorithm of installing the acquired vehicles for operation.

Keywords: *rolling-stock maintenance, rolling-stock reliability, rolling-stock interfailure process, failure intensity, SPLINE method.*

Znaczna część wydatków ponoszonych przez przedsiębiorstwo kolejowe z tytułu eksploatacji taboru kolejowego, to wydatki na konserwację i naprawy. Jednakże, należy pamiętać, że liczba prac remontowych jest wprost proporcjonalna do średniego wieku floty taboru lub jej niezawodności. Przy odnawianiu istniejącej floty kilkudziesięciu pojazdów kolejowych, wprowadzenie nowych pojazdów zmniejsza ogólną ilość awarii floty proporcjonalnie do liczby nowo nabytych pojazdów. W artykule przedstawiono koncepcję stworzenia modelu intensywności uszkodzeń taboru pasażerskiego w funkcji przebiegu kilometrowego. Zgodnie z tym modelem, można stworzyć algorytm odnowy floty pojazdów, z wykorzystaniem którego można maksymalnie ograniczyć wahania średniej intensywności uszkodzeń floty oraz osiągnąć najbardziej dokładną korelację między liczbą uszkodzeń a średnim przebiegiem pojazdów floty. W ten sposób przedsiębiorstwo kolejowe ma szansę uniknąć nieplanowanych wydatków na naprawy pojazdów podczas nieprzewidzianych okresów wzmożonej awaryjności. Zaproponowano metodę SPLINE, za pomocą której można określić zależność awaryjności pojazdu od jego przebiegu kilometrowego. Zastosowanie tej metody pozwala ustalić zmiany w stałym procesie międzyawaryjnym zależne od przebiegu, co z kolei pozwala na modelowanie intensywności uszkodzeń floty według algorytmu wprowadzania nowo nabytych pojazdów do eksploatacji.

Słowa kluczowe: *konserwacja taboru kolejowego, niezawodność taboru, proces międzyawaryjny w taborze, intensywność uszkodzeń, metoda funkcji sklejanej SPLINE.*

Introduction

Railways are made up of huge complex of mechanical and electrical systems, which consist of thousands moving parts. If a railway service is to be reliable, the equipment must be kept in good working order and regular maintenance (repair) is the essential ingredient to achieve this. Rolling-stocks are the most intensive exploited segment of the railway system and they are the most vulnerable if maintenance is neglected. A stalled train will block a railway line immediately and will reduce a timetable on an intensively used system to an uncontrollable shambles for the remainder of the day. Reliability is the key to successful railway operation and maintenance should be the number one priority to ensure reliability is on-going. Lithuanian and many scientists worldwide carry out comprehensive research works to ensure the both technically and economical effective maintenance system of rolling-stock in railway companies [3, 4, 7, 9, 10, 17]. Om the other hand, Swedish scientists assessed the most popular maintenance approaches, i. e. strategies, policies, or philosophies, using a fuzzy multiple criteria decision making (MCDM) evaluation methodology [1]. The fuzzy AHP method proposed is a simple and effective tool for tackling the uncertainty and imprecision associated with MCDM problems, which might prove beneficial for plant maintenance managers to define the optimum maintenance strategy for each piece of equipment [14, 18].

Rolling-stock maintenance can be programmed in one of three ways: by mileage, by time or by conditioning monitoring [5, 11]. Of these three methods, condition monitoring is the most recent [6]. Many railway undertaking administrations adopted a mileage based maintenance system, although this is more difficult to operate as one has to keep records of all rail vehicle mileages and this is time consuming unless they have a modern train control and data gathering system [19]. The maintenance of rail vehicle can be characterized into two types: corrective maintenance and preventive maintenance. The time intervals at which preventive maintenance is scheduled are dependent on both the life distribution of the components and the total cost involved in the maintenance activity, but corrective maintenance cannot be avoided when component failure component occurs. The total cost of rolling-stock maintenance depends on the percentages in performing preventive maintenance and corrective maintenance. In general, more frequent preventative maintenance drives up the total maintenance costs for rolling-stock. On the other hand, proper preventative maintenance can potentially reduce the risks associated with rolling-stock mechanical failure. Thus, railway operators are constantly left weighing the safety risks against the maintenance costs. The railway safety is defined as the most crucial factor for the selection of a rolling-stock maintenance strategy [4].

Researchers Wang and Chen principally for equipment evaluated four maintenance strategies (such as corrective maintenance, time-based preventive maintenance, condition-based maintenance, and predictive maintenance) for different equipment [18]. In order to avoid the fuzzy priority calculation and fuzzy ranking procedures in the traditional fuzzy AHP methods, a new fuzzy prioritization method was proposed. This fuzzy prioritization method can derive crisp priorities from a consistent or inconsistent fuzzy judgment matrix by solving an optimization problem with non-linear constraints. Iranian scientists examined a new approach for selecting optimum maintenance strategy using qualitative and quantitative data through interaction with the maintenance experts [2]. This approach has been based on linear assignment method (LAM) with some modifications to develop interactive fuzzy linear assignment method (IFLAM). The proposed approach is an interactive method which uses qualitative and quantitative data to rank the maintenance strategies.

In contrast to maintenance strategy selection in the manufacturing industry, the maintenance of rolling-stock impacts also both traffic safety and passenger comfort [4, 16]. Because preventative maintenance and corrective maintenance affect these three factors (safety, comfort, and cost), railway system operators must establish a maintenance strategy that strives for an optimum balance. Given this, a method that defines a proper rolling-stock maintenance strategy is invaluable to system operators (railway companies), system safety supervisors (governments), and system customers (passengers). Iranian researchers proposed to apply the fuzzy Delphi method in Simple Additive Weighting (SAW) for solving the maintenance strategy selection problem [8].

Rolling-stock performance in respect of failures can be measured by MTBF (Mean Time Between Failures) or MDBF (Mean Distance Between Failures). It is sometimes measured by numbers of failures per year, month or week, but this may not represent an accurate rate consistent with mileage [18, 19]. On the other hand, rolling-stock does deteriorate rapidly in storage and this, in itself, produces failures, although these may not be the same failures seen under normal service conditions. Scientist Falco described the three case studies for existing rolling-stock, mid-life overhaul and new build rolling-stock [6].

Chinese researchers proposed to permit an approach for selecting a maintenance strategy for rolling-stock and obtaining possible spare parts' quantities and replacement intervals for the components of rolling-stock [4]. The methodology adopts an analytic network process (ANP) technique for the strategy evaluation, because ANP considers the important interactions among evaluation factors. Two Greek researchers introduced a reliability modelling and analysis framework based upon the distinct class of non-stationary Functional Series (FS) models [15]. The FS framework was applied for the modelling and analysis of two rail vehicle reliability series named as Times Between Failures (TBFs). Two models, one based on fuzzy logic (FL) and the other on artificial neural networks (ANN), were developed by Wang to predict the vehicle breakdown duration [19].

In order to improve the rolling-stock's maintenance system, it is favourable when the number of failures is proportional to the vehicle mileage or moto-hours. Then the number of failures can be predicted according to the mileage prognosis. The future needs of works and spare parts can be foreseen according to the prognosis for the number of failures. As seen from experience, the number of rolling-stock's failures is not always proportional to the mileage. Due to the rolling-stock's maintenance, the number of its failures pulsates. Since each rail vehicle is a unique product, the repair of each vehicle is somewhat distinctive, it is impossible to foresee every potential work or complication. Therefore, after the repair, the amount of failures increases for some time. This is the main reason for the pulsation of the number of failures. Each newly-installed vehicle has its own influence on the overall pulsation of the number (as well as the intensity) of the fleet's failures. This means that the pulsation of the overall number and intensity of the rolling-stock's failures depends on the fleet renewal algorithm. If renewing an existing fleet with a few dozen vehicles, the acquisition of the new vehicles reduces the overall failure amount of the fleet proportionally to the number of acquired vehicles. However, for a newly formed fleet, a different consistent pattern applies which is necessary to be studied. When researching the consistent pattern of a renewal vehicle fleet, a model concept for the change of passenger rolling-stock's interfailure according to the mileage has been formed. In accordance with this model, a fleet renewal algorithm is planned to be created in the future in order to limit the fluctuation of the fleet's average failure intensity as much as possible (by limiting the sinusoid amplitudes) and to achieve the most accurate correlation between the failure intensity and the fleet's average mileage. In order to reach this goal, the Lithuanian rail vehicle reliability researches were performed according to the vehicle types: electric multi-unit reliability research and diesel multi-unit reliability research. According to the results of these researches, a mathematical model of the renewal rolling-stock fleet's failures was created. One of the peculiarities of this methodology is that the failure flow's dependency on the vehicle mileage is proposed to be indicated using the SPLINE method [12].

2. Technical background and methodology

2.1. Indicators of the rolling-stock interfailure and reliability

Usually the research focuses on the number of failures per multi-unit or per wagon during a year. If multi-units are operated and recorded without re-forming them, then it is advisable to study the number of failures per multi-unit a year. If the composition of the multi-units constantly changes and the rolling-stock's mileage is recorded for the wagons, then the number of failures per wagon a year is studied. One of the main indicators of the reliability theory is the failure intensity (in the reliability theory it is called an intensity density of a random event). This indicator is characterized by the number of failures (of a multi-unit or wagon) per mileage unit. The period between the repairs of some vehicles is characterized by kilometres (or thousands of kilometres), sometimes – by moto-hours (thousands of moto-hours). This depends on the recommendations from the rolling-stock's manufacturer: the manufacturer provides the recommended type of a maintenance system. The railway companies usually comply with the repair system type recommended by the manufacturer, in order to avoid troubles during the technical operation of a rolling-stock.

The following rolling stock operational parameters were used when modelling the failure intensity of a passenger rolling stock fleet: the number of failures per wagon (or per multi-unit) a year and the failure intensity, the number of failures per mileage unit (per 1000 kilometres) of a wagon (or multi-unit) or the duration of a rolling-stock's operation (per 1000 moto-hours).

2.2. Research on the electric multi-unit reliability

In 2011-2012 the authors conducted a passenger multi-unit fleet's reliability research at State Company ,,Lietuvos geležinkeliai" (*Engl*. "Lithuanian Railways" , hereinafter – LG). The fleet consisted of four RA-2 series diesel multi-units and fourteen electric multi-units. During the research it was assumed that the reliability of the vehicles continuously declines at an established intensity as they age. When the fleet is supplemented with new rolling-stock, the overall failure amount of the fleet declines proportionally to the number of the new rail vehicles. Consequently, the dependence of the number of failures per wagon on the average age of the electric multi-units was measured first. This dependence of the number of failures is shown in Fig. 1.

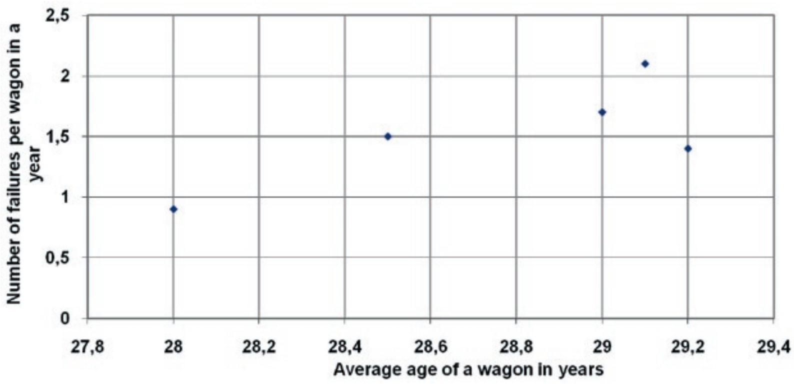


Fig. 1. Dependence of number of electric multi-unit one wagon failures on wagon's average age

The research results showed that, if the vehicles' age increases by one year, the number of their failures increases by (0.5 – 0.7) per electric multi-unit's wagon a year. During the research it was assumed that, after installing new rolling-stock to the electric multi-unit fleet, the average age of the fleet will decrease proportionally to the number of the installed multi-units' wagons. When the rolling-stock fleet's average age decreases, the failure intensity decreases proportionally. The decrease of the electric multi-units average age when their fleet is being renewed (by installing new rolling-stocks) is shown in Fig. 2.

The calculation results, presented in diagram form in Fig. 2, showed that, if the electric multi-unit fleet was renewed each year by adding three wagons (this constitutes one electric multi-unit), after five years the average age of the fleet would decrease by 10 years. The electric multi-unit's failure amount is expected to decrease proportionally to the reduction of the fleet's average age. The dependence of the failure amount's decrease on the number of newly-acquired electric multi-unit wagons is shown in Fig. 3.

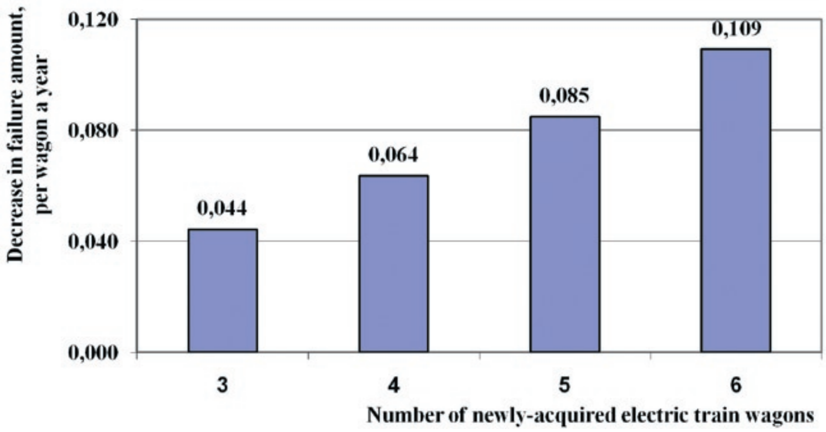


Fig. 3. Dependence of the decrease in electric multi-unit failures on the number of newly-acquired electric multi-unit wagons

The calculation results (the diagram in Fig. 3) show that, if 6 wagons were acquired, the fleet's failure amount would decrease by 0.109 failures per wagon a year, and if 3 wagons were acquired – by 0.044 failures per wagon a year. This means that, on average, each newly-acquired electric multi-unit's wagon reduces the overall electric multi-unit failure amount by (0.015 – 0.018) failures per wagon a year.

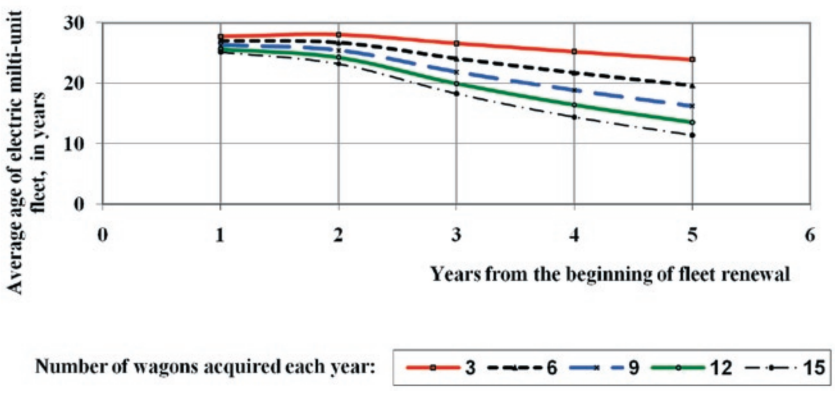


Fig. 2. Decrease of the electric multi-unit fleet average age during the fleet renewal process

To sum up the results, the following preliminary conclusions of the electric multi-unit reliability research are made. According to the results of the vehicle failure intensity research, a mathematical model was made that indicates a consistent pattern between regular renewal of the passenger vehicle fleet and the vehicle failure intensity. The mathematical model was implemented using the LG electric multi-unit fleet as an example. It was learned that, if the existent electric multi-unit fleet of 108 wagons was supplemented by one multi-unit (3 wagons), the fleet's average failure amount (0.6 failures a year) per wagon would decrease by 0.044 failures a year – in other words, if 3% of the vehicle fleet was renewed, its failure amount would decrease by 7.3%. This model can be used for predicting the change of failure amount when planning to acquire a small number of passenger vehicles (1-2 diesel or electric multi-units consisting of 3 wagons) for a period of several years (up to 5). In other words, the model is to be used for predicting the changes of electric multi-unit fleet's failure amount in the beginning of the fleet renewal (in the first decade of the new wagon mileage), when renewing the fleet by (10 – 20) %. This mathematical model should not be relied on when predicting the vehicle fleet's failure intensity for longer periods or bigger number of acquired vehicles. The model could be improved by evaluating more factors and their consistent patterns of influencing the vehicle failure amount. The essential weakness of this model is that it does not consider the aging process of the newly-acquired wagons and its influence on the whole fleet's failure amount.

Further LG passenger vehicle researches showed that the vehicles with an internal combustion engine, i. e. diesel multi-units, have the most complex aging process [7, 17].

2.3. Investigation of diesel multi-unit failure amount

When studying the failure intensity of the vehicles with internal combustion engines, it can be seen that the intensity is closely related to the vehicle maintenance system type. In the beginning of operation, the failure intensity increases due to the peculiarities of installing the vehicles for operation. The rolling-stock manufacturer is not always able to anticipate the real (factual) conditions of the produced vehicles' operation (e.g., load, usage intensity, climate conditions, maintenance work culture etc.). Consequently, in the beginning of the vehicle operation the failure intensity increases for some time (such failures are sometimes called "childhood diseases"). During the unscheduled maintenance, these failures are removed, the most appropriate operational materials are selected (depending on the load, climate), as well as more experience on how to properly operate such vehicles is gained. The failure intensity decreases (stabilizes) for the time being. However, after a while, a permanent repair needs to be done. The vehicle constructions and materials used for manufacturing are constantly improving; therefore each permanent repair of a vehicle is partly unique. After the repair, unforeseen consequences appear. Manufacturer's repair recommendations are not always explicit and specific. Various components and parts of a rolling-stock are often produced by different manufacturers who provide completely different recommendations for operating and repairing the vehicles. For instance, the recommended period for changing engine oil is provided by both a diesel engine manufacturer and an engine oil manufacturer. The recommended period sometimes differs by two or even three times. When planning the works of vehicle's permanent repair, the decision makers of railway companies are not always certain which recommendations to follow. In such cases, the decision makers improvise. Such decisions not always are the best, resulting in the increase in rolling-stock's failure amount when starting its operation after an ordinary permanent repair. When operating the rolling-stock, the mistakes and defects made during the repairs are removed, therefore the resulting vehicle failure intensity decreases (is "contained") for the time being. But due to an elementary deterioration of rolling-stock's parts, the failure intensity starts increasing again until the next permanent repair. Thus forms periodic failure intensity's dependence on rolling-stock's operation as the first permanent repairs are performed on a regular basis. The periodic failure intensity dependence of the LG diesel multi-units RA-2 on the diesel multi-unit mileage is presented in Fig. 4. The failure intensity values provided in the fig. 4 diagram are calculated as a ratio of the failure amount to multi-unit mileage.

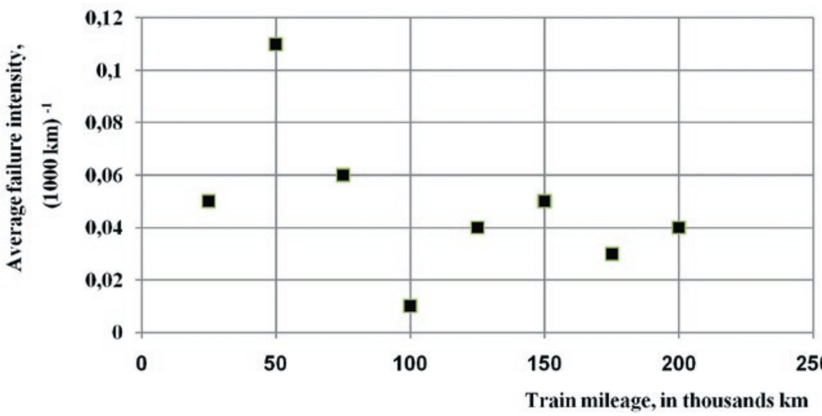


Fig. 4. Failure intensity dependence on diesel multi-unit mileage

The periodic failure intensity dependence on the diesel multi-unit mileage is characterized by the fact that, when the mileage is around 50 000 km, the failure intensity peaks to more than 0.1 failures per 1000 km mileage. When the mileage reaches 100 000 km, the failure intensity falls to 0.01 failures per 1000 km mileage, i. e. by ten times. Later, when the mileage is 150 000 km, the failure intensity rises again to (0.04 – 0.05) failures per 1000 km mileage. Such a variation is determined by the scheduled diesel multi-unit repair warning system used by LG [7]. Until the first permanent repair, the failure intensity stays around 0.5 failures per 1000 km mileage, after the repair it increases more than twofold, later decreases by around 10 times due to the unscheduled repairs. After this cycle has passed, at (100 – 125) thousands km mileage the second cycle starts: due to the repair peculiarities, the failure intensity this time increases not twofold, as in the first cycle, but only by a quarter (from 0.04 to 0.05 failures per 1000 km mileage).After 150 000 km mileage, the failure intensity decreases once more. When the mileage reaches around 200 000 km, the failure intensity steadies around 0.04 failures per 1000 km mileage. The variation consists of 25% – this is trivial compared to the variation during the first 100 000 km mileage. The amplitudes of this variation are likely to be reduced by improving the rolling-stock repair technologies. It should be noted that the failure intensity itself is regular and its periodicity of a sinusoid form cannot be removed. Acquiring diesel multi-units on a regular basis, it would very unacceptable, if the failure intensity maximums of several multi-units coincided – "added up" (e.g., when the mileage of one multi-unit was 50 000 km, the mileage of another multi-unit would be 150 000 km). Such a coincidence would substantially destabilize the fleet's overall failure intensity, i.e. greatly reduce the technical readiness level of a fleet. When modelling

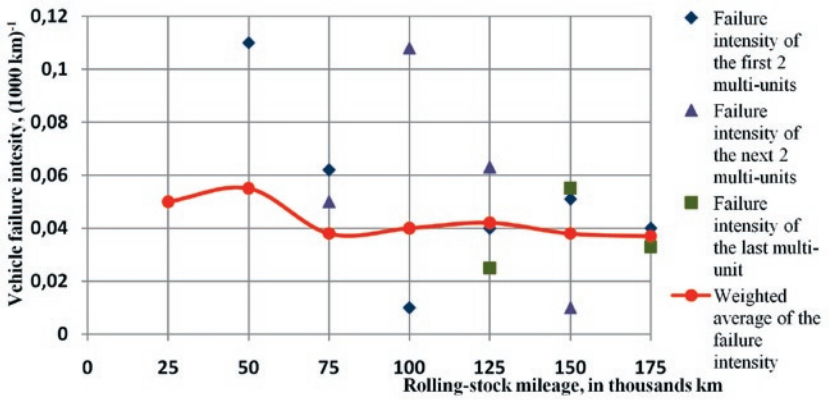


Fig. 5. Failure intensity of diesel multi-unit fleet when the fleet is renewed according to the formula „2+2+1"

the failure intensity, it was learned that the it changes quite consistently, if diesel multi-units are acquired after every 50 000 km mileage under the formula "2+2+1". Such a change of the diesel multi-unit fleet failure intensity is shown in Fig. 5.

When the diesel multi-unit fleet is formed under the formula "2+2+1" (after every 50 000 km mileage), two diesel multi-units are installed at first. In Fig. 5, this moment matches the 0 on the X axis. When the mileage of the first two multi-units reaches 50 000 km (reaches the maximum failure intensity), two more new diesel multi-units are installed. When the mileage of the first two multi-units reaches 100 000 km (the mileage of the other two multi-units then reaches 50 000 km), one more diesel multi-unit is installed. The result of following such a method for rolling-stock renewal (acquisition) can be seen in Fig. 5. When the mileage of the first two multi-units is 50 000 km, the curve in Fig. 5 shows the maximum failure intensity. When the mileage of these multi-units is 100 000 km, the curve in Fig. 5 shows the maximum failure intensity of the other multi-units. When the mileage of the first multi-units is 150 000 km, the curve shows the maximum failure intensity of the last multi-unit. The average failure intensity of a fleet is calculated as a weighted average, taking into account the number of diesel multi-units:

$$\lambda = \frac{\sum \lambda_i \cdot n_i}{\sum n_i}; \quad (1)$$

where: λ_i – the failure intensity of a multi-unit group (instalment) No. i, (1000 km)$^{-1}$; n_i – the number of multi-units of a multi-unit group (instalment) No. i.

After implementing the Formula (1) on the 2+2+1 basis, it would look like this:

$$\lambda = \frac{\lambda_1 \cdot 2 + \lambda_2 \cdot 2 + \lambda_3 \cdot 1}{5}. \quad (2)$$

It should be noted that, in the denominator of the Formula (2), "*5*" appears only if all five multi-units are installed, i.e. if all 3 summands are in the numerator. The multipliers "*2*" , "*2*" and "*1*" in the numerator are weighted coefficients that take into account the number of multi-units. The diagram of Fig. 5 shows the failure intensity after the weighted coefficients are taken into account. That is why the maximum at 150 000 km mileage is two times lower than the two first maximums. The line curving around 0.04 failures per 1000 km mileage is the average failure intensity. The more appropriate modelling approach is when the continuous functions are used. The examples of modelling that have been presented in this article so far are the cases of discrete modelling. The failure intensity's (or another parameter's) values are discretely attributed to the mileage intervals based on which the required actions with the sets are performed. For visualization, the diagram points that represent parameters are connected, thus making the imitation of a continuous function. Such modelling approach is simple, visual and is well-suited for modelling non-complex processes (when one or several extremes are present). However, if there are more extremes or the dependence function is more complex, then it is better to analyse the dependence by using the continuous function. Fig. 6 shows an example of diesel multi-unit failure intensity's approximation by the continuous function.

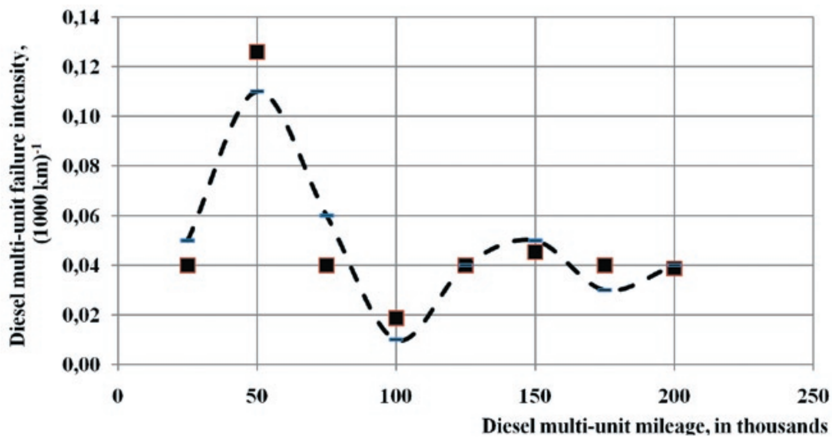


Fig. 6. Approximation of diesel multi-unit failure intensity by continuous function

In Fig. 6, the discrete approximation of points function is made on the basis of the exponentially diverging sine function:

$$\lambda = A \cdot e^n \cdot \sin(C \cdot x - B) + D; \quad (3)$$

where: A – amplitude coefficient; B – phase coefficient; e^n – decrease pattern; C – wavelength ratio; D – average divergence. The measuring units of A and D coefficients matches the measuring unit values marked according to the Y axis.

The units of coefficients B and C are radians per 1000 km or per 1000 moto-hours. The dispersion of the diesel multi-unit failure intensity equals 0.00095. According to these values, a conclusion can be made that the approximation is fairly appropriate.

3. Applying of SPLINE method for multi-unit interfailure modelling

In mathematical science the SPLINE approximation is a known method of compromise [12]. Its basic principle is to divide the dependence function domain into segments where data is approximated by functions in such a way that they would form one consistent pattern. The points at which the diagram goes from one segment to the other are called spline knots. The most simple is the approximation by the splines of the second degree. Coefficients of quadratic equations are calculated from the condition that the derivatives of both functions at the spline knot must be equal. Thus at that point the tangent of both function graphs is the same straight line. If the splines are made from the curves of the third degree (the function coefficients of the third degree are calculated), an additional condition appears, and stating that at the knots the second derivatives of the functions must also be accordingly equal. This method was also applied when researching the LG diesel multi-unit failure amount. At first, the approximation by the linear splines was used, later – the one by the second degree splines. A conclusion was made that the lowest spline degree, when the approximation is getting adequate to the physical phenomena, is the third one. As a basis for creating the third degree splines, the following equation is used:

$$f(x) = y_i + a \cdot (x - x_i) + b \cdot (x - x_i)^2 + c \cdot (x - x_i)^3; \quad (4)$$

where: x and y – point coordinates; a, b and c – coefficients of the third degree equation. These last-mentioned coefficients are calculated according to the following formulas:

$$a = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} - \ddot{f}_{i+1} \frac{x_{i+1} - x_i}{6} - \ddot{f}_i \frac{(x_{i+1} - x_i)}{3}; \quad (5)$$

$$b = \frac{\ddot{f}_i}{2}; \quad (6)$$

$$c = \frac{\ddot{f}_{i+1} - \ddot{f}_i}{6 \cdot (x_{i+1} - x_i)}. \quad (7)$$

The following are the results of the solution found using the MAPLE software package. Practically speaking, this method has one limitation. The diesel multi-unit mileage is indicated in kilometres: 50 000 km, 100 000 km etc. The SPLINE diagram does not accept such indication method as it needs a reference system on the basis of *1, 2, 3, 4*, etc. Mark "*2"* of the SPLINE diagram matches the 50 000 km mileage, mark "*4"* – 100 000 km, etc. Therefore, when reading the formulas and diagrams, the rolling stock mileage has to be calculated separately. Firstly, the splines are indicated by the formulas (third degree functions with corresponding coefficients, within the range of the "*x*" mileage):

$$\begin{cases}
17+\frac{1223}{1991}\cdot x-\frac{51453}{3982}\cdot x^2+\frac{17151}{3982}\cdot x^3, x<2;\\
\frac{238407}{1991}-\frac{305617}{1991}\cdot x+\frac{23217}{362}\cdot x^2-\frac{33989}{3982}\cdot x^3, x<3;\\
-\frac{587901}{1991}+\frac{520691}{1991}\cdot x-\frac{295485}{3982}\cdot x^2+\frac{27219}{3982}\cdot x^3, x<4;\\
\frac{640707}{1991}-\frac{400765}{1991}\cdot x+\frac{165243}{3982}\cdot x^2-\frac{11175}{3982}\cdot x^3, x<5;\\
\frac{94082}{1991}-\frac{72790}{1991}\cdot x+\frac{34053}{3982}\cdot x^2-\frac{2429}{3982}\cdot x^3, x<6;\\
-\frac{704254}{1991}+\frac{326378}{1991}\cdot x-\frac{99003}{3982}\cdot x^2+\frac{4963}{3982}\cdot x^3, x<7;\\
\frac{36663}{181}-\frac{148285}{1991}\cdot x+\frac{36615}{3982}\cdot x^2-\frac{1495}{3982}\cdot x^3, x<8;\\
\frac{779613}{1991}-\frac{289405}{1991}\cdot x+\frac{71895}{3982}\cdot x^2-\frac{2965}{3982}\cdot x^3, x<9;\\
-\frac{2266149}{1991}+\frac{725849}{1991}\cdot x-\frac{153717}{3982}\cdot x^2+\frac{5391}{3982}\cdot x^3, x<10;\\
\frac{160441}{181}-\frac{2671}{11}\cdot x+\frac{8013}{362}\cdot x^2-\frac{2671}{3982}\cdot x^3, otherwise.
\end{cases} \tag{8}$$

The domain is divided into 11 segments (subdomains) each of which has one consistent pattern (see Formula (8)). This mathematical expression is shown in diagram form in Fig. 7.

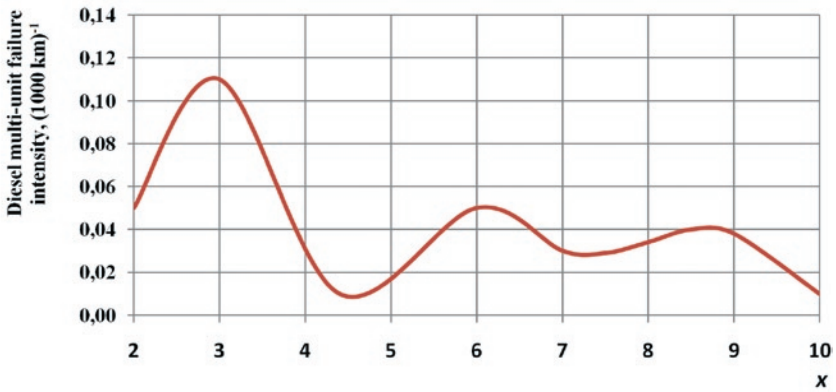


Fig. 7. Approximation of diesel multi-unit failure data by splines of the third degree

From the diagram in Fig. 7, it is obvious that, using the spline method, a fairly complex pattern can be indicated by non-complex mathematical functions. This is very convenient for indicating the failure intensity of a rolling-stock – in this instance, diesel multi-units. The main advantage of this method is that the spline limits (as well as the domain section limits) can be chosen according to the physical phenomena of a process. In the rolling-stock operation this can be the maintenance regularity, in certain cases - the moments of crashes or failures. After indicating the rolling-stock failure patterns by one of the above-mentioned methods (discrete, continuous function or spline), the failure patterns of the whole fleet of a railway company can be modelled in the longer term, taking into account the fleet formation (renewal) algorithm. After making such rolling-stock renewing schedule, a railway operator can avoid the unwanted failure intensity peaks, as well as the unexpected costs of manpower and financial expenditures for repairing the rolling-stock.

4. Conclusions

During this study of the new rail vehicles acquisition influence on the fleet's overall failure amount, it was noticed that this influence depends on whether the fleet is renewed or newly formed. If renewing a large existing fleet with a few dozen or hundred vehicles, then it can be assumed that the acquisition of the new vehicles reduces the overall failure amount of the fleet proportionally to the number of acquired vehicles. This assumption should not be applied when the fleet is newly formed. When forming a new fleet, the rolling-stock is to be acquired in certain cycles. If the rolling-stock failure intensity fluctuates, especially if the fluctuations are regular, the peaks of the newly-acquired vehicle failure intensity can coincide with the peaks of the earlier-acquired vehicle failure intensity. The research showed that, if the failure intensity fluctuation extremes coincide, the unexpectedly large peaks of the fleet failure intensity, as well as the vehicle repair costs, are possible. Therefore, when forming a new rolling-stock fleet, it is necessary to as gradually as possible disperse in time the total failure intensity extremes.

In order to avoid the possible and unexpectedly large peaks of the whole rolling-stock fleet failure intensity, it is necessary to mathematically model the patterns of their failure intensity dependence on operation, and to use these models for choosing the appropriate moment to install a new vehicle. For modelling the vehicle failure intensity dependence on operation, the discrete method, polynomial approximation, and sine pattern approximation are proposed. The more radical proposal of the authors is to approximate this pattern using splines. The latter method is convenient, as the spline limits can be chosen according to the physical phenomena of a process, thus evaluating the moments of maintenance.

After indicating the rolling-stock failure patterns by one of the above-mentioned methods, the failure patterns of the whole vehicle fleet can be further modelled, taking into account the fleet formation (renewing) algorithm. After making such rolling-stock fleet renewing schedule, in the future a railway company can avoid the unexpected failure intensity peaks, as well as the unforeseen costs for unscheduled repair of the rolling-stock.

Acknowledgements

This work has been supported by the European Social Fund within the project "Development and application of innovative research methods and solutions for traffic structures, vehicles and their flows", project code VP1-3.1-ŠMM-08-K-01-020.

Reference

1. Al-Najjara B, Alsyouf I. Selecting the most efficient maintenance approach using fuzzy multiple criteria decision making. Int. J. Production Economics 2003; 84: 85-100.
2. Bashiri M, Badri H, Hejazi T H. Selecting optimum maintenance strategy by fuzzy interactive linear assignment method. Applied Mathematical Modelling 2011; 35: 152-16.
3. Bureika G. Multicriteria evaluation of operational effectiveness of freight diesel locomotives on Lithuanian Railways. Transport 2011; 26(1): 61-68.

4. Cheng Y H, Tsao H L. Rolling stock maintenance strategy selection spares parts' estimation, and replacements' interval calculation. Int. J. Production Economics 2010; 128: 404-412.
5. Cherry C R. Development of duration models to determine rolling stock fleet size. Journal of Public Transportation 2005; 8 (3): 57-72.
6. Falco P J H, MIMechE M C, Dongen L A M. Application of remote condition monitoring in different rolling stock life cycle phases. Proceedings of the 2nd International Through-Life Engineering. ProcediaCIRP 2013; 11. p. 135-138.
7. Gelumbickas G, Vaičiūnas G. Analysis of passenger rolling stock faults and its statistics in Lithuania. Transport 2011; 26 (3): 315-319.
8. Jafari A, Jafarian M, Zareei A, Zaerpour F. Using fuzzy Delphi method in maintenance strategy selection problem. Journal of Uncertain Systems 2008; 2(4): 289-298.
9. Kabo E, Ekber A. Material defects in rolling contact fatigue of railway wheels – the influence of defect size. Wear 2005; 258: 1194-1200.
10. Karlaftis M, Sinha K. Modelling approach for transit rolling-stock deterioration prediction. J. Transp. Eng. 1997; 123(3): 223–228.
11. Kazopoulo M, Kaysi I, El Fadel M A. stated-preference approach towards assessing a vehicle inspection and maintenance program. Transportation Research Part D 2007; 12 (5): 358-370.
12. Lyche T, Morken K. Spline methods draft. University of Oslo. 2008. 226 p.
13. Sawicki P, Zak J. Technical diagnostic of a fleet of vehicles using rough set theory. European Journal of Operational Research 2009; 193 (3): 891-903.
14. Sivilevičius H, Maskeliūnaitė L. The criteria for identifying the quality of passengers' transportation by railway and their ranking using AHP method. Transport 2010; 25(4): 368-381.
15. Stavropoulos Ch N, Fassois S D. Non-stationary functional series modelling and analysis of hardware reliability series: a comparative study using rail vehicle interfailure times. Reliability Engineering and System Safety 2000; 68: 169-183.
16. Sun K, Li H. Scheduling problems with multiple maintenance activities and non-pre-emptive jobs on two identical parallel machines. International Journal of Production Economics 2010; 124 (1): 151-158.
17. Vaičiūnas G, Lingaitis L P. Investigating the dynamics of passenger rolling stock deterioration. Transport 2008; 23(1): 51-54.
18. Wang L, Chu J, Wu J. Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process. Production Economics 2007; 107 (51): 151-163.
19. Wang W, Chen H, Bell M C. Vehicle breakdown duration modelling. Journal of Transportation and Statistics 2005; 8 (1): 75-84.

Gediminas VAIČIŪNAS
Gintautas BUREIKA
Department of Railway Transport
Vilnius Gediminas Technical University
J. Basanavičiaus str.28, LT-03224 Vilnius, Lithuania
E-mails: gediminas.vaiciunas.@vgtu.lt; gintautas.bureika@vgtu.lt

Article citation info:
SZKODA M. Assessment of reliability, availability and maintainability of rail gauge change systems. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 422–432.

Maciej SZKODA

ASSESSMENT OF RELIABILITY, AVAILABILITY AND MAINTAINABILITY OF RAIL GAUGE CHANGE SYSTEMS

OCENA NIEUSZKADZALNOŚCI, GOTOWOŚCI I PODATNOŚCI UTRZYMANIOWEJ KOLEJOWYCH SYSTEMÓW PRZESTAWCZYCH*

The paper provides a comparative assessment of the reliability of two rail gauge change systems: wagon bogie exchange and SUW 2000 self-adjusted wheel sets. In the applied method of assessment, reliability is treated as a comprehensive feature comprising such system characteristics as reliability itself together with availability and maintainability. The calculations of selected reliability ratios, based on operation data, demonstrated that the SUW 2000 system may be an alternative method for overcoming the barrier of different track gauges compared to the current wagon bogie exchange.

***Keywords**: reliability analysis, gauge change system, track gauge change.*

Praca dotyczy porównawczej oceny niezawodności dwóch kolejowych systemów przestawczych: systemu wymiany wózków wagonowych i systemu samoczynnie rozsuwanych zestawów kołowych SUW 2000. W zastosowanej metodzie oceny, niezawodność jest traktowana jaka właściwość kompleksowa obejmującą takie cechy systemów jak: nieuszkadzalność, gotowość i podatność utrzymaniową. Przeprowadzone obliczenia wyselekcjonowanych wskaźników niezawodnościowych, oparte na danych eksploatacyjnych wykazały, że system SUW 2000 może stanowić alternatywną metodę pokonania bariery rożnej szerokości toru w stosunku do aktualnie stosowanej wymiany wózków wagonowych.

***Słowa kluczowe**: analiza niezawodności, system przestawczy, zmiana szerokości toru.*

1. Introduction

Economic development depends largely on an efficient transport system which should enable reliable, safe and efficient cargo transport both domestically and internationally. It is particularly difficult to ensure such conditions for railway transport between Europe and Asia. This relates to the different track gauges on the Euro-Asian continent. Most European states, Poland included, have 1435 mm tracks while the railways within the former Commonwealth of Independent States (CIS) and other countries, such as Lithuania, Latvia and Estonia, are 1520 mm in width. In Asia, trains move along 1520 mm tracks, to go back again, in China and Korea, to normal gauge lines of 1435 mm. In Spain and Portugal, the tracks are even wider – 1668 mm.

The need for research aimed at facilitating the methods of overcoming the barrier of different track gauges may be demonstrated through the fact that the theme is addressed by international consortia under EU programmes and by the International Union of Railways (UIC) [3, 7, 10]. In 1995–2005, at the Institute of Rail Vehicles, Cracow University of Technology, a number of research and development assignments and goal-oriented projects were done on this issue [2, 21, 31, 33, 38]. Recent international writings too, include studies into the transport between railways of different track gauges, and in particular between Europeand Asia [5, 13, 22, 27, 36, 40]. Their authors emphasize that the development of railway transport on the Euro-Asian continent is possible through the implementation of new effective methods of overcoming the barrier of different track gauges. Currently, the time lost at border crossing points to handle goods or replace the vehicle running assemblies, together with the document flow, amounts to as much as 46% of the total time of transport [13, 23].

In cargo transport, two technologies to overcome the barrier of different track gauges are now in place [2, 31]:

- handling, and
- gauge changing.

The handling technology consists in reloading the cargo at border-crossing points from normal to wide gauge wagons or vice versa. Depending on the type of goods, the following can be distinguished in this technology: reloading, pumping or pouring [31].

Rail gauge change systems, which are discussed in this paper, are based on the other possible method of transport between railways with different track gauges, so-called gauge change technology [2]. In such systems, cargo is moved by the same means of transport which is shifted at the border-crossing point from one track gauge to another. The shifting may be done through wagon bogie exchange; exchange of the wheel sets or the use of self-adjusted wheel sets. Currently, cargo transport uses only wagon bogie exchange with the lifting of the wagon body. Methods which consist in exchanging the wheel sets were analysed under research project [21], but have not been put into practical application. In the second half of last century, intensive research was done in Germany, Spain, Russia, Bulgaria, Poland and Japan on automated technologies to overcome the barrier of different track gauges (so-called self-adjusted wheel set systems). The systems which have found practical application include: Talgo (Spain), DB Rafil (Germany), BT (Bulgaria) and the Polish SUW 2000. Under project [3], a comparative analysis of these systems was performed. The main emphasis was placed on legal, economic and logistic aspects, as well as the benefits for railway operators. The research demonstrated that the Polish SUW 2000 system designed by Dr hab. inż. R. Suwalski [5] was most elaborate and technologically advanced. The first years of the system's operation were devoted to work on eliminating the phenomenon of fretting between the axle bearing and the wheel nave bushing. Analysis of different options for the structure and materials enabled the introduction of the self-adjusted wheel sets

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

system into commercial supervised operation. Papers [5, 40] present the results of research into supervised operations of the SUW 2000 system, done in 2003–2008 in passenger and cargo transport between Poland and Ukraine. Currently, the SUW 2000 self-adjusted wheel sets system is applied in passenger transport only between Poland and Ukraine (Wroclaw–Lviv train).

All work done to date on the application of the SUW 2000 system to cargo transport has been limited to comparative analyses of this concept in technical and economic terms, no account being taken of its reliability. Hence, an attempt has been made to provide an assessment of the SUW 2000 self-adjusted wheel sets system compared with wagon bogie exchange which is currently used to transport hazardous materials. Analysis of the existing situation demonstrates that the gauge change transport system requires to be improved, in the transport of hazardous material in particular. The current concepts applied at border crossing points along Poland's eastern border for this cargo group are characterised by low efficiency and pose a serious threat to the environment and the safety of the handling personnel. This allows a supposition that the results represented herein are new and important for further research and development work on innovative methods of overcoming the barrier of different track gauges in cargo transport.

2. Assessment of reliability, availability and maintainability

Rail gauge change systems are amongst renewable facilities. Methods used for items which operate until the first failure, i.e. with the use of the functions of reliability *R(t)* or intensity of failures *λ(t)* are insufficient for such facilities. In the assessment method as herein applied, reliability is treated as a comprehensive characteristic which encompasses such features of the systems as reliability itself together with availability and maintainability (RAM). These may be defined as follows [16, 19]:

- Reliability is understood as the system's capability to perform a required function under stated conditions for a specified period of time;
- Availability is the system's capability of being in an operable state to perform the required functions under stated conditions, at a specified moment or for a specified time, presuming that the required external means are provided;
- Maintainability is defined as a characteristic of adaptation to restoration done in order to restore the item to a specified condition of operation with the use of prescribed methods and resources.

For a system to be in the state of availability means that it is not out of operation due to preventive maintenance or is not incapable of use due to failure. Availability depends not only on maintenance downtimes but also on the probability of the system's failure to perform its functions (unreliability effect) [15]. Maintainability in respect of rail gauge change systems concerns corrective and preventive maintenance. Corrective maintenance enables restoration of the item's capability and its being put back into operation. Preventive maintenance, on the other hand, is done as part of a prescribed item's maintenance cycle in order to improve its reliability and control its wear [25]. Theaim behind effective maintenance is to minimise the Mean Down Time (*MDT*) and the related costs [39].

General guidelines on the analysis of reliability, availability and maintainability are provided in the PN-EN 50126 standard on *Railway Applications – Specification and Demonstration of Reliability, Availability, Maintainability and Safety* [19]. Professional writings on this problem offer a detailed description, definitions and calculation formulae for the different ratios used in the assessment [1, 8, 14, 17, 26]. The reliability, availability and maintainability analysis has been the subject of many research projects in recent decades. It is currently applied in various industries, including aviation, armaments, power engineering, food processing and transport [6, 9, 11, 15, 24, 25, 35]. For instance, in paper [9], the authors describe potential applications of the RAM model in industrial practice to identify equipment which is critical due to frequent failures or high maintenance requirements. Paper [24] presents an application of the RAM analysis as a helpful tool to design systems, introduce constructional changes in order to achieve the minimum number of failures and increase the Mean Time between Failures (*MTBF*). Paper [35] analyses the problem of relation between the availability and maintainability of means of rail transport, and the costs of planned vehicle downtimes and maintenance. A model for optimum inspections and periodical restorations has been set from the viewpoint of costs, taking into account current data on vehicle failures. In the present paper, the reliability, availability and maintainability analysis applies to rail gauge change systems.

3. Comparative analysis of rail gauge change systems

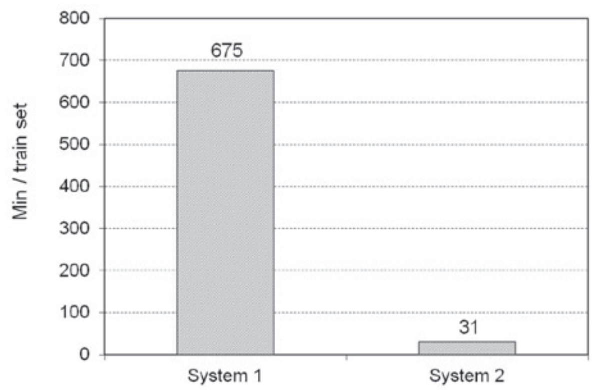
3.1. Systems under analysis

Two systems for transporting hazardous materials are assessed for their reliability, availability and maintainability:

- **System 1** where track gauge change is effected through wagon bogie exchange, with the lifting of the wagon body, as currently applied;
- **System 2** where the track gauge is changed with the use of the prospective method – the SUW 2000 self-adjusted wheel sets.

Figure 1 shows a comparison of the process of operation of the analysed systems. The analysis leaves out the duration of the operations which consist of the train receipt, i.e. checking the securities and conformity of the shipping documents, customs clearance and wagon weighing.

a)



b)

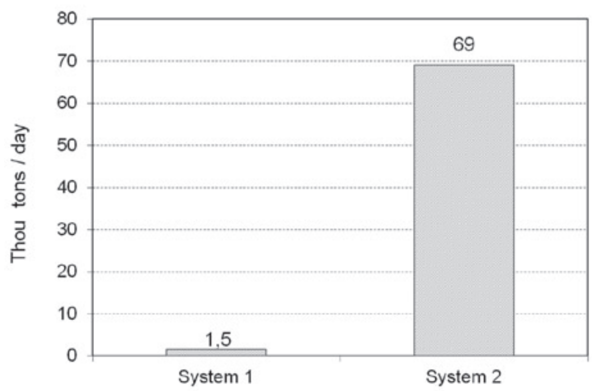


Fig. 1. a) Service time at the border crossing point, b) System performance [30, 32]

Considering the time needed for the servicing of a train at a border crossing point and the resulting performance, the option with the SUW 2000 with self-adjusted wheel sets is unrivalled. The application of the system shortens considerably the time of transport down by up to 18 hours depending on the cargo [4, 29]. There are, however, limitations relating to the universality of the service. This technology requires full-train transports orpreliminary distribution before the contact point of different track gauges.

3.2. Reliability data

The basis for assessing the systems' reliability, availability and maintainability was the operational data gathered in collaboration with PKP Cargo S.A. in actual work conditions covering about 7 years of operation for the wagon bogie exchange system, and almost 4 years for the self-adjusted wheel sets. This enabled observation of the course of operation of system elements in a variety of conditions and thus provided accurate data for reliability assessment. The reliability data was gathered in internal reports and records of PKP Cargo S.A., which performed the function of operation sheets. The documents contained detailed information on:

- date of failure,
- circumstances in which the failure was identified,
- causes for failure,
- maintenance time characteristics, i.e. duration of corrective maintenance, organisational downtime (waiting for the restoration, waiting for collection after restoration),
- labour intensiveness of corrective maintenance,
- figures for the different measurable characteristics before and after the restoration,
- labour intensiveness and duration of scheduled restorations (inspections, periodical restorations),
- consumption of materials and spare parts,
- technologies of restorative operations, and
- other additional information.

The detailed data on the process of operation, number and types of recorded failures of the examined systems are provided in [33].

3.3. Presumptions for and structures of the systems' reliability

The assessment of the reliability, availability and maintainability of the systems concerned was comparative in its nature. Thus, the common elements which have the same effect in both systems, e.g. 1435 and 1520 mm rail infrastructure, traction vehicles and others, were excluded from the analysis and hence from the reliability structure. The interest in the compared systems focused on elements of technical equipment of the points of junction between the different track gauges, and the rolling stock engaged in the transport process.

In system 1, wagon bogie exchange stands together with the cooperating gantry cranes are used to move a wagon from one track gauge to another (Figs. 2a and 2b). In system 2, the extended technical infrastructure of the wagon bogie exchange point is replaced with a track gauge change stand (Fig.3).

As regards the rolling stock, the most significant differences in the reliability assessment concern wagon bogies directly responsible for transport safety. In system 1, two sets of bogies assigned to one wagon are required to effect transport along tracks of different gauges: a 2XTa bogie for a1435 mm track (Fig. 4a) and a 18-100 type bogie for a track of 1520 mm, which are exchanged at the border crossing point. In system 2, on the other hand, bogies of one type – 4RS/N (Fig. 4b) – equipped with adjusted wheel sets are used thus enabling the wagon to move along 1435 and 1520 mm rail tracks.

In order to ensure practical usefulness of the paper, in formulating the presumptions necessary for the assessment of reliability, availabilityand maintainability, reference was made to the current condition

a)

b)

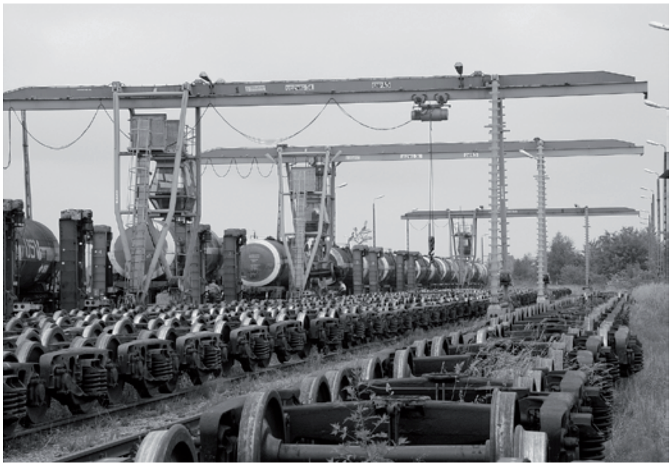
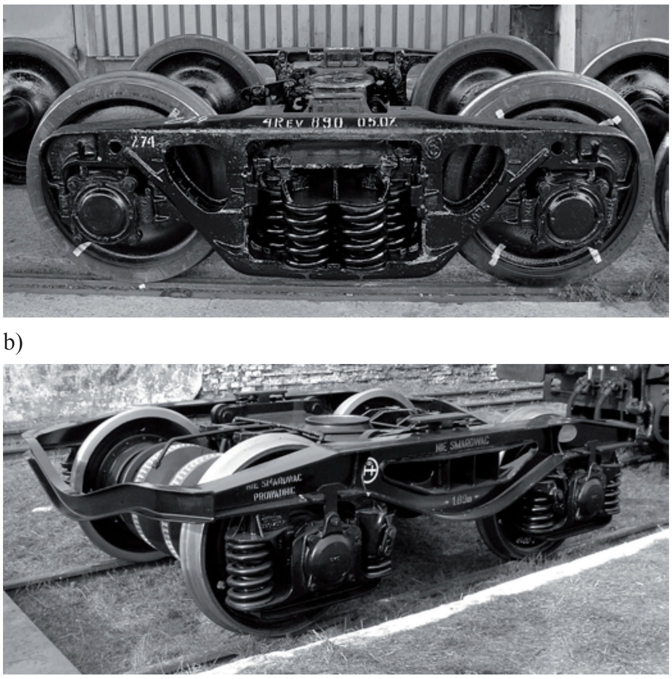


Fig.2. a) Wagon bogie exchange stand b) Gantry cranes to operate the gauge change stands (Photo: M. Szkoda)

Fig. 3. Rail gauge change stand for the SUW 2000 system [28]

Fig. 3. Rail gauge change stand for the SUW 2000 system [28]

a)



b)

Fig. 4. a) 2XTa wagon bogie for a 1435 mm track, b) 4RS/N wagon bogie for a 1435/1520 mm track (Photo: M. Szkoda)

at the bogies exchange point of PKP Cargo S.A., located at the largest Polish-Ukrainian border crossing point at Medyka/Mostiska. The assumptions for the analysis are shown in Table 1.

Table 1. Assumptions for the reliability analysis

No.	ELEMENT	ASSUMPTIONS
1	Type of cargo transported	Hazardous materials in cistern wagons
2	Number of wagons exchanged at the border crossing point	5,483.0 [weight/year]
3	Capacity of the exchanged wagon	48.0 [tonnes]
4	Wagon turnover:	
	- system 1	10.6 [days]
	- system 2	8.0 [days]
5	Transport distance (one way, half along a 1435 mm track and half along a 1520 mm track)	1,100.0 km

In Table 2, for the presumed transport distance, the demand for means of transport was calculated, including the wagon bogies involved in the transport using both systems. The lower demand in system 2 is due to a much shorter servicing process at the border crossing point.

The foregoing presumptions, combined with the analysis of the actual situation, make it possible to determine the reliability structures of the analysed systems. Various structures were used to describe reliability: serial structure, with sliding reserve and the structure k out of n type. The methodical basis for assessing the reliability of systems with sliding reserve and one with the k out of n type are presented, *inter alia*, in [34].

The reliability structure of system 1 (Fig. 5) is mapped through serially connecting four subsystems – P1.1, P1.2, P1.3 and P1.4:
Denotations in Fig.5:

P1.1, P1.2, P1.3, P1.4 –subsystems of system 1,

1.1_i, 1.2_i, 1.3_i, 1.4_i – elements which are comprised in system 1 (1.1 –2XTa wagon bogies, 1.2 –18-100 wagon bogies, 1.3 –bogie exchange stands, 1.4 – gantry cranes).

- subsystem P1.1 comprises a total of 176 bogies of the 2XTa type for a 1435 mm track (element 1.1), making up a reliability structure with sliding reserve with the order of redundancy k = 10. It means that for 160 basic bogies, an operational reserve of 16 elements is presumed, each of which can replace any basic bogie in the event it fails;
- subsystem P1.2 is made of a total of 176 bogies of the 18-100 type for a 1520 mm track (element 1.2), which, by analogy to subsystem P1.1, are mapped by a reliability structure with sliding reserve with the order of redundancy k = 10. Analysis of subsystems P1.1 and P1.2 assumes that the reserve bogies cannot fail when not in operation and that a bogie's non-operating condition does not affect its reliability. It is assumed further that the time during which a destroyed bogie is replaced by a reserve element practically equals zero;
- subsystem P1.3 consists of 14 bogie exchange stands (element 1.3) which are mapped as a threshold structure of the 10 out of 14 type. At least 10 stands are necessary to achieve the assumed number of wagons exchanged in a year at the border crossing point. The 10 out of 14 threshold structure means that subsystem P1.3 is in the state of correct operation when at least 10 out of 14 bogie exchange stands perform the functions they are allocatedin a correct fashion;
- subsystem P1.4 includes 3 gantry cranes (element 1.4) which are mapped by means of the serial reliability structure.

The reliability structure of system 2 (Fig. 6) is mapped through serially connecting two subsystems P2.1 and P2.2:

- subsystem P2.1 consists of a total of 132 bogies of the 4RS/N type for 1435 and 1520 mm tracks (element 2.1), which make up a reliability structure with sliding reserve with the order of redundancy k = 10. It means that for 120 bogies of the 4RS/N type an operational reserve is presumed comprising 12 elements each of which can replace any basic bogie in case of its failure. Like in subsystems P1.1 and P1.2 in system 1, it is presumed that the time during which a failed bogie is replaced with a reserve element is equal to zero;
- subsystem P2.2 comprises one track gauge change stand (element 2.2).

Denotations in Figure 6:

P2.1, P2.2 – subsystems of system 2,

2.1_i, 2.2 – elements comprised in system 2 (2.1–4RS/N wagon bogies with self-adjusted wheel sets, 2.2 – rail gauge change stand).

3.4. Reliability ratios applied in the analysis

Rail gauge change systems may be analysed at various complexity levels. With respect to an element, a subsystem and a system, the relevant reliability ratios relating to reliability, availability and maintainability were assigned, including:

Table 2. Required number of wagon bogies

Parameter / System	Wagon turnover [days]	Demand for wagons [pcs]	Required number of wagon bogies [pcs]	Annual wagon mileage [km/year]
System 1	10.6	80	160 (for 1435 track) 160 (for 1520 track)	75,456.0 km
System 2	8.0	60	120	100,521.0 km

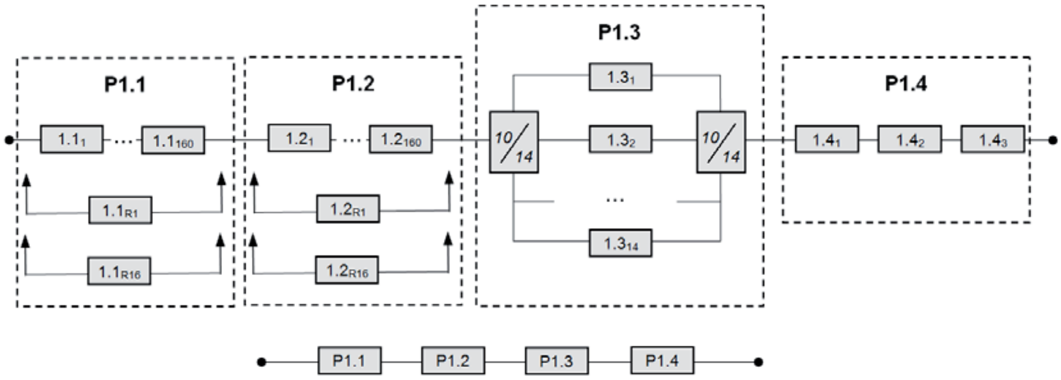


Fig. 5. Reliability structure of system 1

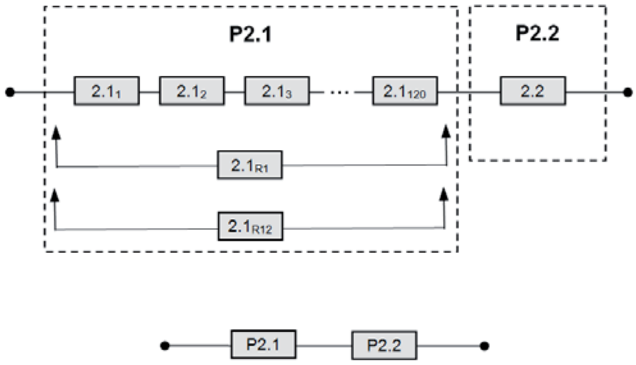


Fig. 6. Reliability structure of system 2

- Cumulative distribution of time of operation until the first failure *F(t),*
- Intensity of failure stream *z(t),*
- Renewal function in the maintenance cycle *H(t),*
- Mean Time To Failure, *MTTF,*
- Mean Time Between Failures $MTBF_k$,
- Operational Availability Ratio A_o,
- Technical Availability Ratio *A,*
- Mean Accumulated Down Time *MADT,*
- Cumulative distribution of restoration *G(t),*
- Mean Time To Restoration $MTTR_B$,
- Mean Time To Maintenance (Periodical Inspection) $MTTM_P$,
- Mean Time To Maintenance (Revision) $MTTM_N$.

The definitions and denotations of the ratios follow the PN-EN 50126 and the PN-EN 61703 standards [19, 20], and the calculations use the capacities of the following packages: Statistica, MiniTab and BlockSim. Owing to the wide range of the analyses, points 3.4.1 ÷ 3.4.3 present only the calculations of selected ratios used to compare the analysed systems.

3.4.1. Reliability ratios

In order to compare system reliabilities, the ratio of the mean number of failures (*MNF*) in a year of operation was applied, which, for a single element, is defined as follows:

$$MNF_i = \left(\frac{H_i(t)}{T_i}\right)\cdot 8{,}760.0 \quad \left[\text{failures}/\text{year}\right] \tag{1}$$

where:

MNF_i – mean number of failures of element “i” in a year of operation,

$H_i(t)$ – renewal function of element “i” in the maintenance cycle,

T_i – duration of operation of element “i” in the maintenance cycle (in hrs).

In the formula above, the renewal function *H(t)* is applied which, assuming that the duration of renewal is negligibly short compared with the duration of correct operation of the item, expresses the expected number of renewals equal to the number of failures until time *t* and is defined as follows [8]:

$$H(t) = \sum_{n=1}^{\infty} F_n(t) \tag{2}$$

where:

$F_n(t)$ – distribution function of the object’s operation until the occurrence of the n-th failure (renewal)

For system 1, the mean number of failures in a year of operation (MNF_1) is the total of failures of four subsystems:

$$\begin{aligned} MNF_1 &= MNF_{P1.1} + MNF_{P1.2} + MNF_{P1.3} + MNF_{P1.4} = \\ &= \left(\frac{H_{P1.1}(t)}{T_{1.1}} + \frac{H_{P1.2}(t)}{T_{1.2}} + \frac{H_{P1.3}(t)}{T_{1.3}} + \frac{H_{P1.4}(t)}{T_{1.4}}\right)\cdot 8{,}760.0 \quad \left[\text{failures}/\text{year}\right] \end{aligned} \tag{3}$$

where:

$H_{P1.1}(t) \div H_{P1.4}(t)$ – renewal functions of subsystems P1.1÷P1.4,

$T_{1.1} \div T_{1.4}$ – time of operation of subsystems P1.1÷P1.4 in the maintenance cycle (in hrs).

With the numerousness of 2XTa and 18-100 bogies together with the reserve elements taken into account, the renewal function of subsystems P1.1 and P1.2 is:

$$H_{P1.1}(t) = H_{P1.2}(t) = \sum_{i=1}^{176} H_{1.1}(t) = \sum_{i=1}^{176} H_{1.2}(t), \quad \textit{for } 0 < t \le 34{,}960 \text{ [hrs]} \tag{4}$$

where:

$H_{1.1}(t)$, $H_{1.2}(t)$ – renewal function of elements 1.1 and 1.2 (2XTa and 18-100 bogie) in the maintenance cycle:

$$H_{1.1}(t)=H_{1.2}(t)=\sum_{n=1}^{\infty}F_n(\mathrm{t})=\mathrm{F_1(t)}+\mathrm{F_2(t)}=$$
$$=\left[0,5\left(1+\Phi\left(\frac{t-18104}{6887\sqrt{2}}\right)\right)\right]+\left[0,5\left(1+\Phi\left(\frac{Ln(t)-11.269}{0.7822\sqrt{2}}\right)\right)\right]\quad for\ 0\le t\le 34,894\ \text{[hrs]} \tag{5}$$

where:

$\Phi(z)=\frac{2}{\sqrt{\Pi}}\int_0^z \exp\left(-t^2\right)dt$ – Gauss distribution function

With the numerousness ofthe bogie exchange stands (element 1.3) taken into account,the renewal function of subsystem P1.3 is:

$$H_{P1.3}(t)=\sum_{i=1}^{14}H_{1.3}(t),\quad for\ 0\le t\le 4.218\ \text{[hrs]} \tag{6}$$

where:

$H_{1.3}$(t) – renewal function of element 1.3 in the maintenance cycle:

$$H_{1.3}(t)=\sum_{n=1}^{\infty}F_n(\mathrm{t})=\mathrm{F_1(t)}+\mathrm{F_2(t)}+\mathrm{F_3(t)}+\mathrm{F_4(t)}=\left[1-1090.06\times10^{-3}\cdot\Gamma\left(1.203;\frac{t}{1441}\right)\right]+\left[1-\exp\left(-4.0973\times10^{-7}\cdot t^{1,7056}\right)\right]+$$
$$+\left[\frac{1}{1+\exp\left(-0.00122\cdot(t-5462.92)\right)}\right]+\left[0.5\left(1+\Phi\left(\frac{t-7382.96}{2102.68\sqrt{2}}\right)\right)\right]\quad for\ 0\le t\le 4.218\ \text{[hrs]} \tag{7}$$

where:

$\Gamma\left(1,203;\frac{t}{1441}\right)=\int_{t/1441}^{+\infty}t^{0,203}\cdot e^{-t}dt$ – incomplete gamma function

$\Phi(z)=\frac{2}{\sqrt{\Pi}}\int_0^z \exp\left(-t^2\right)dt$ – Gauss distribution function

With the numerousness of gantry cranes (element 1.4) taken into account, the renewal function of subsystem P1.4 is:

$$H_{P1.4}(t)=\sum_{i=1}^{3}H_{1.4}(t),\quad for\ 0\le t\le 8.504\ \text{[hrs]} \tag{8}$$

where:

$H_{1.4}$(t) – renewal function of element 1.4 in the maintenance cycle:

$$H_{1.4}(t)=\sum_{n=1}^{\infty}F_n(\mathrm{t})=\mathrm{F_1(t)}+\mathrm{F_2(t)}+\mathrm{F_3(t)}+\mathrm{F_4(t)}+\mathrm{F_5(t)}+\mathrm{F_6(t)}+\mathrm{F_7(t)}=$$
$$=\left[1-\exp\left(\text{-}2.878\times10^{-4}\cdot t^{1,104}\right)\right]+\left[1-9242.48\times10^{-4}\cdot\Gamma\left(2.166;\frac{t}{1465}\right)\right]+$$
$$+\left[0.5\left(1+\Phi\left(\frac{Ln(t)-8.231}{0.5238\sqrt{2}}\right)\right)\right]+\left[1-\exp\left(-1.866\times10^{-10}\cdot t^{2.5681}\right)\right]+\left[0.5\left(1+\Phi\left(\frac{t-6991.75}{2079.25\sqrt{2}}\right)\right)\right]+$$
$$+\left[0.5\left(1+\Phi\left(\frac{Ln(t)-9.0647}{0.4212\sqrt{2}}\right)\right)\right]+\left[1-\exp\left(\text{-}1.393\times10^{-18}\cdot t^{4.4912}\right)\right]\quad for\ 0\le t\le 8.504\ \text{[hrs]} \tag{9}$$

Substituting (5), (7) and (9) to relation (3), we get the mean number of failures in a year of operation of system 1:

$$MNF_1=\left(\frac{201.2}{35,040.0}+\frac{201.2}{35,040.0}+\frac{22.9}{8,760.0}+\frac{16.6}{17,520.0}\right)\cdot 8,760.0=131.8\quad\left[\text{failures}/\text{year}\right] \tag{10}$$

For system 2, the mean number of failures in a year of operation (MNF_2), relates solely to the failures of 4RS/N bogies and is:

$$MNF_2 = MNF_{P2.1} = \frac{H_{P2.1}(t)}{T_{2.1}} \cdot 8,760.0 \quad \left[\text{failures}/\text{year}\right] \tag{11}$$

where:
$H_{P2.1}(t)$ – renewal function of subsystem P2.1,
$T_{2.1}$ – duration of operation of subsystem P2.1 in the maintenance cycle (in hrs).

With the numerousness of 4RS/N bogies (element 2.1) together with the reserve bogies taken into account, the renewal function of subsystem P2.1 is:

$$H_{P2.1}(t) = \sum_{i=1}^{132} H_{2.1}(t) \quad for\ 0 \le t \le 8,140 \text{ [hrs]} \tag{12}$$

where:
$H_{2.1}(t)$ – renewal function of element 2.1 (4RS/N bogie) in the maintenance cycle:

$$\begin{gathered} H_{2,1}(t) = \sum_{n=1}^{\infty} F_n(t) = F_1(t) + F_2(t) + F_3(t) = \left[1 - \exp(-10.9 \times 10^{-4} \cdot t)\right] + \\ + \left[1 - \exp(16.16 \times 10^{-5} \cdot t)\right] + \left[0.5\left(1 + \Phi\left(\frac{t - 21520.9}{10734.2\sqrt{2}}\right)\right)\right], \\ for\ 0 \le t \le 31,191 \text{[hrs]} \end{gathered} \tag{13}$$

where:

$$\Phi(z) = \frac{2}{\sqrt{\Pi}} \int_0^z \exp\left(-t^2\right) dt \quad \text{– Gauss distribution function}$$

Hence, the mean number of failures in a year of operation of system 2 is:

$$MNF_2 = \frac{H_{P2.1}(t)}{T_{2.1}} = \left(\frac{370.9}{8760}\right) \cdot 8,760.0 = 370.9 \quad \left[\text{failures}/\text{year}\right] \tag{14}$$

3.4.2. Assessment of technical availability

Assuming that all elements operating in a system are described with identical probability distribution functions of the duration of operation and the duration of restoration, the system availability *A(t)* can be described with the following function [8]:

$$A(t) = 1 - F(t) + \int_0^t \left[1 - F(t - \tau)\right] h(\tau) d\tau \tag{15}$$

where:

h(t) – renewal density function: $h(t) = \frac{H(t)}{dt}$

The above formula is rarely used in practice due to a considerable degree of complexity of the calculations. Usually, the so-called technical availability ratio is applied, defined as the mean proportion of the time in which the system under consideration is available for use [12]:

$$A(\infty) = \lim_{t \to \infty} A(t) \tag{16}$$

In order to compare the technical availabilities of the analysed systems, the technical availability ratio *A* and the Mean Accumulated Down Time (*MADT*) were applied. The ratio *A* presents the mean technical availability in the maintenance cycle, between the subsequent maintenance (revision) activities of the analysed items. For an individual item, the availability ratio is defined as follows:

$$A_i = \frac{TZ_i}{TZ_i + TN_i + TO_i} \tag{17}$$

where:
TZ_i – mean time of availability of item "i" (in hrs),
TN_i – mean time of unavailability of item "i" due to corrective maintenance (in hrs),
TO_i – mean time of unavailability of item "i" due to preventive maintenance activities (in hrs).

The Mean Accumulated Down Time (*MADT*), in turn, was defined as follows:

$$MADT_i = 8,760.0 \cdot (1 - A_i) \quad \text{[hrs/year]} \tag{18}$$

For system 1, the availability ratio is the product of the availability of four subsystems, P1.1, P1.2, P1.3 and P1.4:

$$A_1 = A_{P1.1} \cdot A_{P1.2} \cdot A_{P1.3} \cdot A_{P1.4} \tag{19}$$

For subsystems P1.1 and P1.2, account being taken of the numerousness of the bogies and the reliability structure with sliding reserve [18]:

$$A_{P1.1} = A_{P1.2} = \prod_{i=1}^{n}\left(1 - \left(1 - A_{1.1}\right)^{k+i}\right) = \prod_{i=1}^{160}\left(1 - \left(1 - 0.9991\right)^{16+i}\right) \approx 1.0 \tag{20}$$

where:
n – number of basic bogies in the structure of subsystem P1.1 (P1.2),
k – number of reserve bogies in the structure of subsystem P1.1 (P1.2),
$A_{1.1}$ – technical availabilityratio for 2XTa bogie (18-100).

For subsystem P1.3, account being taken of the numerousness of the stands and the threshold reliability structure of the 10 out of 14 type [37]:

$$A_{P1.3} = \sum_{i=k}^{n} \binom{n}{i} A_{1.3}^{i} \left(1 - A_{1.3}\right)^{n-i} = \sum_{i=10}^{14} \binom{14}{i} 0.9710^{i} \left(1 - 0.9710\right)^{14-i} = 0.9999 \tag{21}$$

where:
n – number of all bogie exchange stands in the structure of P1.3 subsystem,
k – required number of stands necessary for correct operation of subsystem P1.3,
$A_{1.3}$ – technical availability ratio of the bogie exchange stand.

For the P1.4 subsystem, account being taken of the numerousness of the gantry cranes and the serial reliability structure:

$$A_{P1.4} = \prod_{i=1}^{n} A_{1.4i} = (0.9615)^3 = 0.8889 \tag{22}$$

where:

n – number of gantry cranes in the structure of subsystem P1.4,

$A_{1.4}$ – technical availability ratio for the gantry crane.

Substituting (20), (21) and (22) to relation (19), we get the technical availability ratio for system 1:

$$A_1 = A_{P1.1} \cdot A_{P1.2} \cdot A_{P1.3} \cdot A_{P1.4} = 1.0 \cdot 1.0 \cdot 0.9999 \cdot 0.8889 = 0.8888 \quad (23)$$

In turn, the accumulated down time in a year of operation of system 1 is:

$$MADT_1 = 8{,}760.0 \cdot (1 - A_1) = 8{,}760.0 \cdot (1 - 0.8888) = 973.8 \text{ [hrs/year]} \quad (24)$$

For system 2, the technical availability ratio is the product of the availabilities of two subsystems P2.1 and P2.2:

$$A_2 = A_{P2.1} \cdot A_{P2.2} \quad (25)$$

For subsystem P2.1, account being taken of the numerousness of 4RS/N bogies and the reliability structure with sliding reserve, the ratio is:

$$A_{P2.1} = \prod_{i=1}^{n}\left(1-\left(1-A_{2.1}\right)^{k+i}\right) = \prod_{i=1}^{120}\left(1-\left(1-0.9954\right)^{12+i}\right) \approx 1.0 \quad (26)$$

where:

n – number of basic bogies in the structure of subsystem 2.1,

k – number of reserve bogies in the structure of subsystem 2.1,

$A_{2.1}$ – technical availability ratio for 4RS/N bogie.

For P2.2 subsystem which consists of one rail gauge change stand, the technical availability ratio equals the availability of element 2.2:

$$A_{P2.2} = A_{2.2} = 0.9977 \quad (27)$$

Substituting (26) and (27) to relation (25), we get the availability ratio for system 2:

$$A_2 = A_{2.1} \cdot A_{2.2} = 1.0 \cdot 0.9977 = 0.9977 \quad (28)$$

The accumulated down time of system 2 is:

$$MADT_2 = 8{,}760.0 \cdot (1 - A_2) = 8{,}760.0 \cdot (1 - 0.9977) = 20.2 \text{ [hrs/year]} \quad (29)$$

3.4.3. Maintainability ratios

In order to compare the systems' maintainabilities, the mean maintenance time (*MMT*) in a year of operation was applied, which includes the total time spent on corrective and preventive maintenance of the system. For a single element, this ratio is defined as follows:

$$MMT_i = \left(\frac{\left(H_i(t) \cdot MTTR_{Bi}\right) + \left(\mathrm{NPMA_{Pi}} \cdot \mathrm{MTTR_{Pi}}\right) + \left(\mathrm{NPMA_{Ni}} \cdot \mathrm{MTTR_{Ni}}\right)}{T_i}\right) \cdot 8{,}760.0 \quad \left[\mathrm{hrs}/\mathrm{year}\right] \quad (30)$$

where:

$H_i(t)$ – renewal function of element "i" in the maintenance cycle,

$MTTR_{Bi}$ – mean time to restore element "i" (in hrs),

$NPMA_{Pi}$ – number of periodic maintenance activities on element "i" in the maintenance cycle,

$MTTM_{Pi}$ – mean time to maintain (periodical inspection) of element "i" (in hrs),

$NPMA_{Ni}$ – number of revision maintenance activities of element "i" in the maintenance cycle,

$MTTM_{Ni}$ – mean time to maintain (revision) on element "i" (in hrs),

T_i – time of operationof element "i" in the maintenance cycle (in hrs).

For system 1, the mean maintenance time (MMT_1) in a year of operation is the total of maintenance times of four subsystems and is:

$$MMT_1 = MMT_{P1.1} + MMT_{P1.2} + MMT_{P1.3} + MMT_{P1.4} \quad \left[\mathrm{hrs}/\mathrm{year}\right] \quad (31)$$

With the numerousness of 2XTa and 18-100 bogies taken into account, the mean maintenance time in a year of operation for subsystems P1.1 and P1.2 is:

$$MMT_{P1.1} = MMT_{P1.2} = \left[176 \cdot \left(\frac{(1.1431 \cdot 5.9) + (1 \cdot 6.0) + (1 \cdot 19.5)}{35{,}040.0}\right)\right] \cdot 8{,}760 = 1{,}418.9 \quad \left[\mathrm{hrs}/\mathrm{year}\right] \quad (32)$$

With the numerousness of the bogie exchange stands taken into account, the mean maintenance time in a year of operation for subsystem P1.3 is:

$$MMT_{P1.3} = \left[14 \cdot \left(\frac{(1.634 \cdot 11.2) + (12 \cdot 6.5) + (1 \cdot 29.6)}{8{,}760.0}\right)\right] \cdot 8{,}760.0 = 1{,}762.6 \quad \left[\mathrm{hrs}/\mathrm{year}\right] \quad (33)$$

With the numerousness of the gantry cranes taken into account, the mean maintenance time in a year of operation for subsystem P1.4 is:

$$MMT_{P1.4} = \left[3 \cdot \left(\frac{(5.5263 \cdot 9.3) + (23 \cdot 11.0) + (1 \cdot 36.0)}{17,520.0} \right) \right] \cdot 8,760.0 = 573.1 \quad \left[\mathrm{hrs}/_{\mathrm{year}} \right] \tag{34}$$

Substituting (32), (33) and (34) to relation (31), we get:

$$\mathrm{M}M\mathrm{T}_1 = MM\mathrm{T}_{P1.1} + MM\mathrm{T}_{P1.2} + MM\mathrm{T}_{P1.3} + MM\mathrm{T}_{P1.4} = 5,137.2 \quad \left[\mathrm{hrs}/_{\mathrm{year}} \right] \tag{35}$$

For system 2, the mean maintenance time (MMT_2) in a year of operation amounts to the total of mean maintenance times of two subsystems and is:

$$\mathrm{M}M\mathrm{T}_2 = MM\mathrm{T}_{P2.1} + MM\mathrm{T}_{P2.2} \qquad \left[\mathrm{hrs}/_{\mathrm{year}} \right] \tag{36}$$

With the numerousness of 4RS/N bogies taken into account, the mean maintenance time in a year of operation for subsystem P2.1 is:

$$MM\mathrm{T}_{P2.1} = \left[132 \cdot \left(\frac{(8.4291 \cdot 7.9) + (3 \cdot 7.0) + (1 \cdot 25.2)}{35.040.0} \right) \right] \cdot 8,760.0 = 3,722.1 \quad \left[\mathrm{hrs}/_{\mathrm{year}} \right] \tag{37}$$

For subsystem P2.2, on the other hand:

$$MM\mathrm{T}_{P2.2} = \left[1 \cdot \left(\frac{(2 \cdot 5.0)}{4,380.0} \right) \right] \cdot 8,760.0 = 20.0 \quad \left[\mathrm{hrs}/_{\mathrm{year}} \right] \tag{38}$$

Substituting (37) and (38) to relation (36), we get the mean maintenance time in a year of operation for system 2:

$$\mathrm{M}M\mathrm{T}_2 = MM\mathrm{T}_{P2.1} + MM\mathrm{T}_{P2.2} = 3,742.1 \quad \left[\mathrm{hrs}/_{\mathrm{year}} \right] \tag{39}$$

4. Conclusions

The subject of the paper was a comparative assessment of the reliability, availability and maintainability of two rail gauge change systems: wagon bogie exchange and SUW 2000 self-adjusted wheel sets. Analysing the results of calculations, one should take into account the fact that supervised operation of the SUW 2000 system concerned a prototype solution. During the tests of the solution, a number of failures occurred which were caused by constructional errors to be eliminated in the new upgraded version of the gauge change system. The following observations result from the analysis:

– The SUW 2000 system is characterised by a higher rate of failures compared with the wagon bogie exchange system. The mean number of its failures in a year of operation (MNF) is more than twice as high. This is due to failures of 4RS/N bogies, in particular flat spots on the wheels' rolling surfaces which were most frequent to occur during supervised operation;
– The SUW 2000 system is characterised by a higher technical availability ratio *(A)* and more than 40-times shorter Mean Accumulated Down Time (MADT) compared with the wagon bogie exchange system. The calculations demonstrate, however, that in order to ensure high availabilityof the system in actual operation, it is necessary to have an at least 10-percent operation reserve for bogies with gauge change systems;
– As regards maintainability, thanks to the replacement of the extended and costly technical equipment of the wagon bogie exchange point (Kutruff lifts, gantry cranes and other) with reliable highly available rail gauge change stands, the mean maintenance time (MMT) in a year of operation for the SUW 2000 system is nearly 30% shorter compared with the bogie exchange system.

The analysis, based on reliable data from supervised operation, demonstrated that in future the SUW 2000 system may be an alternative method for overcoming the barrier of different track gauges in the transport of hazardous materials compared with the wagon bogie exchange now in place. The next stage of the work to assess the possibilities of applying the SUW 2000 system should be an assessment of performance with the use of Life Cycle Costs (LCC).

References

1. Adamkiewicz W, Hempel L, Podsiadło A, Śliwiński R. Badania i ocena niezawodności maszyny w systemie transportowym. WKiŁ, Warszawa 1983.
2. Analiza organizacyjno-ekonomiczna wariantów przewozów produktów naftowych na kolejach o różnej szerokości torów. Projekt badawczy nr 186ZS. Politechnika Krakowska, Instytut Pojazdów Szynowych, Kraków 1995.
3. Automatic Gauge Changeover Systems. Project of the International Union of Railways UIC 2006-2013 (available from:http://www.uic.org/spip.php?article2927).

4. Basiewicz T, Gołaszewski A, Towpik K. Nowa technologia na kolejach o różnej szerokości torów. Problemy Kolejnictwa 2007; 144: 5-19.
5. Diomin J, Romanowska K. Techniczne problemy przewozów kolejowych Wschód – Zachód. Przegląd Komunikacyjny 2008; 6: 3-7.
6. DoD Guide for Achieving Reliability, Availability and Maintainability, Department of Defense. United States of America, Washington DC, 2005.
7. Economic study into investment in an automatic rail gauge change system within Pan-corridor 1. Project of the European Eureka Initiative: E!2353 RAIL GAUGE CHANGE, Centrum Naukowo-Techniczne Kolejnictwa, Warszawa 2003.
8. Gniedenko B W, Bielajew J K, Sołowiew A D. Metody matematyczne w teorii niezawodności. Wydawnictwa Naukowo-Techniczne, Warszawa 1968.
9. Herder P M, van Luijk J A, Bruijnooge J. Industrial application of RAM modeling development and implementation of a RAM simulation model for the Lexan plant at GE Industrial, Plastics. Reliability Engineering & System Safety 2008; 93: 501-508.
10. Interoperability, security and safety of goods movement with 1435 and 1520 (1524) mm track gauge railways: new technology in freight transport including hazardous products. Project under the 6th Framework Programme INTERGAUGE No. TST4-2005-516205, 2006-2008.
11. Jackson Y, Tabbagh P, Gibson P, Seglie E. The new Department of Defense (DoD) guide for achieving and assessing RAM. Reliability and Maintainability Symposium. Proceedings Annual IEEE 2005: 1-7.
12. Kopociński B. Zarys teorii odnowy i niezawodności. PWN, Warszawa, 1973.
13. Kwang-Woo Ch, Chul-Su K, Seung-Ho J. Technical Evaluation of Railway Transportation System with the Change of Gauge. Journal of the Korea Academia-Industrial CooperationSociety 2012; 13(5): 1954-1962.
14. Manzini R, Regattieri A, Pham H, Ferrari E. Maintenance for Industrial Systems. Springer, 2010.
15. Martorella S, Villanuevaa J F, Carlosa S, Nebota Y, Sanchezb A, Pitarcha J L, Serradell V. RAMS+C informed decision-making with application to multi-objective optimization of technical specifications and maintenance using genetic algorithms. Reliability Engineering & System Safety 2005; 87: 65-75.
16. Młynarski S, Oprzędkiewicz J. Systemowe rozwiązania zapewnienia bezpieczeństwa i niezawodności obiektów technicznych. ProblemyEksploatacji 2012; 3: 39-54.
17. O'Connor P. Practical Reliability Engineering, 4th Edition. John Wiley & Sons Ltd., 2010.
18. Oprzędkiewicz J. Podstawowe zagadnienia niezawodności narzędzi zmechanizowanych. ZeszytyNaukowe AGH, Kraków, 1972; 48(72).
19. PN-EN 50126 Railway applications - The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS)
20. PN-EN 61703 Mathematical expressions for reliability, availability, maintainability and maintenance support terms.
21. Projekt ujednoliconych wymagań dla kolejowych środków transportowych do przewozu materiałów płynnych po torach o szerokości 1435/1520 mm. Projekt badawczy nr 187ZS. Politechnika Krakowska, Instytut Pojazdów Szynowych, Kraków 1995.
22. Regmi M B, Hanaoka S. Assessment of intermodal transport corridors: Cases from North-East and Central Asia. Research in Transportation Business & Management 2012; 5: 27–37.
23. Review of developments in transport in Asia and the Pacific 2011. Report of The United Nations Economic and Social Commission for Asia and the Pacific (ESCAP) (available from:http://www.unescap.org/).
24. Sharma R K, Kumar D, Kumar P. Performance modeling in critical engineering systems using RAM analysis. Reliability Engineering & System Safety 2008; 93: 891-897.
25. Sharma S P, Kumar D. RAM analysis of repairable industrial systems utilizing uncertain data. Applied Soft Computing 2010; 10(4): 1208-1221.
26. Smith D J. Reliability, Maintainability and Risk: Practical Methods for Engineers, 8th Edition. Butterworth-Heinemann, 2011.
27. Stukalina O, Dzhaleva-Chonkova A. Problems of rail connections between Ukraine and its neigbouring countries. Mechanics Transport Communications 2012; 10(3/2): 3.4-3.8.
28. Suwalski R M. System samoczynnej zmiany rozstawu kół pojazdów szynowych. Rozprawy Monografie. Uczelniane Wydawnictwo Naukowo-Techniczne AGH 2006; 154.
29. Szkoda M, Tulecki A. Decision models in effectiveness evaluation of Europe-Asia transportation systems. The 8-th World Congress on Railway Research WCRR 2008, Seul, Korea, Article number G.3.3.4.2.
30. Szkoda M. Organisational and economic analysis of the possibilities of applying the automatic wheel track change system in the East-West transport. ProblemyEksploatacji 2003; 2: 275-289.
31. Szkoda M. Organizational and economical analysis of East-West variants system with gauge change 1435/1520. Master's thesis No DTT 135/02-SM. Cracow University of Technology, Krakow 2002.
32. Szkoda M. Analysis of the 1435/1520 mm track gauge change systems in the aspects of reliability and efficiency. Problems of Maintenance of Sustainable Technological Systems. Committee on Machine Building, Maintenance Fundamentals Section, Polish Academy of Sciences 2010; 1: 190-202.
33. Szkoda M. The Method of Assessing the Durability and Reliability of Rail Gauge-changing Systems. Doctoral thesis. Cracow University of Technology, Krakow 2008.
34. Szybka J. Methodology for reliability estimation of systems with sliding reserve. Scientific Problems of Machines Operation and Maintenance 2010; 3(163): 65-73.
35. ten Wolde M, Ghobbar A A. Optimizing inspection intervals – Reliability and availability in terms of a cost model: A case study on railway carriers. Reliability Engineering & System Safety 2013; 114: 137-147.
36. Trans-Asian railway and the issue of break-of-gauge. Report of The United Nations Economic and Social Commission for Asia and the Pacific (ESCAP) (available from: http://www.unescap.org/ttdw/common/TIS/TAR/break_of_gauge.asp).
37. Ważyńska-Fiok K, Jaźwiński J. Niezawodność systemów technicznych. PWN, Warszawa, 1990.
38. Zespoły cięgłowo-zderzne kolejowych środków transportowych dla relacji przewozowych Wschód-Zachód. Projekt celowy KBN nr 9T12C 031 96C/2980. Politechnika Krakowska, Instytut Pojazdów Szynowych, Kraków 2000.
39. Zio E. Reliability engineering: Old problems and new challenges. Reliability Engineering & System Safety 2009; 94(2): 125-141.

40. Демин, ЮВ, ТерещакЮВ. Подвижной состав для международных перевозок пассажиров и грузов по направлениям Восток-Запад. Вiсник СНУ iм. В. Даля 2010; 5(147): 167-171.

Maciej SZKODA
Cracow University of Technology
Faculty of Mechanical Engineering
Al. Jana Pawla II 37, 31-864 Kraków, Poland
E-mail: maciek@m8.mech.pk.edu.pl

Article citation info:
IDZIASZEK Z, GRZESIK N. Object characteristics deterioration effect on task realizability – outline method of estimation and prognosis. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 433–440.

Zdzisław IDZIASZEK
Norbert GRZESIK

OBJECT CHARACTERISTICS DETERIORATION EFFECT ON TASK REALIZABILITY – OUTLINE METHOD OF ESTIMATION AND PROGNOSIS

ZARYS METODY OCENY TRWAŁOŚCI I NIEZAWODNOŚCI OBIEKTU Z UWZGLĘDNIENIEM CZYNNIKA LUDZKIEGO I PŁASZCZYZNY LICZB ZESPOLONYCH*

The article introduces the essence of a potential technical object work range. Vital issues connected with the modeling of an object work range deterioration which are influenced by the destructive processes derived from environment, operation and wear of the object, were discussed/described. Typical destructive processes were described and deterministic and probabilistic models which allow for evaluation and prognosis of an object durability were included in the description. An outline of the approach to object work range deterioration adopted by the authors was presented. An outline of an object condition models for evaluation and prognosis of its durability purposes including a complex issues of random influence of the many factors which affect changes in an object work range and influencing the quality of the performed tasks were shown. In the models including randomness, probabilistic tools/apparatus and fuzzy logic were adopted. This kind of approach in modeling the changes in object durability adopted by the authors aims at bringing the models of object durability change closer to operational reality and at the same time at better utilization of their potential work range while maintaining the assumed level of reliability/safety during operations.

Keywords: *object work range, reliability, durability, material consumption, modeling, calculus of probability, fuzzy logic, efficiency, destructive processes.*

W artykule dokonano wprowadzenia w istotę pojęcia tzw. potencjalnego zasobu pracy obiektu technicznego. Opisano istotne zagadnienia związane z modelowaniem zużywania zasobu pracy, na które wpływają procesy destrukcyjne od środowiska, użytkowania i obsług. Wskazano na typowe procesy destrukcyjne i dla nich przedstawiono modele deterministyczne i probabilistyczne umożliwiające ocenę oraz prognozę zużywania potencjalnego zasobu pracy obiektu dla przyjętego poziomu niezawodności lub trwałości obiektu.Przedstawiono też zarys realizowanego przez autorów podejścia w modelowaniu zużywania zasobu pracy obiektu.Pokazano zarys modeli stanu obiektu do oceny i prognozy jego trwałości z uwzględnieniem zagadnień losowego wpływu wielu czynników wpływających na zmianę zasobu pracy obiektu, a tym samym, na jakość realizowanych zadań. W modelach uwzgledniających losowość przyjęto aparat probabilistyczny oraz wykorzystano logikę rozmytą. Tak przyjęte przez autorów podejście w modelowaniu zmian niezawodności/trwałości obiektu, ma na celu lepsze przybliżenie do rzeczywistości eksploatacyjnej, a tym samym lepsze wykorzystanie ich potencjalnego zasobu pracy, przy zachowaniu założonego poziomu niezawodności/bezpieczeństwa w trakcie realizacji działania/uzyskania efektu. Na koniec pokazano nowatorskie na skalę światową podejście, pozwalające na łączenie w jednym modelu technicznych i nietechnicznych aspektów oceny i prognozy zmian jakości obiektów w eksploatacji poprzez wykorzystanie do tego celu płaszczyzny liczb zespolonych.

Słowa kluczowe: *Eksploatacja, zasób pracy obiektu, niezawodność, trwałość, zużycie, modelowanie, probabilistyka, logika rozmyta, efektywność, procesy destrukcyjne*

1. Introduction

For the object holder its work potential, during maintenance process (achieving particular aim) is very important factor. What is more object potential consumption during maintenance procedures, storage and waiting for the execution of the next task is also significant.

For precise defined condition of an object assignment, work potential resource determines its maximum durability achievement (maximum usage of work resource maximum durability)[1]. Object work resource consumption leads to (at the beginning usually in hidden way) object parameters deterioration (necessary during useful object maintenance process).

1 Durability is the ability to endure; it is object ability to maintain its material and structural property (taking into account the maintenance process including parts replacement) which allows to reliable work.

There are two main strategies dealing with a problem. The first one is implementation of the object reconditioning (refurbishment) process (totally or partially). The second strategy is object consumption effects acceptance and use that knowledge in the current assessment and further object work resource prediction. Both strategies need controlling/measuring/ monitoring of the object consumption process. Also need evaluation and prediction methods of the process influence on work resource decreasing and malfunction probability increasing. This approach is significant for objects/systems where reliability and safety are crucial factors.

What we perceive as our material world has one essential property: independent from whether a given technical object[2] is used or not, destructive processes take place in the object and change its properties. It means that the process which leads to diminishing the ob-

2 Further called object.

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

jects potential work range determined during design stage takes place constantly. For various objects (and within the objects for modules/assemblies) depending on environmental conditions change, changes in operational and maintenance intensity, as well as the quality of the object itself (material, constructional and technological) achieved during its development, both the rate and direction of destructive changes are different. Such situation generates various consequences affecting both the object, as well as us that is the operators or owners of the object. This variability can be really significant for the same class objects of the same utility e.g. service life, as well as for an individual object during its task performance measured in time, rotations/revolutions, work cycles, or kilometers. It can lead to more or less serious consequences such as underestimating the rate of the object wear, damage causing stopovers at work, failures generating financial losses, or catastrophes even [3]. The significance of the variability acquires additional meaning, especially when the variability leads to a catastrophe or significant financial or social losses. Predicting the consequences of these changes, preventing them and including them in planning your economic or social activities is one of the essential tasks in design and operation analysis of an object life cycle.

In various technical object definitions, an operator of the object is included or not. It seems indispensable to include an operator/user as an element of the object constituting the whole and analyses operator's capabilities and influence on the object reliability and durability. Especially that usually the objects are integral part of bigger systems, like maintenance systems where technical and human factors are actually connected[3]. Where it is proven/indicated that the capabilities of the operator are not sufficient to perform the tasks in a safe manner, systems replacing human/operator are applied and technical safety engineering deals with the problem [17].

Proper process models are designed for destructing development analysis of assumed (predicted) loads[4]. Variables minimalizing, assuming typical loads and deterministic models (like constant human factor) provide reality simplification, but can be only used in object stationary processes and object environment. In most cases, destructing process models should comply more complicated dependence, taking into account overloads (normal load level exceeded) [13].

Appropriate/adequate, correct model of destructive processes should include weak sides of the object and should be a clue for designers to include inherent object properties [4, 17] leading to minimizing the negative results when encountering excessive loads (e.g. by switching off the object or switching I protective alarm systems). To build such models it is necessary to apply mathematical tools/apparatus [22] which will enable among other things incorporating probability, using partial operational data (diagnostic systems); including influence on human factor process and will include limit values of the process. The issue is not simple. Attempts to implement the above have been undertaken in many works [7, 16, 23] however so far these have been attempts comprising detailed/separate cases of destructive processes, and they lacked comprising both the mechanical causes and the human factor, which can have vital influence on the rate of change (malfunctions, crashes).

Indeterminacy/uncertainty of the object data (where there are no statistical data) cause use fuzzy logic possible in those data evaluation process. There are existing examples of fuzzy logic use in structural reliability analysis, mechanical vibration components [6, 24], reliability improvements estimation during product development [25] and maintenance planning of cold plastic deformation tools [1]. Unfortunately, each of those models do not provide a full picture of object quality and information about all object work resources changing causes. That is why, authors, decided to develop a model and description method based on fuzzy logic theory, probabilistic calculation and the theory of complex numbers[5].

The essence of the model and method is use of:

– probabilistic modeling of changing parameters which decide about technical object evolution (concerned with object inherent properties) to evaluation and prediction of object quality[6],
– fuzzy logic theory (fuzzy inference) for changing parameters description concerns with maintenance organization, environmental condition and standard of use volatility,
– theory of complex numbers to final evaluation/prediction indicator description (quality/use of objects/systems work resources and technical/non-technical object changes influence analysis).

Authors inspiration became searching of more adequate models/methods of objects/systems quality evaluation and prediction which are required, especially in safety reports [4][7]. The main purpose of report [4 p. 8] is presentation that danger of serious failures is identified and all indispensable measures were made to eliminate malfunctions and its influence on people and environment safety. Moreover proper safety and reliability solutions are put into effect during designing, maintenance and conservation of every installation.

2. Outline of methods of assessment and prognosis of object properties deterioration influence on task performance capabilities

The complete model of object/system work resources changing process or transition in state of unable to work (especially malfunctions lead to failures and crashes), should take into consideration inherent and not inherent features.

Object/system work resources changes from inherent features are for example:

– linear or volumetric effects of material deterioration (usage and age);
– deregulation (resulting from vibrations and strikes);
– change in primary characteristic of the object/system after production process and maintenance implementation;
– changes in power supply parameters (electrical, hydraulic).

Object/system work resources changes from not inherent features depend on:

– change in load (as a result of task type change);
– change in operation and maintenance quality,
– change of the working environment, etc.;

what is caused by:

– constancy or inconstancy of usage norms;
– variability of the working agent used in the object (e.g. material parameters of turning-lathe machined parts or types of projectiles used in weapons, voltage and current value for mechatronic and digital devices;
– quality of operation (propriety of starting and shutting down, complying to the accepted usage proprieties);
– natural environment parameters variability gradients (temperature – magnitude and the gradient of change in time, humidity, dusting/sanding;
– artificially induced threats e.g. air defense reaction or surges in the mains caused by switching on and off of big receivers or power suppliers;
– the quality of maintenance (applied strategy of operation, personnel qualifications, diagnostic tools, compliance to and quality of the procedures, used materials).

3 People (operators/users, maintenance personnel, etc.) behave differently and it causes bringing in unreliability in correct object work changes in object work life.

4 In human factor contexts destructive loads could be inappropriate organization change, change of maintenance personnel training level, do not take into consideration changing environmental condition of personnel work which decreasing their work capability.

5 Authors have no knowledge about research concern using of theory of complex numbers in described matters.

6 Object characteristic properties are: reliability, durability, readiness, efficiency, safety, etc.

7 That kind of reports must be realized in Seveso factories [16].

Usually causes changing of not inherent features are defined for normal/typical conditions which are unrealistic and can be estimated only by experts.

Therefore inherent features are changing randomly and are described by many variables random functions. Important are only those which change can be used in diagnostics measurements and during maintenance procedures.

So the Authors main goal is recognition of elements (object) properties, that changes have significant influence on object/system features changes and cause object work variation. Requirement is necessary to record the changes during diagnostic maintenance and provide economically rational profit.

Record the features changes (with use of adequate evaluation and prediction methods) provides information used in decision-making process:

- its work range at the moment of diagnostic examination (its work capacity analysis),
- its residual durability/lifetime (for new objects its overall durability/stability),
- rate of deterioration of the object work range (change of residual durability) for the assumed often changeable operational conditions (change of work standards, operational and environmental conditions),
- when the object should be subjected to maintenance preventing damage (especially the damage leading to failures or crashes/catastrophes) that is to say maintenance which restores completely or partially original object properties,
- the relationship between object/elements properties deterioration (between object maintenance or replacement) and its operational efficiency and the losses generated in relation to a new object.

In conclusion the Authors search the object/system model with changing object parameters (changing because of object features lost depends on its inherent and non-inherent characteristic) on the input. Parameter presents performing tasks possible change, described by evaluation or prediction of an object rest of work resources or changing probability of failures (especially malfunctions lead to failures and crashes) should be on the output.

2.1. Models of processes relevant to object work range

Evaluation or prediction methods of influence object deterioration features on the possibility of tasks implementation request to completion of partial tasks listed below:

- modeling process concerns work resources and its decreasing causes,
- project of mathematical model used to calculate influence of object destructive changes on work resources with incomplete/random data which provides as the effect dependence of work resources change and selected diagnostic parameters,
- project of model transforming measuring and estimating results into hints as possible maintenance decisions,
- project of databases model which provides object data transferring automation process into accepted maintenance/management decisions.

Processes modeling and object maintenance data transferring model designing are presented in this publication. Fig.1 presents schematically representative processes impacting the work range of an object. Object properties, object utility (degree of task performance capability), work and environment load as well as the quality of maintenance and quality of parameters describing properties change in the function of work range deterioration and its influence on performed task efficiency were included.

A scheme, known from automation, of inputs and outputs analysis can be applied here. There are two basic types in the scheme:

I The object is treated as the black box.

II The object model is presented with the use of known mapping/imaging/representation e.g. its reliability, functional structure.

The first model is usually applied where we have no data concerning the internal structure of the object or the structure is so complex/numerous (e.g. a processor) that its analyzing according to the second type is either unattainable or too expensive. Difficulty in adopting this approach lies in proper selection of input and output parameters that is such parameters whose change reflects the factual change of the object properties which are of interest to us.

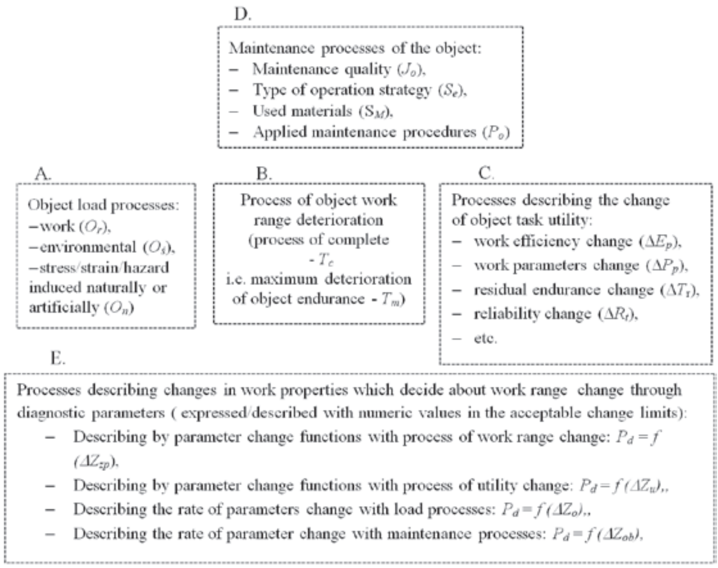


Fig. 1. Representative processes impacting the change of object work range ((ΔZ_{zp}*)*

The second model is used where the change in internal parameters cannot be observed through the analysis of inputs and outputs (the observed effects are stoppages and failures) and the lack of monitoring of the changes can lead to uncontrolled damages which can be the cause of an object failure as well as crashes. It is essential, in both models, to follow changes, transform input parameters into output parameters, steering the rate of output parameter change processes through limiting input/ interference changes.

2.1.1. General outline of the model I

A very general model of object work range change can be expressed by the relationship (1):

$$\Delta Z_{zp} = f\left(\Delta A;\Delta B;\Delta C;\Delta D;\Delta E\right) \tag{1}$$

in which:

ΔZ_{zp} – object work range change;

$f\left(\Delta A;\Delta B;\Delta C;\Delta D;\Delta E\right)$ – function transforming parameters in Fig.1 change to a change of object work range;

$\Delta A;\Delta B;\Delta C;\Delta D;\Delta E$ – Fig.1 parameters change.

Knowing the transformation function and the parameters change it is possible to follow the changes in the object work change. If, in an object population, parameters of change are known then based on that application of objects to different tasks can be predicted. This type of modeling does not allow for strategy realization in accordance with the objects technical condition but only better assessment of its life-time and better utilization of object work range while realizing the strategy of planned prevention.

2.1.2. General outline of model II

Modeling according to type II allows for the realization of operational strategy according to an object technical condition. In this model, mathematical models of Fig.1 processes were represented/ expressed as a set of relationships (2÷6) for (A, B, C, D, E).

$$A = f\left(O_r, O_f, O_n\right) \tag{2}$$

$$B = f(T_c, T_m) \tag{3}$$

$$C = f(\Delta E_p, \Delta P_p, \Delta T_r, \Delta R_t) \tag{4}$$

$$D = f(J_o, S_e, S_M, P_o) \tag{5}$$

$$E = f\left\{\left[P_d = f\left(\Delta Z_u\right)\right];\left[P_d = f\left(\Delta Z_o\right)\right];\left[P_d = f\left(\Delta Z_{ob}\right)\right]\right\} \tag{6}$$

The essence of object condition assessment and prognosis in these models is connecting measured object physical changes with the load causing the changes and the parameter describing the object work range changes e.g. the number of completed work cycles, mileage or object work time/period. Based on the changes and adopted acceptable limit values, life cycle of an object can be managed by introducing it into proper modes e.g. operation, servicing or withdrawal/ retirement/ change of application/ condemnation.

2.1.3. Summary

Modeling of this type is simplified through the adoption of the assumption that changes take place in deterministic unambiguous way and in homogeneous operational conditions and environment load and that all the factors impacting object work range changes are known. When the conditions are scarce or dominating conditions exist, such approach brings sufficient results. In other cases probabilistic model, which deals with random variables in the form of possible events distribution instead of events should be adopted.

2.2. Outline of the method employing probabilistic models

For complex objects (functional and consumption of the objects elements process complexity) state of the elements can cause object transition into different technical state and necessity of probabilistic calculation use for object description and search optimal maintenance strategy [11, 21, 22].

The outline of the method is presented based on the described general [22] and detailed [7, 23] models of objects such as aircraft gun, fast firing automatic cannons as well as operational systems of the object and methods of technical object management [18], audit, endurance and reliability assessment [14].

Of all the elements of object operational process, parameters characterizing them are singled out and their space-time composition is created. The essence of the structure research is defining the mutual relationship and acceptable limits of individual parameters in relation to others changes.

The presented main idea of mathematical modeling of technical object operational process assessment has been based on the following assumptions:

- each isolated element of operational process can be presented in the form of parameters set;
- there are many factors impacting individual parameters changes and none of them is dominating;
- changes in the values of the parameters adopted for the assessment assess the elements unambiguously by defining the brackets of acceptable change values in the process of their operation;
- there exists a result parameter which describes the given element in the operational process in an unambiguous manner.

Because of random character of the changes, a mathematical model which uses differential equation describing the dynamics of technical object condition change (7) [7, 22, 23], has been proposed

$$\frac{\partial U}{\partial N} = -b_1\frac{\partial U}{\partial Z_1} - b_2\frac{\partial U}{\partial Z_2} - \ldots - b_n\frac{\partial U}{\partial Z_n} + \frac{1}{2}\left(a_1\frac{\partial^2 U}{\partial Z_1^2} + a_2\frac{\partial^2 U}{\partial Z_2^2} + \ldots + a_n\frac{\partial^2 U}{\partial Z_n^2}\right) \tag{7}$$

where:

$$\begin{array}{ll} b_1 = \lambda_1 h_1 & a_1 = \lambda_1 h_1^2 \\ b_2 = \lambda_2 h_2 & a_2 = \lambda_2 h_2^2 \\ \vdots & \vdots \\ b_n = \lambda_n h_n & a_n = \lambda_n h_n^2 \end{array}$$

In (7) coefficient b1 means average individual parameters value increase in the work cycle unit e.g. firing and coefficients a1 mean average square of parameters value increase in the unit of firing. The solution of the problem has the form:

$$U\left(Z_1, Z_2, \ldots, Z_n; N\right) = \prod_{i=1}^{n} g_i\left(Z_i, b_i, a_i\right) \tag{8}$$

where:

$$g_i\left(Z_i, b_i, a_i\right) = \frac{1}{\sqrt{2\pi\ a_i N}}\ e^{-\frac{\left(Z_i - b_i N\right)^2}{2 a_i N}} \tag{9}$$

A practical solution can be offered by estimating the parameters of probability distribution with the use/application of e.g. credibility function. Thus for the newly introduced technical object, the final expression of its endurance is described by the formula [7, 23]:

$$N_i = \left(\frac{-\alpha_i \cdot \sqrt{a_i^*} + \sqrt{4 \cdot b_i^* + \alpha_i^2 \cdot a_i^*}}{2 \cdot b_i^*}\right)^2 \tag{10}$$

where: α_i – change limit value.

Using the essence of the model for a technical object, models precisely allowing for:

- rational planning of object maintenance activities in relation to the conditions of its operation,
- predicting stocking of spare parts in relation to operation conditions (maintenance and operation) [7, 23],
- prolonging the life-cycle of serviceable technical objects can be developed [7, 18, 23].

3.3. Outline of fuzzy inference system model

The method utilizing models of fuzzy logic has been developed on general models presented in works/papers [5, 9, 19, 20] and detailed models of objects [5, 15, 26] such as e.g. aircraft guns, operational systems of the objects as well as management methods, audit and evaluation of the technical objects efficiency [1, 6, 24, 25]. Using fuzzy logic, a model of fuzzy reasoning representing properties which are of interest to us can be developed. The basis for the model is the concept of information fuzzy coding. They function/operate on fuzzy sets instead of numbers, which allows for the generalization of the information. There are two basic models of fuzzy inference:

- non-adaptive inference (the parameters and structure of the model established in the design process remain unaltered during its operation);
- adaptive inference (the parameters and structure of the model established in the design process undergo changes during its operation/functioning).

Non-adaptive inference is simpler than the adaptive one but requires greater knowledge about the steered/managed object and can produce worse performance indicators.

Figure 2 presents the scheme of fuzzy inference system.

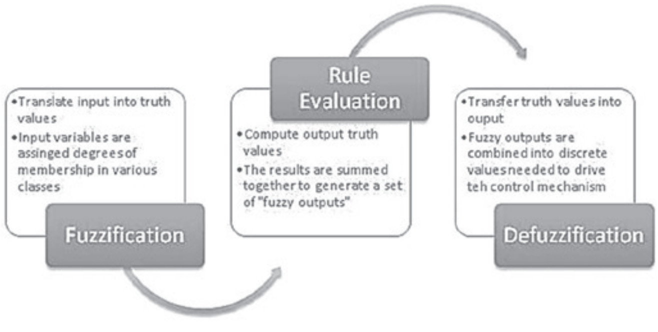


Fig. 2. Fuzzy inference system model

The model of fuzzy inference is based on three major blocks (fig. 3:

- Fuzzification block referred to as fuzzificator,
- Inference block with rules database,
- Defuzzification block referred to as defuzzificator.

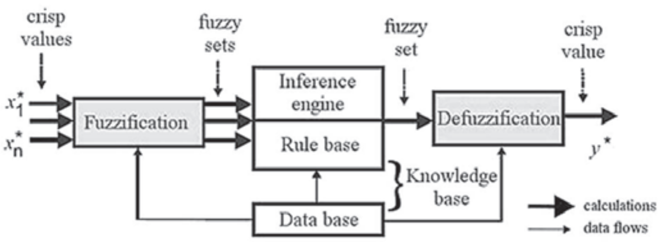


Fig.3. Fuzzy inference system with fuzzificator and defuzzificator blocks

Usually on the **fuzzificator** input (fig. 3) determined values are given/fed (crisp: $x_1 \div x_n$), which are transformed into fuzzy variables i.e. the numerical value of membership function is obtained, $\mu(x_A)$, $A \in \{1, N\}$ for $x_A \in X_A$. Calculated and given on the output, values of membership degree provide information about how high is the membership of input values in relation to individual fuzzy inputs sets.

Inference block contains/includes (fig.3):

- rules database (contains the main part of knowledge about the system being modeled, therefore the capability to design this part properly is essential);
- inference algorithms;
- variable membership functions and generates fuzzy set for variable y.

Result membership function often assumes a complex shape and it is calculated by inference which can be mathematically realized in many different ways.

Methods of inference can be derived from a number of sources:

- expertise (an expert based on his accumulated prior experience, defines modus operandi for individual cases, which may take place during the process – the expert's task then will be to design the inference rule itself as well as to select membership function for each individual case;
- qualitative model;
- automatic knowledge accessibility/extraction algorithms.

Inference based on expertise is predicate on knowledge and experience of a person familiar with the idiosyncrasy of the designed system. Here the explicit and tacit knowledge can be differentiated. The explicit one is characterized by the fact that it can be expressed verbally by the expert and thus transferred to another person. Tacit knowledge on the other hand cannot be formulated [5,19]. This knowledge is manifested during practical maintenance activities of a system (e.g. using aircraft weapons). By interviewing experts only formal part of knowledge about the system can be obtained from them in the form of verbal rules illustrating the input/output relationships of type:

$$\text{When } (x_1 \text{ is } A_n) \text{ and } (x_2 \text{ is } B_n) \text{ then } (y \text{ is } C_m), \tag{11}$$

where: x_1, x_2 – system inputs, y – output,
A_n, B_n, C_m – fuzzy sets applied in linguistic assessment of system inputs and outputs [5].

The example of the inference process realization (based on MODUS PONENS rule) is presented in the table 1.

Table 1. Inference process realization based on MODUS PONENS rule.

A	highly efficient aircraft armament
IMPLICATION	if highly efficient aircraft weapons are used then the probability of combat task execution increases
B	high probability of combat task execution

The set of verbally formulated rules defining the input/output relationship and the set of verbal information of linguistic values as used by an expert is called a verbal model. Verbal model is usually more modest than mental model as it does not include tacit knowledge about the system, which an expert is not able to transfer [5, 19]. The information flow taking place in the process of fuzzy linguistic system model creation is presented in Fig. 4.

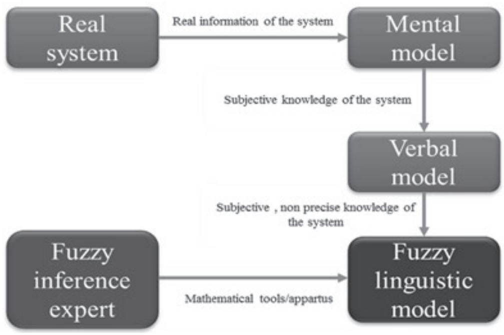


Fig.4. Process of creating fuzzy linguistic model of the realistic system

Result function in the defuzzificator (fig. 3) is converted into determined variables (defuzzification) y.

Among many defuzzification methods the most known ones are:

- *"Middle of Maximum"* – MOM,
- *"Smallest of Maximum"* – SOM,
- *"Largest of Maximum"* – LOM,
- *"Center of Gravity"* – COG,
- *"Center of Sums"* – COS,
- *"Height Method"* – HM.

Modeling of the type allows for the assessment and prediction of the objects condition in the situations when because of the lack of other possibilities we need to seek help in experts' opinions and especially the so called intuitive aspects of these opinions deriving more from the combination of their accumulated experience and inner intuition in the given field. In some situations it is the only method in some other it is the most efficient or the fastest method to assess and predict object work range deterioration for the preliminarily defined rules of the object operation, maintenance and given environmental conditions.

2.4. Object maintenance quality model with use of complex numbers

According to European data [4], the importance of limitations/threats comes from human factor is increasing in system designing processes. And that is why civilian and technical safety engineering starts developing.

Very important conclusion is provided in this publication: the theory about mathematical dependence between technical and non-technical aspects of object work resource consumption is required for further model analysis and research of object work resource effective use process with assumed/accepted/required level of reliability or durability.

To do so, Authors propose (the world innovation) using complex numbers theory [10] in object maintenance quality modeling. It consists of technical and non-technical maintenance object quality evaluation and change prediction connection.

Formula (12) describes generalized quality object model. Object ability parameter shows how its value, changing in time, affects object durability (T) and reliability (N), as two primary object ability state characteristics. Proposed parameter is complex number (the real part describes durability resource T and object material and technological features; the imaginary part describes reliability resource N and object features concerns human decisions (named „human factor").

$$Z_u = T - i N \tag{12}$$

where:

$Z_u = T + i N$ – generalized reliability-durability object model,

$T = \int_{x_p}^{x_d} \sum_{x=1}^{n} x_{pt}$ – object durability reserve,

x_{pt} – any diagnostic durability parameter,

x_p – initial value of diagnostic durability parameter,

x_d – acceptable value of diagnostic durability parameter.

$N = \int_{x_p}^{x_d} \sum_{x=1}^{n} x_{pn}$ – object reliability reserve, reliability redundancy when object meet the planned before expectations (or changed during),

x_{pn} – any diagnostic reliability parameter,

x_p – initial value of diagnostic reliability parameter,

x_d – acceptable value of diagnostic reliability parameter.

Therefore, maintenance factors, raw materials, environment, the pace and load of an object changes have influence on the real part (formula 12). Variability of human/operator competence, accepted maintenance strategy and maintenance/organizational procedures have influence on the imaginary part.

Changing object state during maintenance is natural, unavoidable process. The particular maintenance situation have only influence on dynamic changing state parameters (material and intellectual factors). So, the object must be seen as: technical object, maintenance situation, human resources and relation between them determining the object dynamic changing state.

Durability resource depends on:
- parameters acceptable changes of length intervals (initial durability resource),
- completion of recovery processes;

whereas its rate of decrease depends on:
- possibility of prophylactic service implementation,
- object life for its worse than designer predict condition ,
- payloads, environmental and materials changes.

While reliability resource mainly depends on:
- initial reliability resource,
- completion of recovery processes,
- reliability redundancy.

Reliability parameter can be analyzed in two aspects:
- work reliability for specific conditions depends on impact resistance and object counteract damage ability,
- reliability to meet operators expectations (expectations increasing and decreasing for new and used object – meeting operators different use expectations analysis),

Reliability depends on:
- preservation of diagnostic parameters in acceptable limits,
- preservation of required parameter values within the existing limits identified by designer during modernization process,
- completion of different expectations configuration and cooperation with other objects tasks,
- keeping price competitiveness with other same class objects,
- safety,
- risk (safety loss, costs prediction, profitable recovery, modernization etc.).

Object which is considered as able to use needs to have specific level of reliability and durability resource, if not the object will be withdraw from use.

Reliability-durability selected individual models:
- $Z_u = T$ which means that $iN = 0$; it means that object durability resource was expended or object expectations has been changed that object has no capabilities to meet the expectations despite having durability resource or it means that object is durable in all spectrum of use T or there is no possibility to have an effect on its parameters and use (e.g. autonomous system after operator control disengagement like Pershing missile);
- $Z_u = iN$ which means that $T = 0$; it means that object meets the durability expectations in the whole range of life and its output is in line with the designer.

When $N = 0$ in maintenance reality, it could mean that human decisions have no influence on object state (object is no serviceable, changing life standards etc.) which means that that reliability–durability model transformed into durability model:

$$T = \int_{x_p}^{x_d} \sum_{x=1}^{n} x_{pt} \tag{13}$$

where:

T – object durability reserve as sum of durability reserves of object individual elements described by x_{pd} – diagnostic durability parameters in their ability limitations (from x_p to x_d),

x_{pt} – any diagnostic durability parameter,

x_d – acceptable value of diagnostic durability parameter,

x_p – initial value of diagnostic durability parameter.

$T = 0$ in case, when all parameters reached a limiting state and there is no possibility to conduct renewing.

When $N = 1$ it means that object is reliable (object meets the expectations independently from human decisions – usually in specified time – so is assumed to maintenance by service life with no predicted servicing). It means, that reliability-durability model transformed into reliability model:

$$N = \int_{x_p}^{x_d} \sum_{x=1}^{n} x_{pn} \quad (14)$$

where:

$N = \int_{x_p}^{x_d} \sum_{x=1}^{n} x_{pn}$ or $N = \sum_{x=1}^{n} \left(x_d - x_p\right) \sum_{x=1}^{n} x_{pn}$ – object reliability reserve as sum of reliability reserves of object individual elements described by x_{pd} – diagnostic reliability parameters in their ability limitations (from x_p to x_d),

x_{pn} – any diagnostic reliability parameter,

x_d – acceptable value of diagnostic reliability parameter,

x_p – initial value of diagnostic reliability parameter.

That kind of modeling is clear to understand especially when particular cases are considered, like: situations, when an object is in one-piece and is unrecoverable and its ability depends on keeping diagnostic parameters in borders limited by designer. Any further decisions are not considered. Therefore prediction of object ability takes into account the technical, organizational and i management relations.

That kind of modeling allows to directly observe the changes of individual parameters on complex plane and durability and reliability optimization in view of any material and human parameter. It is important because, for the some parameters, change its value depends on load variation, which can be result of human factor or changes in the technologic or climate conditions. It is hard to determine which factor is the most important at the moment. However, we are able to continuously observe the changes if we consider individual decisions in general context. In that case use of complex numbers in maintenance changes process description, which do not lose technical and non-technical relations. What is more the description allows to observe and capture any maintenance relations. Single change of condition durability can be described as change results from adding ΔT and ΔN :

$$\Delta Z_u = \Delta T + i\Delta N \quad (15)$$

The sum of changes:

$$\sum \Delta Z_u = \sum \Delta T + i \sum \Delta N \quad (16)$$

Formula (12) after taking changes (16) into account is:

$$Z_u + \sum \Delta Z_u = T + \sum \Delta T + iN + \sum \Delta N \quad (17)$$

Therefore general parameter of the object ability Z_u taking changes into account for moment i is:

$$Z_{ui} = \frac{T + \sum \Delta T + iN + \sum \Delta N}{\sum \Delta Z_u} \quad (18)$$

2.4.1. Practical implication from (12) and (17) models

As a result of (12) i (17) models are very important, practical observations like:

- Two maintenance systems (or two maintenance states in the same system) are equal, when $\operatorname{Re} z_i = \operatorname{Re} z_j$ and $\operatorname{Im} z_i = \operatorname{Im} z_j$, or, when states concerning systems material parts are equal and at the same time states concerns elements come from human factor (decision-making) are equal.

Formula (12) allows:

- To evaluate and to predict, in generalized suitability indicator of the system analysis, the role of material part and human factor, and thereby if object maintenance system is equable (if we know what proportion of the real part and imaginary part should be for effective implementation of the maintenance process).
  - When we put two systems into one (two objects into one) we have clear view of system total rate, because new system or object addition could improve (deteriorate) the real part (material) and imaginary part of the rate as well. If we are interested in general profit, simple calculation of the profit or lost rate of the systems connection is possible. Moreover it gives us rate for the adequacy of the applied prevention evaluation to balance of the system, because if the imaginary part deteriorate increasing the real part will be pointless.
  - If the general parameter of the object ability Z_u combines in a relationship with transferring possibilities or probable corporation profit, the value of the parameter will present the potential of the corporation.
  - If we associate the imaginary part with corporation capability of market adaptation (intellectual capital) and the real part with new technologies and financial capital we can observe change of the potential and development of the corporation and its capability of taking on challenges in new markets, determining intellectual reserve to challenging of the new task or capital reserve to increase material production.
  - If we are capable to evaluate task (projects) needs by general parameter of the object ability Z_u then simple transformation of the rates in space[8] (C,+,) allows to analyze corporation ability to execute and searching the most effective ways of use corporation resources simulation (material and intellectual).

3. Summary

A proposed approach to modeling of object work range deterioration and especially to assessing the impact of properties change (as a result of work range deterioration) on task performance efficiency is a n attempt to include complex problem of many factors random influence which deteriorates object work range and their random influence on the quality of performed tasks. Adopting probabilistic tools/apparatus and fuzzy logic in modeling (at adopted model assumptions of an object) appears to be the right research direction when designing efficient and cost effective ways of solving problems of connecting variable factors with/of operation, maintenance, environment and

8 C - complex numbers space.

safety conditions in operational technical objects reality. It is assumed that models of the type allow for better approximation to operational reality and thereby better utilization of the object work range while maintaining the assumed level of their reliability/safety performance/effect achievement.

The problems presented in the paper do not exhaust the considered issue but only indicate the area of the planned by the authors scientific publications on this problem in the nearest time. Subsequent articles will present detailed developments in proposed methods and show their applications for e.g. comparing the obtained results, determining the ranges of a given method use, as well as their implementation in database systems to provide support for object administrators/commanders/owners in the decision making process.

References

1. Baban M, Baban CF, Blaga FS. Maintenance planing of cold plastic deformation tools using fuzzy logic. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2010; 3: 21-26.
2. Biegus A. Podstawy projektowania konstrukcji według PN-EN 1990. Materiały szkoleniowe. Poznań 2010.
3. Cuerden RW, Edwards MJ, Pittman MB. Effect of Vehicle Defects in The Road Accidents. Transport Research Laboratory Published Project Report (2011).
4. Fabbri L, Struckl M, Wood M. Guidance on the Preparation of a Safety Report to Meet the Requirements of Directive 96/82/EC as Amended by Directive 2003/105/EC (Seveso II). Opublikowane przez Wspólnotowe Centrum Badawcze (Joint Research Centre) European Communities. Luxemburg: Biuro Urzędowych Publikacji Wspólnot Europejskich 2005.
5. Grzesik N. Podstawy sterowania rozmytego. Projektowanie rozmytych systemów eksperckich w środowisku Matlab-Simulink. WSOSP. Dęblin 2012.
6. Huang H-Z. Structural reliability analysis using fuzzy sets theory. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2012; 4: 284-294.
7. Idziaszek Z. Zarys metody oceny trwałości szybkostrzelnych armat automatycznych wykorzystującej zmiany parametrów diagnostycznych zasadniczych zespołów. Zagadnienia Eksploatacji Maszyn 2004; Z.2: 97-109.
8. Jarosiński W. Periodic technical inspections of vehicles and road traffic safety with the number of road accidents involving fatalities. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (1): 105–111.
9. Kacprzyk J. Wieloetapowe sterowanie rozmyte. WNT. Warszawa 2001.
10. Leja F. Funkcje zespolone. Wydawnictwo naukowe PWN. Warszawa 2006. Wydanie 6.
11. Liu Y, Huang H-Z. Optimal replacement policy for multi-state system under imperfect maintenance. IEEE Transactions on Reliability 2010; 59.3: 483-495.
12. Miranda V. Fuzzy reliability analysis of power systems. 12-th PSCC. Dresden, August 1996.
13. Nowakowski T. Metodyka prognozowania niezawodności obiektów mechanicznych. Oficyna Wydawnicza Politechniki Wrocławskiej. Wrocław 1999.
14. Olearczuk E, Idziaszek Z. Audyt trwałości w eksploatacji szybkostrzelnych armat automatycznych z uwzględnieniem bezpieczeństwa. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2004; 3: 15-23.
15. Orkisz M, Grzesik N. Wykorzystanie teorii zbiorów rozmytych w lotniczych układach wspomagająco – decyzyjnych w celu zwiększenia efektywności wykonania zadania lotniczego. Journal of Aeronautica Integra 2009; 2: 9-23.
16. Paska J. Niezawodność systemów elektroenergetycznych. Oficyna Wydawnicza Politechniki Warszawskiej. Warszawa 2005.
17. Pichowicz W. Inżynieria bezpieczeństwa technicznego. Warszawa: Wydawnictwa Naukowo-Techniczne 2008.
18. Piechna S, Idziaszek Z. Teoretyczne podstawy oceny trwałością eksploatacji obiektów. Mechanik 2010; 7: 145-152.
19. Piegat A. Modelowanie i sterowanie rozmyte. Akademicka Oficyna Wydawnicza EXIT. Warszawa 2003.
20. Rutkowska D, Piliński M, Rutkowski L. Sieci neuronowe, algorytmy genetyczne i systemy rozmyte. PWN. Warszawa, Łódź 1997.
21. Soro IW, Nourelfath M, Aït-Kadi D. Performance evaluation of multi-state degraded systems with minimal repairs and imperfect preventive maintenance. Reliability Engineering & System Safety 2010, 95.2: 65-69.
22. Tomaszek H. Modelowanie procesów zużycia elementów mechanicznych urządzeń o obciążeniu impulsowym w aspekcie niezawodności. ITWL, Warszawa 1981.
23. Tomaszek H, Idziaszek Z. Zarys metody oceny trwałości luf działek lotniczych. Zagadnienia Eksploatacji Maszyn 2004; Z.1: 99-110.
24. Wu W, Huang H-Z, Wang Z-L, Li Y-F, Pang Y. Reliability analysis of mechanical vibration component using fuzzy sets theory. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2012; 2: 130-134.
25. Yadav OP, Singh N, Chinnam RB, Goel PS. A fuzzy logic based approach to reliability improvement estimation during product development. Reliability Engineering and System Safety 80 (2003) 63-74.
26. Żurek J, Grzesik N. Fuzzy expert aircraft onboard control systems assistant. Safety, Reliability and Risk Analysis: Beyond the Horizon. CRC Press/Balkema, Tylor & Francis Group (pp 250). ESREL 2013, Amsterdam, Holland, 29.09.÷02.10.2013.

Zdzisław IDZIASZEK
Mechatronics and Aviation Faculty
Military University of Technology, 49, ul. Kaliskiego 2, 00-908 Warsw 49, Poland

Norbert GRZESIK
Aviation Faculty
Polish Air Force Academy, ul. Dywizjonu 303 35, 08-521 Dęblin, Poland
E-mails: zidziaszek@wat.edu.pl, norbertgrzesik@op.pl

Article citation info:
DĄBROWSKI JR, KLEKOTKA M, SIDUN J. Fretting and fretting corrosion of 316L implantation steel in the oral cavity environment. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 441–446.

Jan Ryszard DĄBROWSKI
Marcin KLEKOTKA
Jarosław SIDUN

FRETTING AND FRETTING CORROSION OF 316L IMPLANTATION STEEL IN THE ORAL CAVITY ENVIRONMENT

FRETTING I FRETTING-KOROZJA STALI IMPLANTACYJNEJ 316L W ŚRODOWISKU JAMY USTNEJ*

Processes of mechanical destruction of implants, dental prosthetics elements, and orthodontic apparatus considerably limit their operating lifetime and the comfort of patients. Processes of destruction of kinematic joint elements caused by fretting and fretting corrosion processes are an important problem, albeit one that is not yet fully understood. This paper presents the results of fretting and fretting corrosion studies conducted on 316L implantation steel, which is used in dentistry, particularly in prosthetic and orthodontic applications. Tests were performed by means of an original device of the authors' own design, with the application of methodology developed by the authors. Fretting and corrosion tests were carried out in phosphate buffered saline (PBS) as well as in the presence of natural saliva and its substitutes. Own compositions of artificial saliva were developed for the purposes of studies. Observations of sample surfaces were performed using a scanning electron microscope (SEM) and a confocal microscope. Test results indicate a significant influence of fretting on the corrosion of 316L steel (fretting corrosion) as well as the important role of the studied fluids (saliva and its studies) in these processes. It was stated that the saliva substitute containing mucin III was characterized by the most favorable tribological characteristics. During fretting tests, intensive phenomena of materials conveyance into the friction contact area were observed.

Keywords: *tribology, fretting, fretting corrosion, artificial saliva, implants, dentistry.*

Procesy destrukcji metalicznych implantów, elementów protetyki stomatologicznej i aparatów ortodontycznych znacznie ograniczają ich trwałość eksploatacyjną i komfort pacjentów. Szczególnym zagadnieniem, aczkolwiek dalece niepoznanym, są procesy niszczenia elementów połączeń kinematycznych wywołane procesami frettingu i fretting – korozji. W pracy przedstawiono wyniki badań frettingu i fretting-korozji stali implantacyjnej 316L – używanej w stomatologii, szczególnie w zastosowaniach protetycznych i ortodontycznych. Badania realizowano za pomocą oryginalnego urządzenia własnej konstrukcji, z wykorzystaniem metodyki opracowanej przez autorów. Badania frettingu i korozji przeprowadzone zostały w buforze fosforanowym (PBS) jak również w obecności śliny naturalnej i jej substytutów. Na potrzeby badań opracowano własne kompozycje sztucznych ślin. Obserwacje powierzchni próbek prowadzone były z wykorzystaniem skaningowego mikroskopu elektronowego (SEM) oraz mikroskopu konfokalnego. Wyniki badań wskazują na znaczący wpływ frettingu na niszczenie korozyjne stali 316L (fretting-korozja), a także na istotną rolę badanych płynów (śliny i jej substytutów) w tych procesach. Stwierdzono, że najkorzystniejszymi charakterystykami tribologicznymi charakteryzował się substytut śliny zawierający mucynę III. W trakcie testów frettingu obserwowano intensywne zjawiska przenoszenia materiałów w strefie kontaktu tarciowego.

Słowa kluczowe: *tribologia, fretting, fretting-korozja, sztuczna ślina, implanty, stomatologia.*

1. Introduction

Fretting is defined as a complex process of relative, oscillating micro-displacements of contacting surfaces, as a result of which destruction of the surface layers of these contacting elements takes place [33, 34]. Adhesive damaging of surfaces and the formation of fatigue cracks causes the creation of wear particle as well as their subsequent oxidation and hardening. These products act as an abrasive and are broken up, and their amount increases until the contacting surfaces are separated by a layer of oxide particles and wear conditions are stabilized. Materials conveyance processes, which take place with intensive oxidation, are also observed in the contact area [19]. Four primary mechanisms are responsible for fretting wear processes: adhesion, fatigue, abrasion, and corrosion [20]. Depending on the type and direction of motion, fretting is divided into: tangential, radial, torsional, and rotational fretting [36–38]. Its main consequence is a drastic reduction of the durability and the operational reliability period of devices.

The phenomenon of fretting applies to most biomaterials, including metals, polymers, as well as ceramic materials [5, 14, 22, 37]. Advances in implantology impose the application of materials fulfilling ever-greater requirements concerning biofunctionality, with preservation of full biotolerance of implants in the human body [10, 25]. Metallic materials are widely used in dental prosthetics and surgery. They are used, among other things, to reconstruct or replace missing teeth and as elements of orthodontic apparatus for correction of malocclusions [2, 23]. The most commonly used materials are precious metals (gold, platinum, palladium), alloys of cobalt, titanium, and nickel, as well as austenitic steels [3, 5, 6, 11]. It should also be mentioned that, due to the toxic properties of some of these materials

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

(nickel, vanadium), new compositions of metallic alloys with similar characteristics but which cause less harm in the human body are sought for. Components made from implantation steels have found wide applications in dentistry and orthodontics because of their good technological and strength properties as well as their low production costs [25, 8]. AISI 316L and AISI 316LVM stainless steels are mainly deserving of attention. They contain molybdenum, which stabilizes the chromium compounds forming the passive layer on the surface of the alloy, thus increasing its resistance to pitting and crevice corrosion, which are responsible for the degradation of multi-component implant systems [15, 35].

Many works concerning fretting and fretting corrosion relate to the field of orthopedics and are focused on components of prostheses and bone fixations [12, 16, 30]. Only a few works pertain to processes occurring in the oral cavity [4, 24, 28, 30]. Processes of destruction of metal prosthetic elements (intraoral distractors of maxillofacial bones, dental implants, orthodontic apparatus) resulting from their fretting corrosion are still little known. There are known cases in which wear products of orthodontic elements (mainly ligatures) are adsorbed onto dental plaque, causing stains. A part of the wear products is sent to the digestive system, from which they pass through to the organism in the form of metal ions (iron, chromium, nickel), which exhibit toxic activity [10, 13, 29].

The influence of the environment is of large significance in processes of fretting and corrosion destruction. Its aggressiveness is largely caused by active chemical substances, particularly chlorine, sulfur, oxygen, and phosphorus compounds [9, 17]. Human saliva, which fulfills a series of protective and supporting functions, including lubrication functions, also has a large influence on corrosive and fretting wear processes in the oral cavity [7, 26, 27]. Persons suffering from abnormal salivary gland secretion (xerostomia) are exposed to the effects of excessive grinding down of teeth and are forced to use artificial preparations that act as a substitute for natural saliva. This is why more and more new preparations simulating the functions of human saliva are being developed and introduced onto the market. Studies in this scope have a high cognitive and utilitarian significance. The results of these studies can be used for better selection of biomaterials and saliva substitutes, oriented towards increasing the durability and extending the operating life of dental prosthetics components and, at the same time, improving the health and comfort of patients.

The goal of this paper was to study the processes of fretting and fretting corrosion of 316L implantation steel in the presence of human saliva and its substitutes – developed at the Department of Materials and Biomedical Engineering of Białystok Technical University. A special testing station and research methodology, as described further in the paper, were developed in order to achieve this goal.

2. Materials and testing methodology

316L implantation steel with the chemical composition presented in table 1 was subjected to tests.

Test samples were subjected to grinding operations in order to achieve a surface roughness on the order of Ra=0.4 μm. To eliminate impurities, samples were rinsed in ethanol and placed in a sonifier for 10 minutes before every test. Human saliva solutions acquired from a 25 year old male and its substitutes, developed at the Department of Materials and Biomedical Engineering of Białystok Technical University (table 2), were used in tests. All artificial saliva preparations were based on phosphate buffered saline (PBS) with pH=7 and had the chemical compositions presented in table 3.

Table 1. Chemical composition of 316L steel

Alloy	Alloying components (maximum content, % mass)					
316L steel	Fe	Cr	Ni	Mn	Mo	Si
	remainder	22,00	15,00	4,25	3,00	1,00
	Nb	N	Cu	C	P	S
	0,80	0,50	0,50	0,08	0,03	0,01

Table 2. Solutions used in tests

Designation	Composition
A	human saliva
B	phosphate buffered saline (PBS)
C	mucin III (2% mass) in PBS
D	mucin III (2% mass) + xanthan gum (0.35% mass) in PBS

Table 3. Chemical composition of PBS (content of ingredients in 1 dm^3 of water)

Ingredient	Content [g]
Sodium chloride (NaCl)	6,72
Disodium phosphate (Na_2HPO_4)	2,27
Monopotassium phosphate (KH_2PO_4)	1,11

Analysis of data from the literature and the results of preliminary studies conducted by the authors were the basis for selection of model saliva substitutes. The selection of mucins was related to the favorable tribological characteristics of this substance. Xanthan gum is a natural ingredient in many compositions destined for use in the oral cavity. Its addition enabled correction of the rheological properties of the developed compositions relative to natural saliva.

A diagram of the kit used for fretting and fretting corrosion tests is shown in fig. 1. Friction processes were performed at a small amplitude on the order of 100 μm, a frequency of 0.8 Hz, and with unit pressing forces of: 5, 15 and 30 MPa. The moving table of the device, on which disk-shaped samples with a diameter of 8 mm were fastened, was in reverse motion. A countersample in the shape of a truncated code was pressed to the surface of the disk, and the diameter of the contact surface was equal to 1.3 mm. Disks and cones were made from 316L steel.

Assessment of fretting wear was performed according to a method developed by the authors by measuring the volume of material decrement as well as of the material accumulated during the friction process (excess material). A LEXT OLS4000 with 3D imaging capability was used in these studies. It enabled accurate measurement of the surface and of the volumes of decrements and excess material. The applied method makes it possible to measure only areas of interest, the statistics of which are calculated in real time. An illustration of the developed method of fretting wear assessment is presented on the diagram in fig. 2.

Fretting corrosion tests were conducted at a pressure of 15 MPa. The tester was connected to a PGP201 potentiostat from the Radiometer company along with a tri-electrode system. The working electrode was the tested sample, and the reference electrode was a saturated calomel electrode (SCE). The auxiliary electrode was a platinum electrode with a surface of 128 mm^2.

Pitting corrosion resistance tests were carried out using the potentiodynamic method according to standard PN-EN-ISO 10993-15 [20]. Every test was started by determining the open circuit potential E_{OCP}.

Fretting corrosion tests were divided into three trials. All trials were repeated five times. The first trial

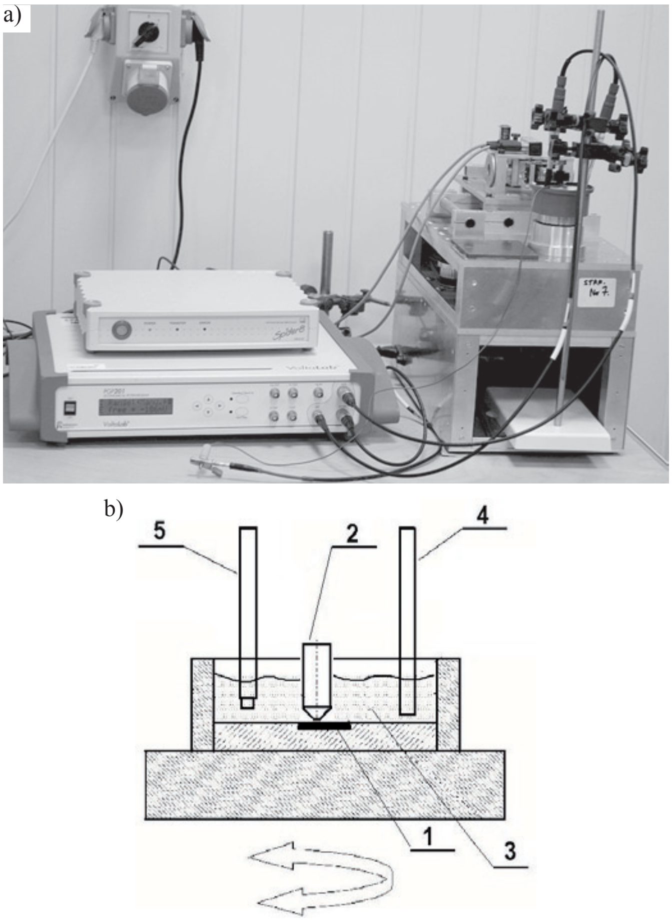


Fig. 1. Testing station: a) general view, b) diagram of the testing system: 1,2 – sample and countersample, 3 – tested solutions, 4 – saturated calomel electrode, 5 – platinum electrode

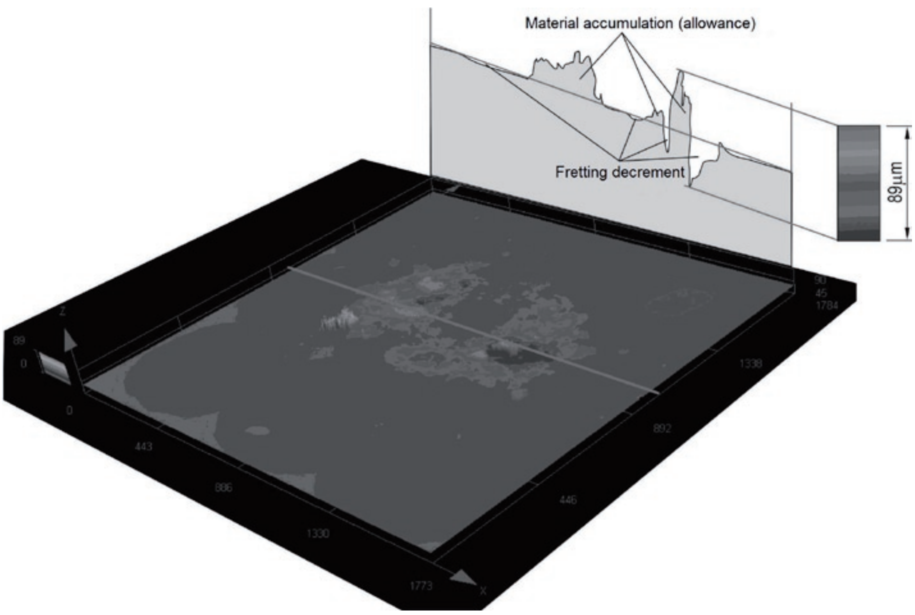


Fig. 2. Diagram of the fretting wear assessment method

was a reference test, during which pitting corrosion resistance was assessed. At the start of potentiodynamic tests, potential was set at the level $E_{init} = E_{OCP} - 100$ mV. Potential change took place in the direction of the anode at a rate of 3 mV/s. After the achievement of the maximum value in the measuring range +4094 mV or after the achievement of anode current density at 1 mA/cm^2, a change in the polarization direction took place. In the second trial, the fretting process was started after the determination of E_{OCP} potential, and changes of potential were recorded over one hour (2880 cycles). Next, a potentiodynamic test of samples with damaged surfaces was conducted. The final test involved the registration of the electrochemical processes occurring during fretting (approx. 900 cycles).

The registration of anode polarization curves during individual stages of fretting corrosion made it possible to determine characteristic quantities describing the resistance of steel to pitting corrosion in the environment of saliva and its substitutes. Open circuit potential (E_{OCP}), corrosion potential (E_{cor}), repassivation potential (E_{rep}), and polarization resistance (R_p) were registered. The value of polarization resistance was determined using Stern's method by analyzing the range of ±10 mV relative to the corrosion potential. This was conditioned by the requirements of maintaining a linear dependency between current density and sample potential. The value of repassivation potential was read at the coordinates of the point of intersection of the anode curve with the return curve.

Observations of sample surfaces were carried out using a Hitachi S-3000N (SEM) scanning electron microscope and a LEXT OLS4000 confocal microscope.

3. Test results and discussion

Measurements of friction forces and estimated friction coefficients were performed under stable conditions during fretting tests. Tests were performed in the environment of natural saliva and developed substitutes. Obtained results are presented in fig. 3.

The data shown in fig. 3 indicates that friction pairs in the environment of mucin III and natural saliva are characterized by the lowest resistances to motion. The greatest resistances to motion, which were reduced as pressure increased, were observed under dry friction conditions. This is probably related to the oxides that are formed on the surface of steel, which create a protective anti-adhesive layer that reduces frictional resistance after being broken up. The presence of lubricant substances in the friction pair limits oxygen access, which reduces the capability of protective oxide layer formation. In this case, the rheological properties of lubricant fluids and the content of anti-friction modifiers (mucin) capable of forming adsorbent boundary layers are decisive. These layers prevent direct metal-metal contact, thus reducing resistance to motion and wear. The worsening of the tribological characteristics of compositions containing xanthan gum, particularly under small loads, may be related to the antagonistic effect of this additive to the adsorptive capabilities of mucin and to an increase of the viscosity of the tested lubricant composition (D). A more precise explanation of the observed phenomenon requires further study.

It should be emphasized that weak removal of wear products from the friction area is characteristic of fretting processes. Disk wear test results are presented in fig. 4. This data indicates that the greatest wear occurs during dry friction. This is expressed by the accumulation of a large amount of wear products on the surface of friction (excess material). In a large degree, these are oxidized products of primary wear, mainly in the form of iron and chromium oxides (components of the steel) [32]. As a result of intensive oxidation processes, the volume of wear products in comparison

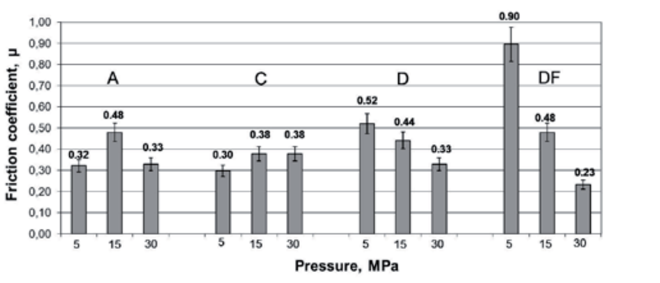


Fig. 3. Influence of saliva and its substitutes on friction coefficient values: A – human saliva, C – mucin III in PBS, D – mucin III + xanthan gum in PBS, DF – dry friction

to the native material and their hardness are increased. This has an effect on further friction and tribological destruction processes.

The application of lubricant fluids significantly reduced the amount of wear, particularly in the presence of saliva and of the mucin solution. The addition of xanthan gum (solution D) led to increased wear. As mentioned earlier, this is probably caused by a reduction of mucin adsorption, which limits the formation of protective boundary layers. Furthermore, an increase of fluid viscosity has an unfavorable impact on capabilities of wear product removal from the friction area. This may lead to intensification of wear processes (secondary wear).

As mentioned earlier, the significant effect of fretting on the intensification of corrosion processes is indicated in many publications [4, 17, 30]. The results of conducted tests shown in table 4 confirm these constatations.

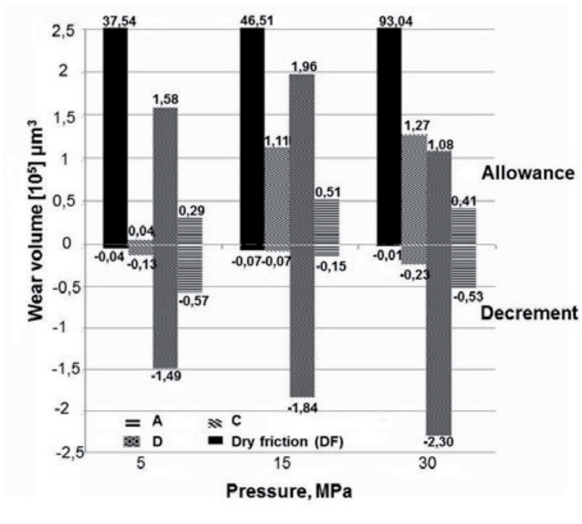


Fig. 4. Volumetric wear in the friction zone accounting for decrement – excess of material

Testing of sample potential in an open system enables preliminary assessment of the resistance of the material to corrosion processes. The influence of fretting on changes of potential is clearly visible in fig. 5.

After potential was stabilized over one hour, the fretting friction process was initiated, which led to the destruction of the protective passive layer and a sudden drop of the corrosion potential value. After friction is discontinued, the oxide layer is reconstructed and potential rises in the direction of the initial state. However, the destroyed protective layer was not fully reconstructed, which has an unfavorable influence on the corrosion resistance of the tested material. Lower open circuit and corrosion potentials of studied samples are indicative of this.

Table 4. Results of corrosion tests

Environment	Trial*	E_{OCP} [mV]	E_{cor} [mV]	R_p [kΩ·cm²]	E_{rep} [mV]
A	I	−114	−191	303,61	−23
	II	−205	−267	161,1	26
	III	−125	−354	4,42	−94
B	I	−105	−134	118,70	−76
	II	−113	−181	32,51	−165
	III	−80	−273	5,04	−109
C	I	−144	−212	97,12	−155
	II	−189	−256	44,52	−176
	III	−172	−343	3,00	−93
D	I	−151	−217	135,05	−148
	II	−249	−315	13,12	−114
	III	−240	−355	10,13	−178

**I – reference sample; II - corrosion test after fretting; III - corrosion test during fretting*

Analysis of the obtained data shows that all studied saliva substitutes exhibit similar corrosion aggressiveness. However solution D had the worst characteristics, as indicated by the lowest values of corrosion potentials as well as by the slightly greater rate of current density increase. Despite similar potentials, the greatest polarization resistances are present in the environment of natural saliva - which indicates is favorable anti-corrosion properties. Registered anode polarization curves in the environment of saliva and its substitutes are presented in fig. 6. The return curve in solution A visible on the chart indicates the occurrence of repassivation processes - the formation of a secondary protective layer. This formed layer is quickly destroyed, however, and the revealed metal surface is subject to corrosion. No clear repassivation processes were observed in other environments.

Obtained data shows that, for the studied fluids, repassivation potentials are similar to corrosion potentials, which indicates a low intensity of surface repassivation processes. The processes of passive layer destruction and reconstruction are visible in the case of simultaneous testing of fretting and corrosion. This indicates a typical, jagged course of the anode polarization curve (fig. 7).

Microscope observations of the surfaces of tested samples constituted unequivocal evidence of the intensive processes of fretting and corrosion wear. Indicative images of such surfaces are shown in fig. 8. Fretting wear processes are focused in the central friction contact area (fig. 8a). Wear products remaining in the friction area are clearly visible.

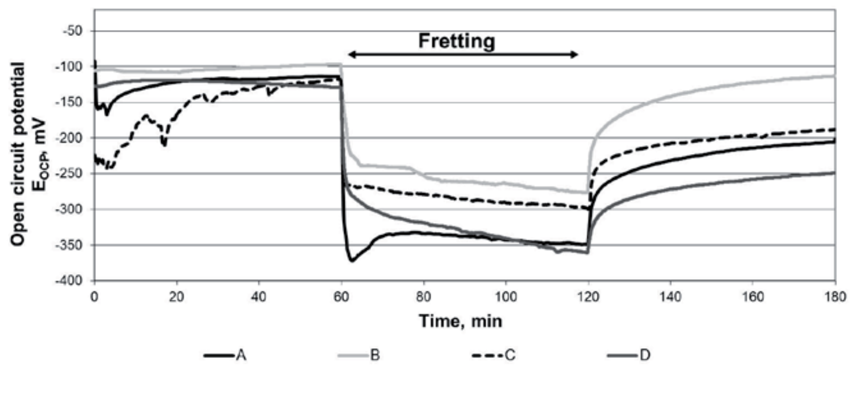


Fig. 5. Changes of the potential of the 316L implantation steel sample in an open system with fretting

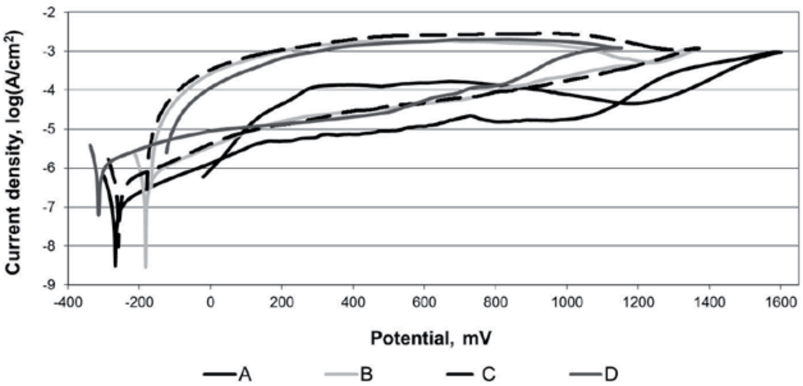


Fig. 6. Anode polarization curves in the environment of studied preparations after fretting

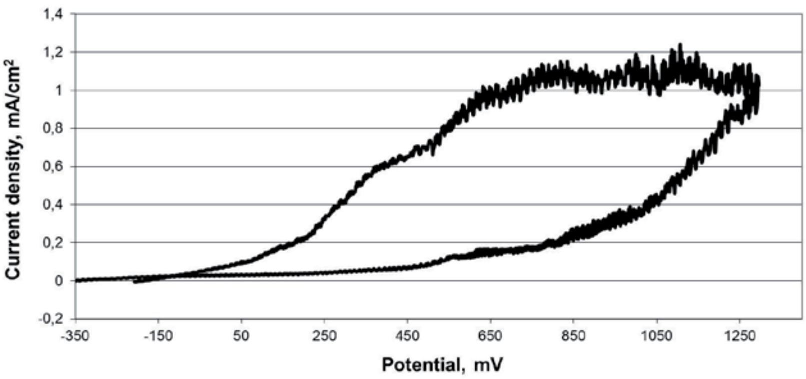


Fig. 7. Anode polarization curve in the saliva environment during fretting

Despite the difficulties in identification of the simultaneously occurring processes of tribological and corrosive wear (fretting corrosion), it was possible to observe propagation of the corrosion beyond the friction area (fig. 8b).

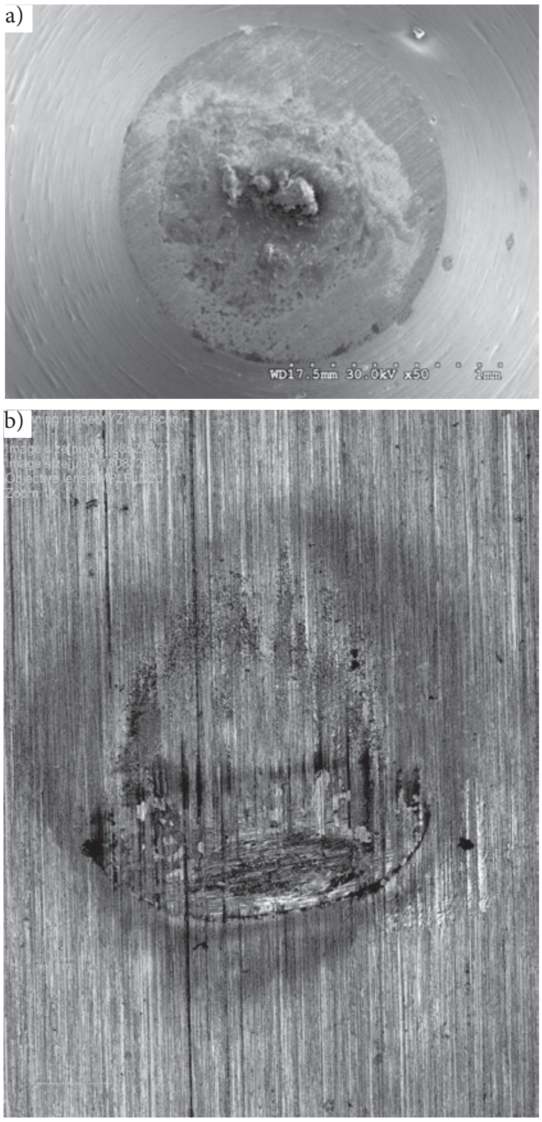


Fig. 8. Microscope photographs of the surfaces of 316L implantation steel samples after tests in the saliva environment: a) after fretting (SEM), b) after fretting corrosion (confocal microscope)

4. Conclusions

The results of studies confirmed the high susceptibility of 316L implantation steel to wear as a result of fretting and fretting corrosion. Test results indicate differing effects of saliva and its substitutes on the course of these processes. The lowest corrosive aggressiveness and the best lubricant properties, relative to natural saliva, were exhibited by the mucin III solution in phosphate buffered saline. The addition of xanthan gum had an unfavorable effect on tribological characteristics and corrosion resistance. 316L implantation steel exhibits the greatest resistance to fretting corrosion processes in the environment of natural saliva. This is indicated by high polarization resistances and visibly greater values of potential at which reverse polarization takes place.

Despite the fact that natural saliva still remains to be the least corrosive environment for 316L steel, it seems that the base solution of mucin III in PBS may constitute a basis for development of saliva substitutes with favorable tribological characteristics. Such preparations can be used to reduce the effects of affections in the oral cavity, such as bruxism, and may also be useful in improving the operational lifetime of dental prosthetics and orthopedic apparatus components.

Acknowledgements: *This work was realized with financial support within the framework of research project NCBiR NR15 0117 10.*

References

1. Andrysewicz E, Mystkowska J, Dąbrowski J R. Krawczyk-Dembicka E, Investigations of tribological and physicochemical properties of human saliva and its substitutes. Engineering of Biomaterials, 2013, 118: 23-29.
2. Andrysewicz E, Mystkowska J, Kolmas J, Jałbrzykowski M, Olchowik M R, Dąbrowski J R. Influence of artificial saliva compositions on tribological characteristics of Ti-6Al-4V implant alloy. Acta of Bioengineering and Biomechanics 2012; 14: 71-79.
3. Bhola R, Bhola S M, Mishra B, Olson D L. Corrosion in titanium dental implants/prostheses – a review, Trends in Biomaterials and Artificial Organs 2011; 25: 34-46.
4. Bieira A C, Ribeiro A R, Rocha L A, Celis J P. Influeance of pH and corrosion inhibitors on the tribocorrosion of titanium in artificial saliva. Wear 2006; 261: 994-1001.
5. Bronzino J D, ed. The Biomedical Engineering HandBook, Second Edition. Boca Raton: CRC Press LLC, 2000.

6. Dobrzański L A, Reimann Ł. Influence of Cr and Co on hardness and corrosion resistance CoCrMo alloys used on dentures. Journal of Achievements in Materials and Manufacturing Engineering 2011; 49: 193-199.
7. Dodds M W J, Johson D A, Yeh C. Health benefits of saliva: a review, Journal of Dentistry 2005; 33: 223-233.
8. Ducheyne P, Healy K, Hutmacher D E, Grainger D W, Kirkpatrick C J. Comprehensive biomaterials (6 volume set). Oxford: Elsevier, 2011.
9. Duisabeau L, Combrade P, Forest B. Environmental effect on fretting of metallic materials for orthopaedic implants. Wear 2004; 256: 805-816.
10. Eliades T, Athanasiou A E. In vivo aging of orthodontic Alloys: Implications for corrosion potential, nickel release and biocompatibility. Angle Orthodontist 2002; 72: 222-237.
11. Fraunhofer J A. Corrosion of orthodontic devices. Seminars in Orthodontics 1997; 3: 198-205.
12. Hallab N J, Messina C, Skipor A, Jacobs J J. Differences in the fretting corrosion of metal-metal and ceramic-metal modular junctions of total hip replacements. Journal of Orthopaedic Research 2004; 22: 250-259.
13. Hansen D C. Metal corrosion in the human body: the ultimate bio-corrosion scenario. The Electrochemical Society Interface 2008; 17: 31-34.
14. He D, Zhang T, Wu Y. Fretting and galvanic corrosion behaviors and mechanisms of Co-Cr-Mo and Ti-6Al-4V alloys. Wear 2002; 249: 883-891.
15. Krasicka-Cydzik E, Mystkowski J, Ciupik F L. Materiały implantowe: stal a stopy tytanu, System Dero: Rozwój technik operacyjnego leczenia kręgosłupa.
16. Kumar S, Narayanan T S N S, Raman S G S, Seshadri S K. Evaluation of fretting corrosion behaviour of CP-Ti orthopaedic implant applications. Tribology International 2010; 43: 1245-1252.
17. Lgied M, Liskiewicz T, Neville A. Electrochemical investigation of corrosion and wear interactions under fretting conditions. Wear 2012; 282-283: 52-58.
18. Mystkowska J, Jałbrzykowski M, Dąbrowski J R. Tribological properties of selected self-made solutions of synthetic saliva. Solid State Phenomena 2013; 199: 567-572.
19. Neyman A. Fretting w elementach maszyn. Wydawnictwo Politechniki Gdańskiej, 2003, Gdańsk.
20. Norma PN-EN ISO 10993-15:2009. Biologiczna ocena wyrobów medycznych. Identyfikacja i oznaczanie ilościowe produktów degradacji metali i stopów.
21. Paulin C, Fouvry S, Meunier C. Finite element modeling of fretting wear surface evolution: Application to a Ti-6Al-4V contact. Wear 2008; 264: 26-36.
22. Pellier J, Geringer J, Forest B. Fretting-corrosion between 316L SS and PMMA: influence of ionic strength, protein and electrochemical conditions on material wear. Application to orthopedic implants. Wear 2011; 271: 1563-1571.
23. Rajendran S, Uma V, Krishnaveni A, Jeyasundari J, Shyamaladevi B, Manivannan M. Corrosion behavior of metals in artificial saliva in presence of D-Glucose. The Arabian Journal for Science and Engineering 2009; 34: 147-158.
24. Rapiejko C, Fouvry S, Grosgogeat B, Wendler B. A representative ex-situ fretting wear investigation of orthodontic arch-wire/bracket contacts. Wear 2009; 266: 850-858.
25. Ratner B D, Hoffman A S, Schoen F J, Lemons J E. Biomaterials science: An introduction to materials and medicine. San Diego: Academic Press, 1997.
26. Sajewicz E. Effect of saliva viscosity on tribological behaviour of tooth enamel. Tribology International 2009; 42: 327-332.
27. Schipper R A, Silletti E, Vingerhoeds M H. Saliva as research material: Biochemical, physiocochemical and practical aspects. Archives of Oral Biology 2007; 52: 1114-1135.
28. Sharma M, Kumar A V R, Singh N, Adya N, Saluja B. Electrochemical corrosion behavior of dental/implant alloys in artificial saliva, Journal of Materials Engineering and Performance 2008; 17: 695-701.
29. Shih C C, Shih C M, Chen Y L, Su Y Y, Shih J S, Kwok C F, Lin S J. Growth inhibition of cultured smooth muscle cells by corrosion products of 316 L stainless steel wire. Journal of Biomedical Materials Research 2001; 57: 200-207.
30. Sivakumar B, Kumar S, Narayanan T S N S. Fretting corrosion behaviour of Ti-6Al-4V alloy in artificial saliva containing varying concentrations of fluoride ions. Wear 2011; 270: 317-324.
31. Sivakumar M, Shanadurai K S K, Rajeswari S, Thulasiraman V. Failures in stainless steel orthopaedic implant devices: A survey, Journal of Materials. Science Letters 1995; 14: 351-354.
32. Szymański K, Olszewski W, Satuła D, Rećko K, Waliszewski J, Kalska-Szostko B, Dąbrowski J R, Sidun J, Kulesza E. Characterization of fretting products between austenitic and martensitic stainless steels using Mossbauer and X-ray techniques. Wear 2013; 300: 90-95.
33. Tritschler B, Forest B, Rieu J. Fretting corrosion of materials for orthopaedic implants: a study of a metal/polymer contact in an artificial physiological medium. Tribology International 1999; 32: 587-596.
34. Vadiraj A, Kamaraj M. Effect of surface treatments on fretting fatigue damage of biomedical titanium alloys. Tribology International 2007; 40: 82-88.
35. Xulin S, Ito A, Tateishi T, Hoshino A. Fretting corrosion resistance and fretting corrosion product cytocompability of ferritic stainless steel. Journal of Biomediacal Mterials Research 1997; 34: 9-14.
36. Yu H Y, Quan H X, Cai Z B, Gao S S, Zhu M H. Radial fretting behavior of cortical bone against titanium. Tribology Letters 2008; 31: 69-76.
37. Zhu M H, Yu H Y, Zhou Z R. Radial fretting behaviours of dental ceramics. Tribology International 2006; 39: 1255-1261.
38. Zhu M H, Zhou Z R. On the mechanisms of various fretting wear modes. Tribology International 2011; 44: 1378-1388.

Jan Ryszard DĄBROWSKI
Marcin KLEKOTKA
Jarosław SIDUN
Department of Materials and Biomedical Engineering
Białystok Technical University
Ul. Wiejska 45 C, 15-351 Białystok, Poland
E-mails: j.dabrowski@pb.edu.pl, m.klekotka@pb.edu.pl, j.sidun@pb.edu.pl

Article citation info:
ODHIAMBO S, DE MEY G, HERTLEER C, VAN LANGENHOVE L. Reliability testing of PEDOT:PSS capacitors integrated into textile fabrics. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 447–451.

Sheilla ODHIAMBO
Gilbert DE MEY
Carla HERTLEER
Lieva VAN LANGENHOVE

RELIABILITY TESTING OF PEDOT:PSS CAPACITORS INTEGRATED INTO TEXTILE FABRICS

BADANIE NIEZAWODNOŚCI KONDENSATORÓW PEDOT: PSS WBUDOWANYCH W TKANINY

Textile-based capacitors have been made from polyethylene dioxythiophene, polystyrene sulphonate (PEDOT:PSS) as the electrolyte and pure stainless steel filament yarns as the electrodes. The capacitor is well integrated into the textile structure, small in size and of light weight. Although they experience a self-discharge, the reliability of the PEDOT:PSS capacitors has been investigated by repeating up to 14 cycles of charging and discharging. Initially, the voltage output turns out to be higher with increasing number of cycles. However, after the fifth cycle, degradation of the cell starts occurring and a decreasing behaviour in the voltage output is observed. One can roughly say that these capacitors could be used up to 10–15 cycles.

Keywords: *capacitors, conductive yarns, PEDOT:PSS, voltage discharge, cycling.*

Kondensatory tekstylne wytwarza się z mieszaniny poli(3,4-dioksyetylenotiofenu) z polistyrenem sulfonowanym (PEDOT: PSS), pełniącej rolę elektrolitu oraz włókien ciągłych z czystej stali nierdzewnej, pełniących funkcję elektrod. Kondensatory tego typu są dobrze zintegrowane ze strukturą tkaniny, są lekkie i mają niewielkie rozmiary. Chociaż kondensatory PEDOT:PSS ulegają samorozładowaniu, przeprowadzono badania ich niezawodności powtarzając 14 cykli ładowania i rozładowywania Początkowo napięcie wyjściowe zwiększało się wraz ze wzrostem liczby cykli. Jednakże po piątym cyklu, dochodziło do degradacji ogniwa i obserwowano zmniejszanie się napięcia wyjściowego. Można orientacyjnie powiedzieć, że omawiane kondensatory nadają się do użytku przez maksymalnie 10–15 cykli.

Słowa kluczowe: *przędze przewodzące, PEDOT: PSS, rozładowanie napięcia, próba cykliczna.*

1. Introduction

In recent years a lot of effort is put into integrating electronic components into textiles, a new discipline called smart textile design. The applications of smart textile systems can be found in many fields: protective clothing, medical applications and sports clothing. An overview can be found in literature [2, 27, 29, 30].

Flexibility of the electronic components is desirable with these recent developments and can be achieved from organic and inorganic materials in smaller forms like microstructures or nanostructures [23].

If electronics have to be integrated into a textile garment, one is dealing with all possible electronic components like conductors, resistors, capacitors, transistors and displays. Electric conductors can be made by inserting electrically conducting yarns into a fabric [9], or by suitable coating of conductive compounds on a non-conducting yarn [6, 21, 24, 25, 26]. These conductive yarns can be made from materials like stainless steel yarns or hybrids of conducting and non-conducting yarns. Screen printing has also been successfully used to deposit conducting layers on a fabric [11]. However, textile being a flexible and porous material, one must pay special attention to the mechanical properties and their influence on the electric characteristics [22, 26]. Besides electric conductors and resistors, other components like transistors, capacitors or displays have been integrated into a textile material [3, 15, 17, 28].

It must be clearly pointed out that full integration into a fabric means that the electrical component is only made out of textile material and/or polymers embedded into the textile during the production process and not added as detachable in the final assembly of the garment. As a consequence, these components cannot be removed. Other applications involve electronic components which are attached to a fabric. Garments equipped with LED lights are a typical example of these. Maintenance and reliability of these devices is very important in their proper functioning and life span. In this respect, various experiments on reliability are conducted on the developed capacitor to determine these aspects. Care and washability are also essential if at all the device is fully compatible with the textile. Influence of several washing cycles on the electrical performance of electronic textiles like sensors and antennas have been performed by a number of researchers.

A flexible and lightweight energy storage device which is either a capacitor or a battery is described in the papers [4, 5, 7, 8, 12, 16, 18, 19]. All of them involve a textile or a textile material in either fibrous form, or in the textile structure.

In this paper, we focus on the reliability and stability of an electric energy storage device – capacitor intended to supply power to the integrated electronic components and circuits. The type I capacitor (both the anode and the cathode are made of the same material) we are investigating in this contribution uses the PEDOT:PSS polymer as the “dielectric” or “electrolyte” material between the two electrodes, which are made from pure stainless steel filament yarns sewn on a textile substrate. A first report of such a device was published by Bhattacharya [1]. Their device used silver coated polybenzoxazole (PBO) yarns as the electrodes. However, the performance of the stainless steel filament yarns is shown to be better than the silver coated yarn electrodes, as described in our article [20].

2. Sample preparation

A three layered laminate of textile substrate (woven cotton/polyester) with the same specifications as used in our paper [20] was adopted. The electrodes were pure stainless steel filament yarns from Bekinox® Bekaert. The electrodes were sewn at a close distance to each other into the fabric substrate. Therefore, there is no relative movement between the parts of the capacitor i.e the solid electrolyte and the electrodes, which may interfere with functionality. The upper surface of the fabric (except for a left out region of 10 mm by 6 mm including part of the electrodes) was made hydrophobic by using a thermoplastic polyurethane (TPU) layer from SunChemical. The TPU prevented the PEDOT:PSS from spreading too much in the fabric. Water based PEDOT:PSS from Ossila (of PEDOT:PSS ratio of 1:6, approximately) was coated in layers on the left out region. The definition of this ratio of PEDOT:PSS is important, because the product exists in different component ratios from different companies with different conductivities. The performances of the PEDOT:PSS brands as electrolyte for our capacitor are different from each other, based on the aspect ratio and may be from any other additives within the polymer solution.

A schematic view of the capacitor design is shown in Figure 1.

The PEDOT:PSS was applied on the foreseen area with a syringe while the fabric was in the oven. Each layer of PEDOT:PSS was left to dry and cure in the oven for 15 minutes at temperatures of 90-100^{0}C, before applying the next layer.

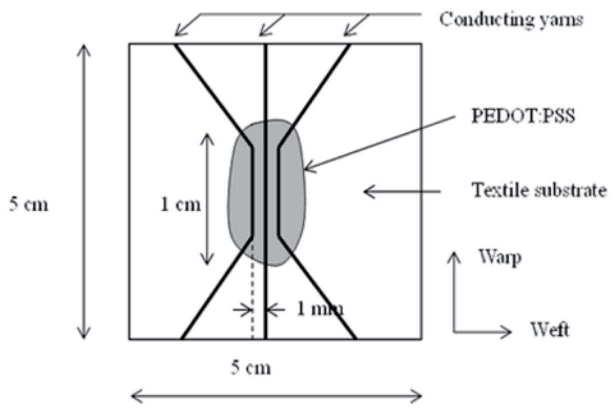


Fig. 1. Cross section view of the device

3. Material investigation

PEDOT:PSS is a conjugate polymer, namely polyethylene dioxythiophene (PEDOT) and a polystyrene sulphonate (PSS). The formulae for both polymers and how they interact are shown in Figure 2. The PEDOT molecule can lose one or more electrons whereas the PSS receives those electrons. The PEDOT has several S^+ (positive) ions whereas the PSS molecule will have then several $SO_3{}^-$ (negative) ions as shown in Figure 2. The material behaves then like a solid electrolyte. Under influence of an externally applied electrical field the charged PEDOT and PSS polymer chains will move in opposite directions so that the material will be electrically polarised and the capacitor becomes charged. After removal of the applied electrical field the ions will move back to their original position so that the material loses its polarisation.

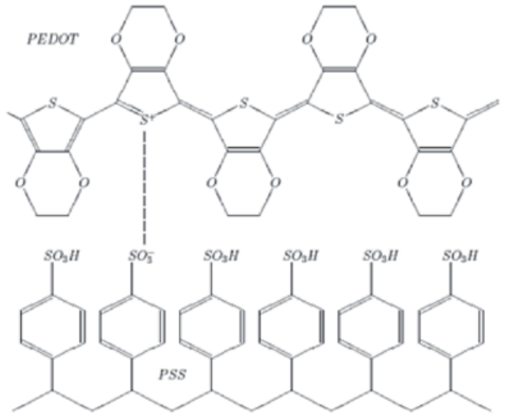


Fig. 2. Schematic of the chemical structure of PEDOT:PSS showing the ion sites

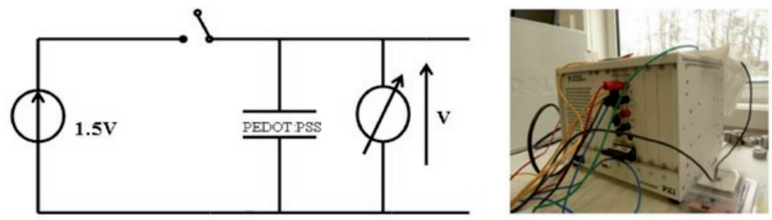


Fig. 3. Charge-discharge circuit and the NI PXI

4. Experimental electrical measurements & discussion

First of all the fabricated PEDOT:PSS textile capacitor was charged with a constant voltage of 1.5 V for 2 hours with the circuit shown schematically in Figure 3. After opening the switch the self-discharge of the PEDOT:PSS capacitor was recorded with a voltage meter having a high input resistance of 10MΩ. Since each measurement lasts for several hours, the apparatus NI PXI from National Instruments was used to carry out the operations automatically. The NI PXI 1033 is a chassis equipped with several voltage generators, a digital voltage meter and a computer interface. For the switch, a relay was used, which was controlled by one of the voltage generators. A special software package running on LabVIEW was written to carry out all the measurements automatically, including the transfer of data to an external computer.

After the charging time of 2 hours, the switch was opened and the first observed discharge characteristic was measured. The PEDOT:PSS capacitor was not connected to any voltage for at least 10 hours before the next cycle was started. A day later, the second charging cycle of 2 hours was applied followed by measuring the second discharge char-

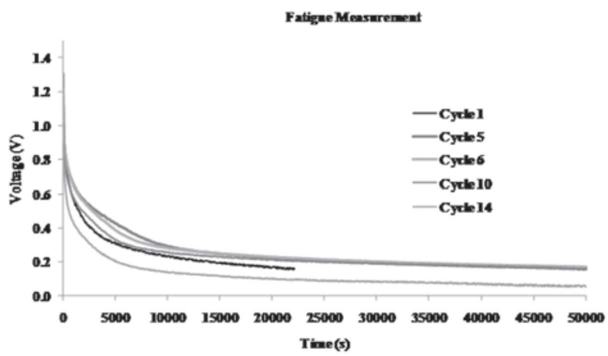


Fig. 4. Fatigue test measurements

acteristic. This procedure was repeated up to 14 times. The results are shown in Figure 4.

The output voltages have been drawn as functions of time up to 50,000 s (about 14 hours). One remarks that during the first 5 to 6 cycles the output voltage is increasing, which means that the device is improving per each subsequent cycle. This could be due to the residual charge in the device from a subsequent previous charging. But when more cycles are applied, the device seems to get worse. This could also be attributed to the onset of the degradation of the electrolyte since the experiments are conducted in the ambient environment of normal humidity and temperature, this is known to have an influence on PEDOT:PSS activity and degradation.

A closer look at the characteristic of one single graph in Figure 5 reveals the following: an immediate observation is that the voltage drops rapidly in the beginning. But after some time (100 s) the voltage tends to be more stable around a value of 0.4 V for a rather long time (up to several hours).

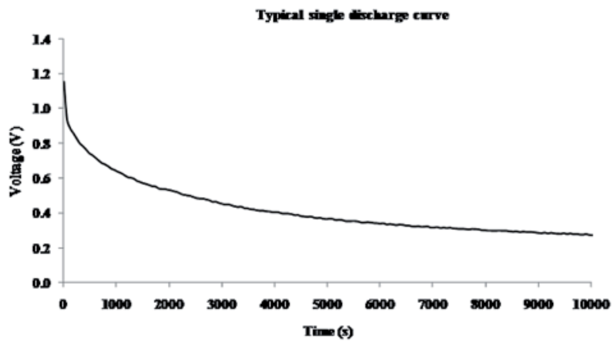


Fig. 5. Discharge characteristic of single cycle

If one takes into account that a voltage of 0.4 V is rather small as compared to the initial charging voltage of 1.5 V, then the efficiency of our fabricated device is rather low. Also the number of charging/discharging cycles is rather limited. But on the other hand, we are dealing with a device which is fully integrated into a textile fabric. This is the price one has to pay to have a completely integrated component.

The main purpose of this research is to investigate the reliability of the PEDOT:PSS textile capacitors. Other authors reported that a similar device with silver coated PBO yarn electrodes could be charged/discharged up to 4 times [1]. Accordingly our results presented in Figure 4 show that devices equipped with stainless steel electrodes can be charged and discharged up to 14 times. At least one day elapsed between each two cycles. Also Figure 4 clearly shows the degradation of the cells after 5 to 6 cycles. Up to 5 cycles the output voltage is increasing but for more cycles the decreasing behaviour is clearly observed. After 5 charging/discharging cycles, the devices started to get lower output voltages. A clearer view of this phenomenon is shown in Figure 6, where the recorded output voltage is displayed as a function of the number of cycles N at several times after opening the switch *S* (t = 3000 s, t = 6000 s, t = 12000 s, t = 18000 s and t = 36000 s). Remark that t = 36000 s corresponds to 10 hours of discharging time.

One can roughly say that these capacitors can be used up to 10-15 charging/discharging cycles. This number is rather small and one might have the impression that these components are inapplicable in practice. However, for wearable textiles, electric components with a limited reliability have proved their applicability [10]. Besides this, the study of PEDOT:PSS capacitors integrated into fabrics started very recently, therefore this topic is still in the initial phase of fundamental research.

The voltage measurements were done with a digital instrument (National instrument) with a 3 digit accuracy. From Figure 6, one might have the wrong impression that the measurements contain large errors because the curves are far from being smooth. This phenomena is entirely due to the (still unknown) physical mechanisms inside the PEDOT:PSS material.

Taking into account that the discharge curves were recorded with a device having a 10 MΩ input impedance, the current could be easily evaluated. A numerical integration gave the total charge. The ratio of this charge with respect to the applied voltage yields a capacitance value around 18000μF. By adding resistors in parallel, the internal series resistance was measured to be 300 kΩ.

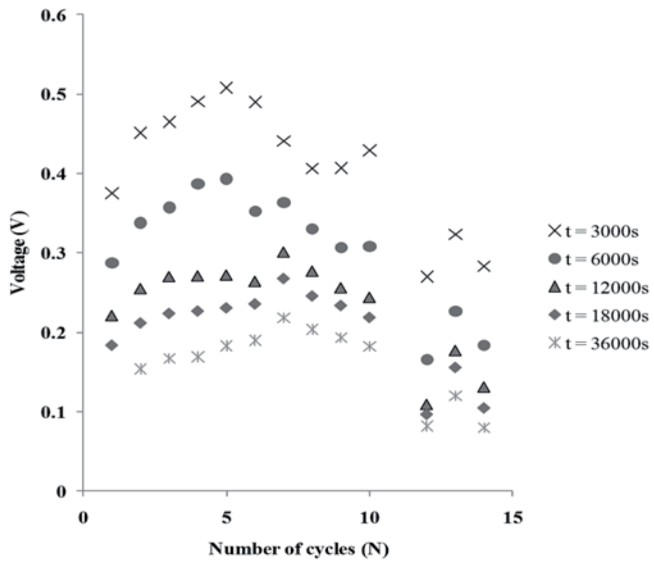


Fig. 6. Graph showing discharge behaviour of the capacitors at specific times for different number of cycles (N)

A typical problem related to PEDOT:PSS is that the electric conduction mechanism is still not well understood. As some authors claim it is still under debate [13, 20] or in other words a lot of research has to be done to fully understand the fundamental phenomena happening in this material. It was mentioned before that ions are responsible as shown in Figure 1. The charge and discharge is expected to involve cation transport [14], where migration is expected to occur. But some authors found that by using silver coated yarn electrodes electrolytic phenomena occur i.e deposition of silver ion that moves from the anode electrode to the cathode, this observation was done using SEM [1].

We observed that with silver coated yarn electrodes the output voltage was even lower (up to 50 %) than with the stainless steel contacts [20]. All these experiments have proved that other phenomena like electrolysis cannot be excluded. Hence, one can start the discussion whether we are dealing with a capacitor or a battery or a mixture of these. Obviously, when the conduction mechanism will be better understood, it will be easier to search for devices with better characteristics and performance.

5. Conclusion

A capacitor well integrated into the textile structure that is small and light weight has been made. The device shows some robustness and can withstand up to 15 cycles of each 7200 seconds charging at 1.5V. However, the efficiency of energy storage is still very low due to the self-discharge. One can roughly say that these capacitors could be used up to 10–15 cycles, with no significant difference in the output energy level for the first 10 cycles. This shows the limited level of reliability of the capacitor. Consequently, the decay of the discharge characteristic has to be taken into account during the design phase of the application if the capacitor will be used for a more efficient performance. More fundamental research will still be necessary in the future. The self-discharge of the capacitors has to be improved.

Acknowledgements:
Sheilla A. Odhiambo, on study leave from Moi University, Eldoret, Kenya, wants to thank the coordinators of MU-VLIR UOS (Flemish inter university council project) for the financial support for her research and stay at Ghent University. She also thanks the management of Moi University for granting her the study leave.

References

1. Bhattacharya R, de Kok MM, Zhou J. Rechargeable electronic textile battery. Applied Physics Letters 2009; 95: 22.
2. Cao Lihui, Li Yinglin. The Development and Application of Smart Garment Materials. Materials and Manufacturing Technology, Pts 1 and 2. X. Yi and L. Mi. Stafa-Zurich. Trans Tech Publications Ltd 2010; 129-131: 472-475.
3. Gu JF, Gorgutsa S and Skorobogatiy M. Soft capacitor fibers for electronic textiles. Applied Physics Letters 2010; 97: 13.
4. Horng Ying-Ying, Lu Yi-Chen, Hsu Yu-Kuei, et al. Flexible supercapacitor based on polyaniline nanowires/carbon cloth with both high gravimetric and area-normalized capacitance. Journal of Power Sources 2010; 195(13): 4418-4422.
5. Hu L, La Mantia F, Wu Hui, Xie X, James M and Cui Y. Lithium-Ion Textile Batteries with Large Areal Mass Loading. Advanced energy material 2011;1(6): 1012-1017.
6. Irwin Michael D, Roberson David A, Olivas Richard I, Wicker R B and MacDonald E. Conductive Polymer-Coated Threads as Electrical Interconnects in e-Textiles. Fibers and Polymers 2011; 12(7): 904-910.
7. Jost K, Perez CR, McDonough JK, Presser V, Heon M, Dion G, Gogotsi Y. Carbon coated textiles for flexible energy storage. Energy & Environmental Science 2011; 4(12): 5060-5067.
8. Kaltenbrunner M, Kettlgruber G, Siket C, Schwodiauer R, Bauer S. Arrays of Ultracompliant Electrochemical Dry Gel Cells for Stretchable Electronics. Advanced Materials 2010; 22(18): 2065-2067.
9. Kayacan O, Bulgun E, Sahin O. Implementation of Steel-based Fabric Panels in a Heated Garment Design. Textile Research Journal 2009; 79(16): 1427-1437.
10. Kazani I. Study of screen-printed electroconductive textile materials. PhD dissertation, Gent university 2012.
11. Kazani I, Hertleer C, De Mey G, Guxho G, Van Langenhove L. Dry cleaning of electroconductive layers screen printed on flexible substrates. Textile Research Journal 2013; 83(14): 1541-1548.
12. Kazani I, Hertleer C, De Mey G, Schwarz A, Guxho G, Van Langenhove L. Electrical Conductive Textiles Obtained by Screen Printing. Fibres & Textiles in Eastern Europe 2012; 20(1): 57-63.
13. Laforgue A. All-textile flexible supercapacitors using electrospun poly(3,4-ethylenedioxythiophene) nanofibers. Journal of Power Sources 2011; 196(1): 559-564.
14. Lee Hyo Joong, Lee Joowok, and Park Su-Moon. Electrochemistry of Conductive Polymers. 45. Nanoscale Conductivity of PEDOT and PEDOT:PSS Composite Films Studied by Current-Sensing AFM. Journal of Physical Chemistry B 2010; 114(8): 2660-2666.
15. Li GC, Pickup PG. Ion transport in poly(3,4-ethylenedioxythiophene)-poly(styrene-4-sulfonate) composites. Physical Chemistry Chemical Physics 2000; 2(6): 1255-1260.
16. Liu Y, Gorgutsa S, Clara S, Skorobogatiy M. Flexible, solid electrolyte-based lithium battery composed of LiFePO4 cathode and Li4Ti5O10 anode for applications in smart textiles. Journal of the Electrochemical Society 2012; 159(4): A349-A356.
17. Meng CZ, Liu C, Chen L, Hu C, Fan S. Highly Flexible and All-Solid-State Paper like Polymer Supercapacitors. Nano Letters 2010; 10(10): 4025-4031.
18. Muller Christian, Hamedi M, Karlsson R, Jansson R, Marcilla R, Hedhammar M and Inganas O. Woven Electrochemical Transistors on Silk Fibers. Advanced Materials 2011; 23(7): 898-901.
19. Nishide, H. and Oyaizu K. Materials science - Toward flexible batteries. Science 2008; 319(5864): 737-738.
20. Nyholm L, Nystrom G, Mihranyan A and Stromme M. Toward Flexible Polymer and Paper-Based Energy Storage Devices. Advanced Materials 2011; 23(33): 3751-3769.
21. Odhiambo S, Heertleer C, Schwarz A, Van Langenhove L, De Mey G. Discharge characteristics of PEDOT: PSS textile batteries; comparison of silver coated yarn electrodes devices and pure stainless steel filament yarn electrodes devices. Textile Research Journal 2013; Accepted.
22. Ouyang J, Xu QF, Chu CW, Yang Y, Li G, Shinar J. On the mechanism of conductivity enhancement in poly (3,4-ethylenedioxythiophene): poly(styrene sulfonate) film through solvent treatment. Polymer 2004; 45(25): 8443-8450.
23. Rattfalt L, Linden M, Hult P, Berglin L, Ask P. Electrical characteristics of conductive yarns and textile electrodes for medical applications. Medical & Biological Engineering & Computing 2007; 45(12): 1251-1257.
24. Rogers JA, Someya T, Huang Y. Materials and Mechanics for Stretchable Electronics. Science 2010; 327(5973): 1603-1607.
25. Schwarz A, Cuny L, Hertleer C, Ghekiere F, Kazani I, De Clercq G, De Mey G, Van Langenhove L. Electrical circuit model of elastic and conductive yarns produced by hollow spindle spinning. Materials Technology 2011; 26(3): 121-127.
26. Schwarz A, Hakuzimana J, Kaczynska A, Banaszczyk J, Westbroek P, McAdams E, Moody G, Chronis Y, Priniotakis G, De Mey G, Tseles D, Van Langenhove L. Gold coated para-aramid yarns through electroless deposition. Surface & Coatings Technology 2010; 204(9-10): 1412-1418.
27. Schwarz A, Hakuzimana J, Westbroek P, De Mey G, Priniotakis G, Nyokong T, Van Langenhove L. A study on the morphology of thin copper films on para-aramid yarns and their influence on the yarn's electro-conductive and mechanical properties. Textile Research Journal 2012; 82(15): 1587-1596.
28. Stead L, Goulev P, Evans C, Mamdani E. The Emotional Wardrobe. Personal and Ubiquitous Computing 2004; 8(3-4): 282-290.
29. Tao Xuyuan, Koncar V, Dufour C. Geometry Pattern for the Wire Organic Electrochemical Textile Transistor. Journal of the Electrochemical Society 2011; 158(5): H572-H577.
30. Van Langenhove L, Hertleer C. Smart clothing: a new life. International journal of clothing science and technology 2004; 16: 63-72.

Sheilla ODHIAMBO
Department of Textiles
Ghent University
Technologiepark 907, 9052 Zwijnaarde, Belgium
Department of Textiles
Moi University
Eldoret, Kenya

Gilbert DE MEY
Department of Electronics and Information Systems
Ghent University
Sint Pietersnieuwstraat 41, 9000 Ghent, Belgium

Carla HERTLEER
Lieva VAN LANGENHOVE
Department of Textiles
Ghent University
Technologiepark 907, 9052 Zwijnaarde, Belgium
E-mails: SheillaAtieno.Odhiambo@UGent.be, Gilberd.DeMey@UGent.be

Article citation info:
MICHALSKI R, GONERA J, JANULIN M. A simulation model of damage-induced changes in the fuel consumption of a wheeled tractor. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 452–457.

Ryszard MICHALSKI
Jarosław GONERA
Michał JANULIN

A SIMULATION MODEL OF DAMAGE-INDUCED CHANGES IN THE FUEL CONSUMPTION OF A WHEELED TRACTOR

MODEL SYMULACYJNY ZMIAN ZUŻYCIA PALIWA CIĄGNIKA KOŁOWEGO W ASPEKCIE USZKODZEŃ*

The existing diagnostic systems are applied to monitor and optimize a tractor's performance and effectiveness, while there is no consumption monitoring systems in terms of damage. A malfunction can be analyzed at different levels of complexity, including systems, kinematic pairs and components. Based on the generated consequences, defects can be classified into the following groups. Are the following classes of damage: damage to functional, emission, unsafe deteriorating dynamics, which are assigned to certain effects. The study was prepared simulation model wheeled tractor, which incorporates traction characteristics describe physical phenomena associated with the operation of the tractor and having an impact on the process of degradation under certain loading cycles. An algorithm for determining fuel consumption during simulated defects in a wheeled tractor is presented in this paper. Most tractor malfunctions affect fuel consumption. Fuel consumption is one of the key diagnostic parameters in evaluations of a vehicle's technical condition. The magnitude of changes in fuel consumption varies subject to the type of defect. The effects of simulated malfunctions of a wheeled tractor on its fuel consumption are discussed in this paper.

Keywords: *wheeled tractor, model simulation, damage, fuel consumption.*

Stosowane obecnie elektroniczne systemy w ciągnikach kołowych służą do monitorowania i optymalizacji efektów pracy pod kątem jego wydajności i efektywności, natomiast brak jest systemów monitorujących zużycie paliwa w aspekcie uszkodzeń. Uszkodzenie można rozpatrywać na różnych poziomach złożoności maszyny, np. układów, zespołów węzłów konstrukcyjnych lub elementów. W pracy przyjęto klasyfikację uszkodzeń ze względu na ich skutki. Wyróżniono następujące klasy uszkodzeń: uszkodzenia funkcjonalne, emisyjne, zagrażające bezpieczeństwu, pogarszające dynamikę, którym przyporządkowano określone skutki. W pracy przygotowano model symulacyjny ciągnika kołowego, w którym uwzględniono charakterystyki trakcyjne opisujące zjawiska fizyczne związane z funkcjonowaniem ciągnika i mające wpływ na proces jego degradacji w określonych cyklach obciążeń. Przygotowano algorytm służący do określania zmian zużycia paliwa przy symulowanych uszkodzeniach ciągnika kołowego. Większość z uszkodzeń ciągnika kołowego ma wpływ na zużycie paliwa. Zużycie paliwa może być jednym z podstawowych parametrów diagnostycznych podczas oceny stanu technicznego pojazdu. W zależności od rodzaju uszkodzenia zmiany zużycia paliwa mogą być różne. W pracy przedstawiono przykładowe przebiegi symulacyjne niektórych uszkodzeń ciągnika kołowego na zużycie paliwa.

Słowa kluczowe: *ciągnik kołowy, model symulacyjny, uszkodzenie, zużycie paliwa.*

1. Introduction

The agricultural tractor is rather a heavy machine and is used for a variety of operations from tillage to haulage and under diverse conditions. Whatever the case may be, one of the most important considerations is to ensure safety in operation or in other words a hazard free operation [8]. Statistical data shows tractor-related accidents cause approximately 300 fatalities each year [17].

This paper proposes a simulation model for diagnosing defects in a wheeled tractor's components and evaluating the consequences of the resulting damage. A simulation model of a wheeled tractor should support evaluations of:

- tractor's functionality during transport and operation,
- performance parameters,
- operating safety in field and road driving modes,
- exhaust gas emissions.

A tractor has to be maintained in good operating condition during seasonal field works in farming and forestry. The technical condition of a wheeled tractor has to be regularly monitored to ensure its full functionality, to lower repair costs and minimize down time. The existing diagnostic systems are applied to monitor and optimize a tractor's performance and effectiveness, and they are often limited to analyzing engine performance and comparing the measured parameters with standard values [1, 16, 18].

In this study, a simulation model was developed on the assumption that defect diagnosis in a wheeled tractor is a process of detecting, isolating and describing malfunctions:

- detection of defects (determining the moment of damage),
- localization of defects (determining the type and place of damage),
- identification of defects (determining the magnitude and variability of diagnostic parameters over time).

The physical, chemical and mechanical properties of arable soil in north-eastern Poland vary significantly. For this reason, the efficiency of tractor performance is an important and a complex consideration. Tractor performance is determined by various dynamic input

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

functions that change over time. At present, there are no methods for continuously monitoring tractor efficiency that account for all energy losses. In order to support the optimization of tractor performance, the above indicators have to be determined on-line [7].

Tractor performance on actual soil conditions differs substantially from the results of laboratory track testing. The physical laws governing movement and behaviour in general vary from one surface to the other, while constant changes in terrain and drawbar pull generate ongoing variations in the dynamic load on each wheel. [3].

Traction properties of a wheeled tractor largely depend on the interaction between the wheel and the ground. The optimal use of a wheel consists of ensuring maximum efficiency of a wheel, not exceeding the permissible values of wheel slip, and obtaining the maximum tractive effort [15, 19].

A new approach to simulations of wheeled vehicle operation needs to be adopted to ensure compliance with environmental protection requirements and to reduce fuel consumption. Harmful substances emitted with exhaust gas have a highly detrimental effect on plant production and the environment, and they create serious environmental risks [4, 5, 9].

In the past decades, significant advances have been made in intelligent machine diagnostic systems. Modern wheeled tractors are highly complex vehicles which have to be monitored in real time with the involvement of automatic damage detection systems [6].

2. Classification of defects in a wheeled tractor

A defect is reported when at least one of the measurable or immeasurable features characterizing the tractor's technical condition ceases to support correct functioning of the machine [14]. A defect is defined as the loss of the tractor's ability to perform the assigned functions [PN-93/N-50191]. A malfunction can be analyzed at different levels of complexity, including systems, kinematic pairs and components [12].

Based on the generated consequences, defects can be classified into the following groups [10]:

- functional defects (u_f) which inhibit performance (torque, towing force, working speed, fuel consumption),
- exhaust defects (u_e) which increase toxic emissions (and noise) and fuel consumption due to a malfunction of the fuel supply system, layout of the diesel engine and the power transmission system,
- defects that jeopardize driving safety (u_s) can affect the following tractor systems: brake, suspension, steering and lights.
- defects that affect engine performance (u_d) and driving parameters in a tractor, including decreased acceleration, delayed response to changes in movement parameters, unequal power levels, significant loss of power and moment of force.

The presence of defects and malfunctions should be signaled in the monitoring system. The operator should be provided with information about the type and location of the malfunction.

In a wheeled tractor, the following symptoms are associated with the discussed categories of defects:

- **functional defects (u_f):** overheating or slipping of friction clutch, gearbox overloading and overheating, damage to gearbox bearings which increases system temperature, damage to reduction gear shaft bearings which increases system temperature, overloading of the final drive which causes overheating and seizing of differential satellite gear, damage to ring gear and shaft bearings which increases system temperature, overloading of rear portal axle which causes overheating, damage to reduction gear housing, bearing damage which increases system temperature, uncontrolled loss of air in front and rear axle tires, leak in induction system, turbocharger failure, engine overheating, cooling system failure, engine overload, wear of cylinder liners, wear of piston rings, loose sockets and valves, engine wear, unequal power levels in cylinders, oil pump damage, incorrect oil pressure reading, significant bushing clearance in the crankshaft and piston assembly, abnormal oil pressure drop in the engine lubrication system,
- **exhaust defects (u_e):** damage to gear housing, loss of gear oil, damage to reduction gear housing, loss of gear oil, damage to final drive housing, uncontrolled loss of air in front and rear axle tires, leak in induction system, engine overheating, cooling system failure, loss of coolant, wear of piston rings, loose sockets and valves, engine wear, unequal power levels in cylinders, fuel injector leak, combustion problems, malfunctioning fuel dosing system, injector failure, loss of engine oil, engine oil burning,
- **defects that jeopardize safety (u_s):** excessive clearance in steering system, light bulb damage in the lighting system, wiper malfunction, horn malfunction, loss of brake fluid, reduced brake force, reduced pressure in brake system, air in brake system, damaged brake pump, worn-out brake lining, uncontrolled loss of air in front and rear axle wheels, tire puncture, loose valve stem, oil pump damage
- **defects that affect performance (u_d):** overheating or slipping of friction clutch, gear tooth damage, damage to reduction gear teeth, damage to portal axle teeth, leak in induction system, turbocharger failure, wear of cylinder liners, wear of piston rings, loose sockets and valves, engine wear, unequal power levels in cylinders, fuel injector leak, combustion problems, malfunctioning fuel dosing system, injector damage.

3. Model for simulating defects in a wheeled tractor

A system for simulating damaged-induced changes in a tractor's fuel consumption has been proposed for the identified categories of defects: (u_f), (u_e), (u_s) and (u_d). The presented model of a wheeled tractor relies on traction values describing physical phenomena which are associated with a vehicle's operation and which determine its wear in a given load cycle (Fig. 1).

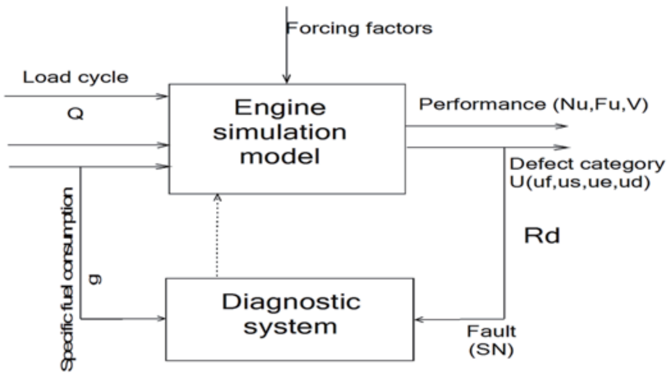


Fig. 1. Diagnostic diagram of a wheeled tractor based on the proposed mechatronic system for monitoring the condition of wheeled tractor (MSMC), where: g – specific fuel consumption, N_U - pulling power; pulling power; FU – pulling force; V – speed; SN - fault, Q – load; u – defect category [13]

A tractor's operating status is determined by: driving speed, resistance to motion (of the tractor and implements), mass, gear ratio in the power transmission system, layout of the power transmission system, rolling radius of drive wheels, etc. The severity of damage to a tractor's parts and assemblies is determined by the load cycle and material strength.

A simulation model of a wheeled tractor was developed based on a system of functional correlations presented in Figure 2. A structural diagram of a wheeled tractor is presented in Figure 3.

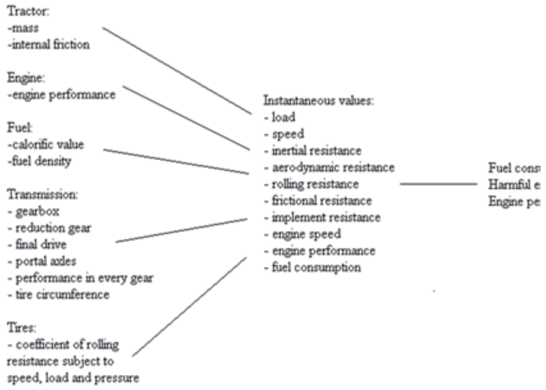


Fig. 2. Correlations between a tractor's technical parameters in a defect simulation model

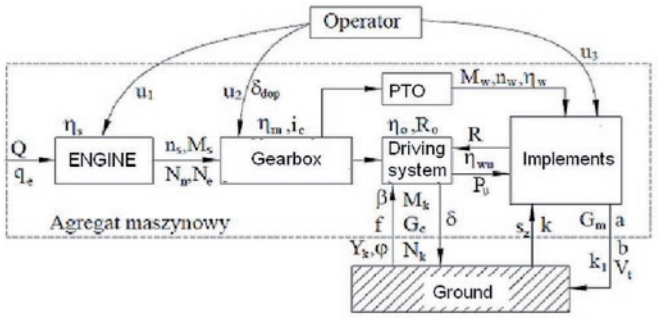


Fig. 3. Structural diagram of a wheeled tractor in the process of monitoring operating parameters [11]: β – ground inclination (slope), δ – slip ratio of drive wheels, ϕ – coefficient of tractive adhesion, η_m –mechanical efficiency, η_o – tractive efficiency (total), η_s – engine efficiency, η_u – traction efficiency, η_{wu} – traction efficiency rating, a – width of implement, b – depth of implement, f – coefficient of rolling resistance, M_w – PTO torque, n_w – PTO rotational speed, η_w – PTO efficiency, g_e –specific fuel consumption, G_c – tractor mass, G_m – implement mass, i_c –overall gear ratio, k – soil strength, k_1 –rolling resistance of implement, M_k – driving torque, M_s – net torque, N_e – effective power, N_k – driving power, N_n – nominal power output, n_s – engine rotational speed, P_n – driving force, Q – fuel consumption per hour, s_z – presence of field stones, u_1 – fuel dose, u_2 –applied gear ratio, u_3 – operating mode, Y_k – load on drive wheels, V_t – operating speed of implement operation, R – resistance of implement, R_0 – resistance in the drive , PTO – power takeoff

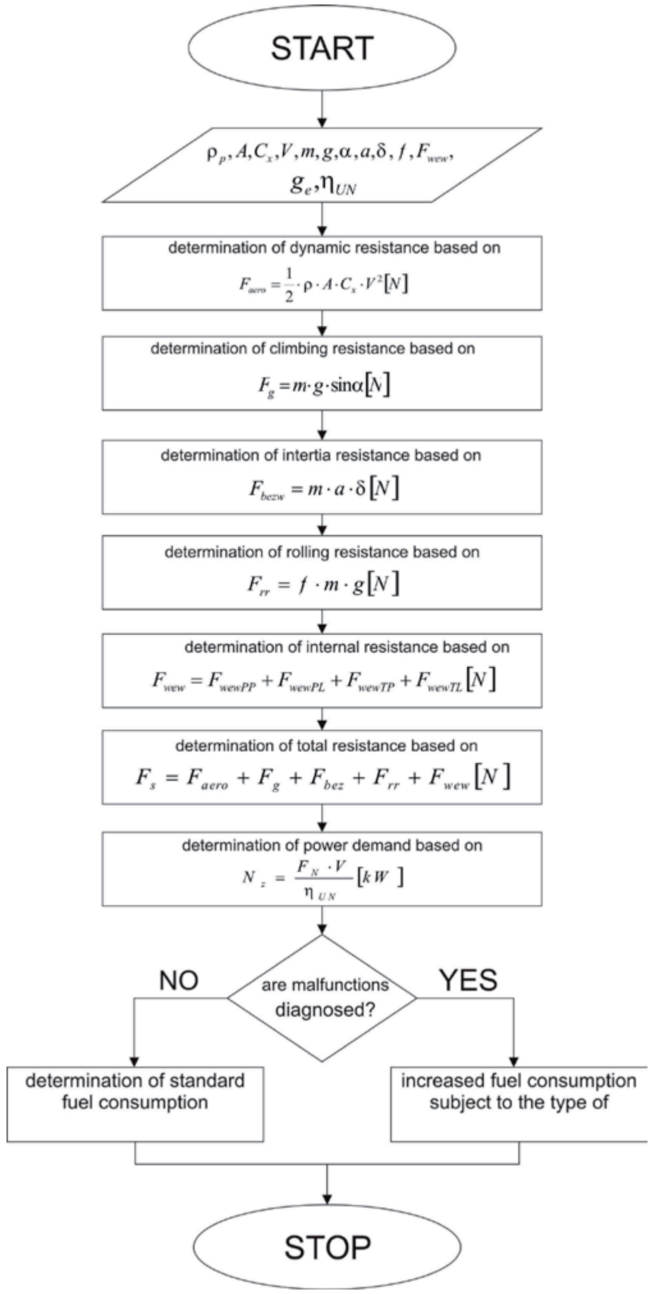


Fig. 4. An algorithm for determining fuel consumption in simulated tractor malfunctions g_e – specific fuel consumption, ρ_p – fuel density, ρ – air density, A – area of the tractor's front face, C_x – coefficient of aerodynamic drag, V – velocity, m – mass, g – gravitational acceleration, α – ground inclination, a – acceleration, δ – rotational mass coefficient, f – rolling resistance coefficient, F_{wew} – internal resistance on each wheel, η_{UN} – power transmission efficiency.

An algorithm for determining fuel consumption during simulated defects in a wheeled tractor is presented in Figure 4.

The model for simulating malfunctions in a wheeled tractor supports the identification of defects in view of their consequences (fuel consumption), as illustrated by the following relationship:

$$G_V = \frac{g_e}{1000 \cdot \rho_p}\left(\frac{1}{2} \cdot \rho \cdot A \cdot c_x \cdot V^2 + m \cdot g \cdot \sin\alpha + m \cdot a \cdot \delta + f \cdot m \cdot g + F_{wewPP} + F_{wewPL} + F_{wewTP} + F_{wewTL}\right) \cdot \frac{V}{\eta_{UN}} \quad (1)$$

Based on the analysis of the relationship defined by formula (1) can see the effect of specific resistance to motion (aerodynamic, resulting from the slope of the road, inertia, and internal resistance of rolling wheels) changes in fuel consumption.

New generation tractors are equipped with automatic control systems which replace the operator. This solution supports effective vehicle operation, reduces fuel consumption and optimizes torque and engine power settings required to perform the tasks of agricultural. The system rapidly detects any vehicle malfunctions to prevent damage and an increase in long-term fuel consumption [2].

4. The use of a simulation model to evaluate the effect of vehicle malfunctions on fuel consumption

Most tractor malfunctions affect fuel consumption. Fuel consumption is one of the key diagnostic parameters in evaluations of a vehicle's technical condition. The magnitude of changes in fuel consumption varies subject to the type of defect. Certain malfunctions significantly increase fuel consumption in a wheeled tractor, including induction system leaks, turbocharger failure, combustion problems and fuel injector damage. Defects such as engine overheating result in a medium increase in fuel consumption. Malfunctions which lead to a minor increase in fuel consumption include incorrect oil pump pressure, significant bushing clearance in the crankshaft and piston assembly, loss of gear oil. Some defects increase fuel consumption subject to engine load, including wear of piston rings, loose sockets and values, oil pump damage, overheating or slipping of friction clutch. The effects of simulated malfunctions of a wheeled tractor on its fuel consumption are discussed in this part of the paper.

Gear tooth damage can be identified based on the following symptoms: deteriorating performance of the gearbox and the power transmission system, increased fuel consumption during attempts to maintain the same wheel torque and identical performance parameters (Fig.5 and 6).

$$\eta_{UN} = \eta_{sprz.} \cdot \eta_{skrz.pl.} \cdot \eta_{skrz.przek.} \cdot \eta_{red} \cdot \eta_{przek.gl.} \cdot \eta_{zwol.}$$

$$G_V = \frac{g_e}{1000 \cdot \rho_p} \left(F_{aero} + F_g + F_{bez} + F_{rr} + F_{wew} \right) \frac{V}{\eta_{UN}}$$

Fig. 5. Formula presenting the effect of gear tooth damage on changes in fuel consumption

Fig. 6. Diagnostic symptoms of gear tooth damage

Induction system leaks and turbocharger failure are identified based on the following symptoms: reduced engine performance, lower torque values, increased specific fuel consumption and increased fuel consumption during attempts to maintain the same wheel torque and identical performance parameters (Fig. 7–9).

Malfunctions of the fuel dosing system are diagnosed based on the following symptoms: reduced engine performance, reduced torque, increased specific fuel consumption and increased fuel consumption during attempts to maintain the same wheel torque and identical performance parameters (Fig. 10 and 11).

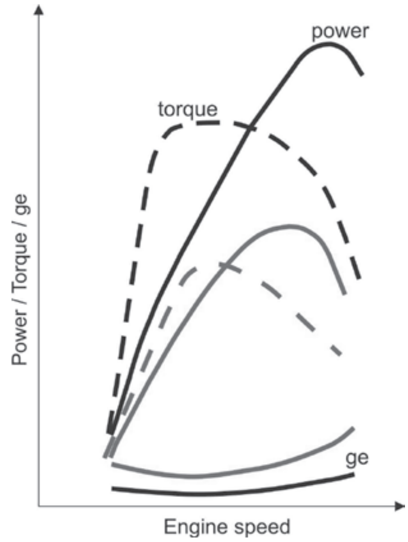


Fig. 7. The effect of induction system leaks and turbocharger failure on engine performance

$$G_V = \frac{g_e}{1000 \cdot \rho_p} \left(F_{aero} + F_g + F_{bez} + F_{rr} + F_{wew} \right) \frac{V}{\eta_{UN}}$$

Fig. 8. Formula presenting the effect of induction system leaks and turbocharger failure on engine performance

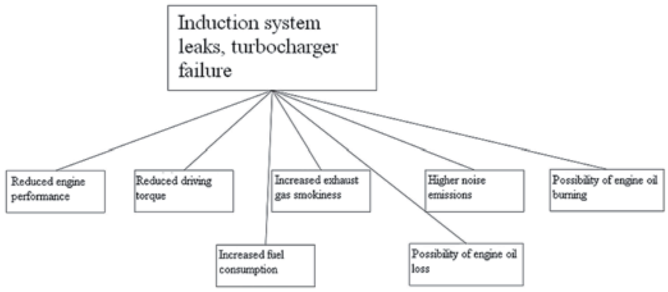


Fig. 9. Diagnostic symptoms of induction system leaks and turbocharger failure

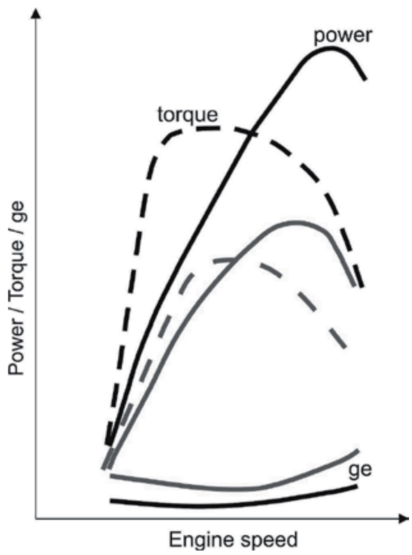


Fig. 10. The effect of fuel dosing system malfunctions on engine performance (black – optimally functioning engine, red – malfunctioning engine)

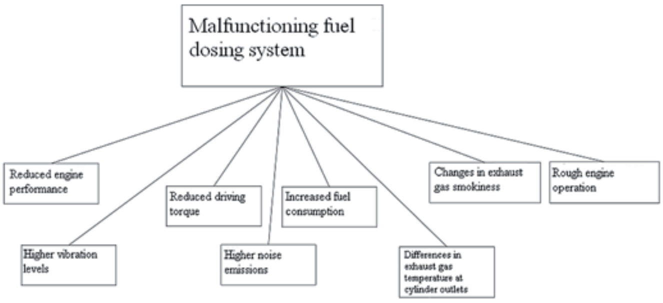


Fig. 11. Diagnostic symptoms of fuel dosing system malfunctions

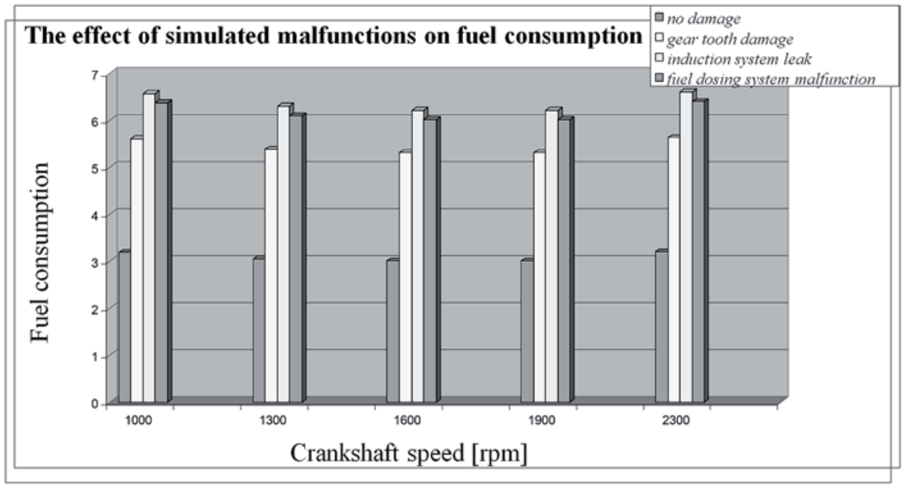


Fig. 12. Correlations between changes in fuel consumption and the simulated malfunctions: no damage, gear tooth damage, induction system leak, fuel dosing system malfunction.

Changes in hourly fuel consumption at different engine speeds during the simulated malfunctions are presented in Figure 12.

6. Conclusions

The presented simulation model supports the development of a diagnostic system which identifies four categories of defects in a wheeled tractor.

Simulation models can identify functional defects, exhaust defects, defects that jeopardize safety and defects that affect performance. The above has been illustrated on the example of several simulations.

Developed diagnostic model enables the identification of the technical condition of wheeled tractors by identifying changes in fuel consumption which can be used to control operation of the vehicle. Fuel consumption is the primary diagnostic parameter in identifying the condition of the vehicle.

Simulation models support the selection of the most appropriate software and hardware for a wheeled tractor's diagnostic system.

References

1. Alvarez I, Huet S, Automatic diagnosis of engine of agricultural tractors: The BED experiment, BIOSYSTEMS ENGINEERING 100 (2008).
2. Carrera González A, Alonso García S, Gómez Gil J. Design, Development and Implemementation of a Steering Controller Box for an Automatic Agricultural Tractor Guidance System, Using Fuzzy Logic, Technological Developments in Education and Automation, 2010.
3. Catalan H, Linares P, Mendez V. Tractor_PT: A traction prediction software for agricultural tractors, Computers and Electronics in Agriculture, Vol. 60, 2008.
4. Dyer J, Desjardins R. Carbon Dioxide Emissions Associated with the Manufacturing of Tractors and Farm Machinery in Canada. Biosystems Eng. 93 2006: 107-118.
5. Dyer J, Desjardins R. Simulated Farm Fielwork, Energy Consumption and Related Greenhouse Gas Emissions in Canada. Biosystems Eng. 85 2003: 503-513.
6. Gao L, Ren Z, Tang W, Wang H, Chen P, Intelligent Gearbox Diagnosis Methods Based on SVM, Wavelet Lifting and RBR, Sensors 2010.
7. Kolator B, Białobrzewski I. A simulation model of 2WD tractor performance, Computers and Electronics in Agriculture, Vol. 76, 2011.
8. Kumar A. Pandey KP. A device to measure dynamic front wheel reaction to safeguard rearward overturning of agricultural tractors, Computers and Electronics in Agriculture, Vol. 87, 2012.
9. Lindgren M. A Transient Fuel Consumption Model for Non-road Mobile Machinery Biosystems Eng. 91 2005: 139-147.
10. Michalski R, Gonera J, Janulin M, Arendt R. Structural analysis of a wheeled tractor oriented towards damage diagnostics; Monografie, studia, rozprawy nr M29; Wydawnictwo Politechniki Świętokrzyskiej; Kielce 2012.
11. Michalski R, Gonera J. Monitoring the operation of a wheeled tractor; Zeszyty naukowe Nauki Techniczne – Budowa i eksploatacja maszyn nr 17; Wydawnictwo Politechniki Świętokrzyskiej; Kielce 2012.
12. Michalski R, Janulin M. 2012, Analiza diagnostyki użytkowej ciągników kołowych, Monografie, studia, rozprawy nr M29.
13. Michalski R. Mechatroniczny system monitorowania stanu ciągnika kołowego. Biuletyn WAT Vol. LX, nr 1. 2011.
14. Michalski R. pod red.: Diagnostyka maszyn roboczych. Wyd. ITE Radom-Olsztyn, 2004.
15. Pranav PK. Pandey K.P., Tewari V.K. Digital wheel slipmeter for agricultural 2WD tractors, Computers and Electronics in Agriculture 73 2010.
16. Pranav P.K., Tewari V.K., Pandey K.P., Jha K.R. , Automatic wheel slip control system in field operations for 2WD tractors, Computers and Electronics in Agriculture 84 (2012).
17. Rautiainen R.H., Reynolds S.J., Mortality and morbidity in agriculture in the United States, Journal Agricultural Safety Health, Vol. 8 (3), 2002.

18. Singh CD, Singh RC. Computerized instrumentation system for monitoring the tractor performance in the field, Journal of Terramechanics 48 (2011).
19. Żebrowski J. Traction efficiency of a wheeled tractor in construction operations, Automation in Construction, Vol. 19, 2010.

Ryszard MICHALSKI
Jarosław GONERA
Michał JANULIN
Department of Vehicle and Machine Construction and Operation
Faculty of Technical Sciences
University of Warmia and Mazury in Olsztyn, Poland
E-mails: Ryszard.Michalski@uwm.edu.pl,
Jaroslaw.Gonera@uwm.edu.pl, Michal.Janulin@uwm.edu.pl

Article citation info:
BURDZIK R. Implementation of multidimensional identification of signal characteristics in the analysis of vibration properties of an automotive vehicle's floor panel. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 458–464.

Rafał BURDZIK

IMPLEMENTATION OF MULTIDIMENSIONAL IDENTIFICATION OF SIGNAL CHARACTERISTICS IN THE ANALYSIS OF VIBRATION PROPERTIES OF AN AUTOMOTIVE VEHICLE'S FLOOR PANEL

IMPLEMENTACJA WIELOWYMIAROWEJ IDENTYFIKACJI CHARAKTERYSTYCZNYCH CECH SYGNAŁU W ANALIZIE WŁASNOŚCI DRGANIOWYCH PANELU PODŁOGOWEGO POJAZDU SAMOCHODOWEGO*

The article provides a proposal of software application of a method and an algorithm developed to identify signal characteristics in the analysis of vibration properties of an automotive vehicle's floor panel. Due to the complexity resulting from nonlinear and random nature of vibration phenomena in automotive vehicles, the analysis in question is multidimensional. The property table being established consists of numerous measures and estimators, both dimensional and dimensionless ones, in the domains of amplitudes, time, frequency and time-frequency. The foregoing enables observation and separation of signal components in multiple domains, but it also makes it possible to define signal measures depending on stationary and non-stationary characteristics as well as accurate time positioning of resonant frequencies. Multicriterial approach to identification of vibration enables determining the table of vibration properties measures of floor panel. The table is numerical form of characteristics properties of the vibration signal.

Keywords: *vibration signal processing, wavelet transform, FFT.*

W artykule przedstawiono programową aplikację opracowanej metody i algorytmu matematycznego identyfikacji charakterystycznych cech sygnału w analizie własności drganiowych panelu podłogowego pojazdu samochodowego. Z uwagi na złożoność, wynikającą z nieliniowości i losowości, zjawisk drganiowych w pojazdach samochodowych analiza ma charakter wielowymiarowy. Wyznaczana tabela właściwości składa się z wielu miar i estymatorów wymiarowych i bezwymiarowych w dziedzinach amplitud, czasu, częstotliwości i czasowo-częstotliwości. Pozwala to na obserwację i separację składowych sygnału w wielu dziedzinach. Umożliwia definiowanie miar sygnału w zależności od cech stacjonarności i niestacjonarności oraz precyzyjną lokalizację czasową częstotliwości rezonansowych. Wielokryterialne podejście do identyfikacji drgań umożliwia wyznaczenie zbioru właściwości drganiowych panelu podłogowego, który jest numerycznym odzwierciedleniem charakterystycznych cech sygnału drgań.

Słowa kluczowe: *analiza sygnałów drganiowych, transformata falkowa, FFT*

1. Introduction

The vehicle vibration are results from many kind of dynamic interactions. The proper identification of the vibration is very difficult research and scientific problem. It requires good knowledge fundament and correct measurement tools and signal processing. An automotive vehicle, being a complex mechanical system, includes a set of specific free vibrations frequencies depending on the direction of the oscillatory wave propagation. From the most general perspective of vibration phenomena that one may consider, what matters most is the free vibration frequency bands for both sprung and unsprung masses, arranged in a vertical direction. Various publications mention different ranges for these resonant bands. The free vibration frequency of an automotive vehicle's sprung masses is assumed to be contained within the range from 1 to 2.5 [Hz]. Such dynamics of vibration phenomena does not essentially exert any negative effects on passengers, since it corresponds to man's natural frequency of making steps. Vibrations of the frequency below 1 [Hz] cause effects similar to seasickness in people, whereas those of the frequency exceeding 2.5 [Hz] bring prompt weariness and pain. The first resonant frequency for a man in a sitting position comes to ca. 4–6 [Hz] depending on individual body build features [14]. Input functions with the frequency of 3–4 [Hz] trigger strong vibrations in the abdominal cavity organs. The amplitude maximisation of the effects caused by these vibrations occurs at the frequency of 5–8 [Hz]. Close to these frequencies are those causing resonance in a human chest (i.e. 7–8 [Hz]). Organs of the head resonate in the band of 20–30 [Hz], whereas eyeballs at 60–90 [Hz]. However, it is the nervous as well as the cardiovascular system that are the most sensitive to the whole organism vibrations. The responses of these systems and their respective organs manifest themselves in their functions being disturbed, in poor physical and mental state, and even in certain forms of damage on higher amplitudes of effects and long exposure times. Some interesting investigation on influence of chosen driving parameters on vibration comfort according to Human-Vehicle-Road (HVR) model and vibration exposure metric described in the ISO 2631 have been presented in [18]. In a wide variety of transport environments the vibration transmitted through seats is associated with discomfort [14]. Seats can either reduce vibration discomfort or increase vibration discomfort [29]. The paper [29] presents results of the study on determine how factors, as age, gender, physical characteristics, backrest contact, and magnitude of vibration affect seat transmissibility. The paper presents analysis of the vibration registered on vehicle floor panel in location when it penetrate to the human organism via feet. Based on empirical studies, resonant

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

phenomena at higher frequencies, even exceeding 5 [Hz], have been identified, namely those which may cause considerable discomfort. In terms of unsprung masses, free vibration frequencies assume values within a range from several to more than a dozen hertz (i.e. 8–18 [Hz]). While an automotive vehicle is moving, free vibrations of sprung and unsprung masses occur simultaneously and overlap. Designers of automotive vehicles in mass production strive to limit the vibrations of sprung masses, trying to maintain sufficient rigidity of the suspension system at the same time, so that suitable steerability is ensured [1, 3, 4, 6, 7, 11, 16, 19, 22, 30]. Consequently, material properties and metallurgical technologies applied in the automotive industry are gradually growing in importance [2, 12, 13, 17, 23, 31] with the many analysis on influence of some parameters on physical and chemical properties [9, 10, 15].

As the results of observing and acquisition of vibration phenomena are received signals of displacement, velocity or acceleration of vibration. A vibration signal is a carrier of information on the state, the changes or the process to which the given physical or technical system is subject [24, 33]. Vibroacoustic signals are characterised by the largest information carrying capacity and they enable observation of changes occurring in a broad frequency band.

Numerous measuring problems may be considered on a general level of a signal, perceiving the signal as an entirety in the course of observation. They may be examined in the domains of amplitudes, time and frequency [8]. As far as random vibration phenomena are concerned, the signals recorded are of non-stationary nature which requires that the signal distribution is observed in the domains of time and frequency simultaneously. However there are some methods, for example as conjugate-pair decomposition (CPD) for signal decomposition, dynamics characterization, and nonlinearity identification in the time domain only [26]. The paper [25] presents novel time–frequency signal processing methodology based on Hilbert–Huang transform (HHT) and a new conjugate-pair decomposition (CPD) dedicated for characterization of nonlinear normal modes and parametric identification of nonlinear multiple-degree-of-freedom dynamical systems.

A signal is represented in the domain of frequency by application of the discrete Fourier transform. In the sphere of signal processing, it is mainly used to transform the *y(t)* function, being continuous in the domain of time, into the *Y(f)* function, continuous in the domain of frequency. The discrete Fourier transform is based on an assumption that every signal may be obtained by adding sinusoid properties with appropriate phases and amplitudes. Therefore, a result of the discrete Fourier transform may be interpreted as a set of properties of the signal being examined in the function of frequency of component sinusoids [20]. The fast Fourier transform (FFT) is more frequently applied in practice, since it is a computational algorithm of the discrete Fourier transform as well as of an inverse transform, making use of the sine function symmetry.

In the field of technical diagnostics, time implementations of physical quantities may be perceived as a sum of two components: the determined and the random one. It is assumed that the determined component carries information on the wear of the given device being examined, whereas the random one is a measure of noises and interferences. The only data relevant from the technical diagnostics' perspective are those contained in the determined component, and the data must necessarily be separated [21, 27, 28, 30, 32]. One of the mathematical instruments enabling separation of non-stationary signal components is a wavelet transformation which consists in distinguishing a part of the *f(t)* signal being similar to a present template, i.e. the part which corresponds to the determined component. The template role is performed by basic wavelet $\psi(t)$. The wavelet functions as a transformation kernel. A single wavelet is used in the given transformation, however, due to modification of scale coefficient a and modification coefficient b, it forms what is referred to as a *wavelet family*. A continuous wavelet transform in the domain of time and frequency is defined as follows:

$$\tilde{s}_{\Psi}(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} s(t)\Psi\left(\frac{t-b}{a}\right)dt \tag{1}$$

gdzie:

a – scale coefficient,
b – modification coefficient,
s(t) – value of the signal examined in the function of time,
$\tilde{s}_{\Psi}(a,b)$ – wavelet coefficient dependent on a and b,
ψ – wavelet function,
$\Psi((t-b)/a)$ – transformation kernel.

The value of wavelet coefficient $\tilde{s}_{\Psi}(a,b)$ established by means of the above formula is generally understood as a measure of similarity between the signal examined and the chosen wavelet [20].

Furthermore, due to dimensional estimates' sensitivity to the stationary nature of operating conditions, in the process of identification of signal characteristics, besides dimensional estimates, one applies quotients of these measures being dimensionless amplitude discriminants. They are obtained by dividing moments of various ranks by one another.

3. Method of multidimensional identification of vibration signal characteristics of an automotive vehicle's floor panel – WSA WIBROCAR

For the sake of identification of signal characteristics in the analysis of vibration properties of an automotive vehicle's floor panel, a complex mathematical algorithm was developed to be subsequently implemented in the MatLab environment, and a user interface was created named WIBROCAR. The programme developed was given the name of WSA, and it was then extended with several modules dedicated to analysis, monitoring and diagnostics of selected vehicle systems and structural assemblies. Procedure of testing is starting by vehicle data and research parameters entry (Fig. 1).

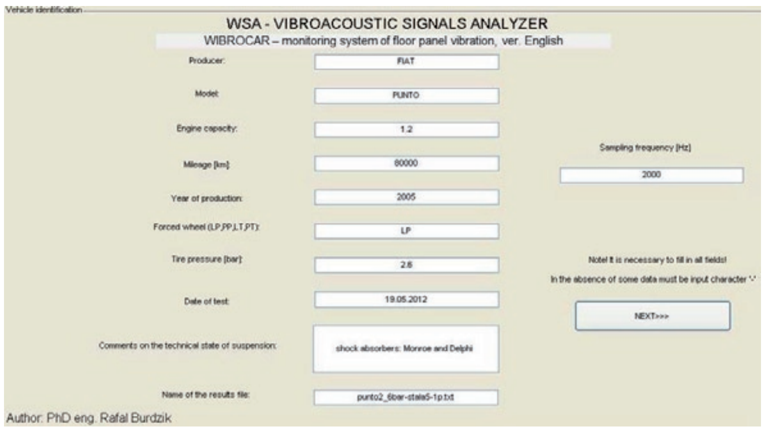


Fig. 1. First window of WSA program

The implementation of the WSA program was assumed the utilitarian of the software. For this purpose it is very important to communicate to the user with clear orders and information reports. The work in the WSA should be close to intuitive. Some examples of the communication windows have been depicted in Figure 2.

Due to the complexity resulting from nonlinear and random nature of vibration phenomena in automotive vehicles, the analysis in ques-

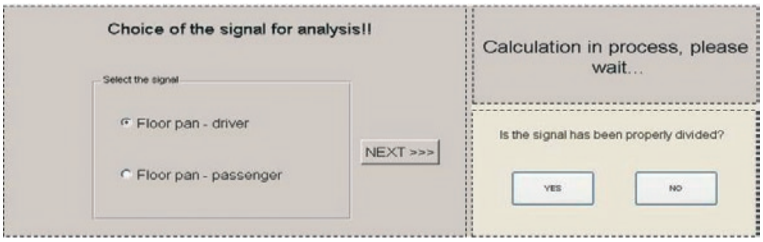


Fig. 2. User – WSA program communication windows

tion is multidimensional. The property table being established consists of numerous measures and estimators, both dimensional and dimensionless ones, in the domains of amplitudes, time, frequency and time-frequency. In order to accurately identify signal characteristics, one needs appropriate analytical methods depending on the stationary and non-stationary nature of the signal. An automatic algorithm was developed for positioning of stationary and non-stationary signal cycles. For this purpose identification of next cycles of forced machine working there were next phases of vibration inductor working identification measure formulated. The markers of next cycles of forced machine working measures based on STFT (Short Time Fourier Transform) transformation were used. The main reason of choosing this transformation was short realization time. There was 21-22 Hz frequency band cut out from STFT spectrum for analysis. Based on time function of cut off frequency band identifying algorithm of end of stand run up and start of stand coasting time coordinates was created. Elaborated algorithm is based on comparing next value of analysed frequency band ("analysis of edge") around set parameters. Locating of end of stand run up and start of stand coasting enables to divide signal on three time windows. First window for fragment of signal growing according to constant frequency increase of the forced system. Second window for signal with constant frequency and the third one for coasting stand – decrease of signal amplitudes according to constant frequency decrease of the forced system. This method and algorithm has been depicted in Figure below.

An example of such a division has been provided in Fig. 4. It is the very first step towards identification of signal characteristics using dedicated methods in the analysis of stationary and non-stationary signals.

For the purposes of analysis of the stationary signal part, an algorithm based on FFT was developed. The signal characteristics are then identified by amplitude based correlation of successive signal harmonics which have been accurately separated from non-stationary signal components. Results of this algorithm have been depicted in Figure 5. Preliminary tests of a car's floor panel proved various sensitivities to deviation of vibration damping parameters of successive harmonics from a constant input function.

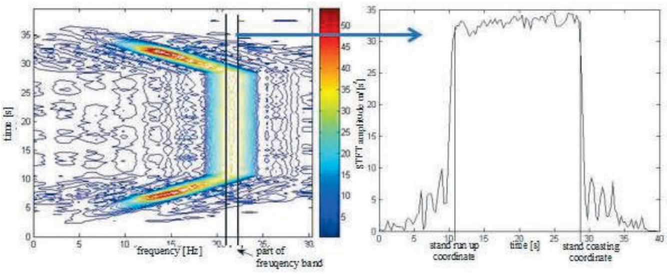


Fig. 3. Calculation and analysis of time function of STFT coefficients for identification of stationary and non-stationary parts of the signal

In order to analyse predominant components of resonant frequencies of sprung and unsprung masses, a transformation algorithm was developed for the non-stationary signals recorded during a rundown of the vibration forcing station and once it was completely shut down. Finally, for the purposes of identification of the signal characteristics, a vehicle free vibration suppression window was chosen, where the vibrations of a system subject to free suppression were recorded. It enabled the system's free vibration frequency bands to be accurately observed and defined. The window used to analyse and define the range

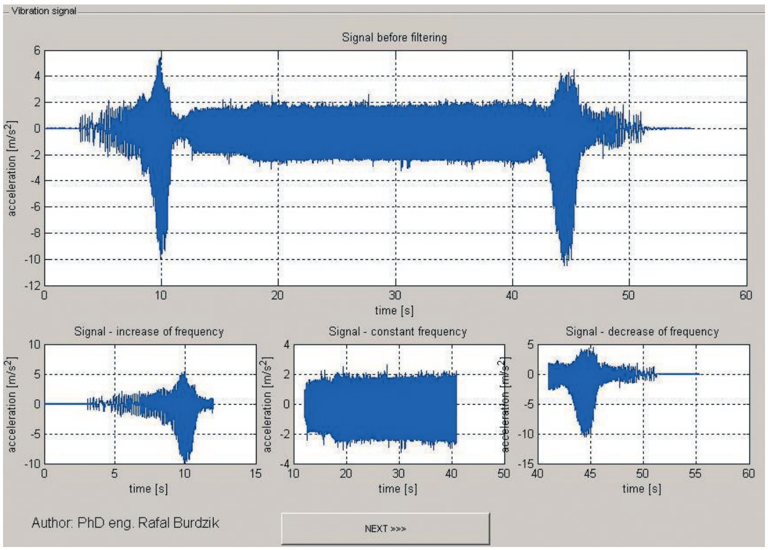


Fig. 4. Vibration of the floor panel - automatic algorithm for positioning of stationary and non-stationary signal cycles

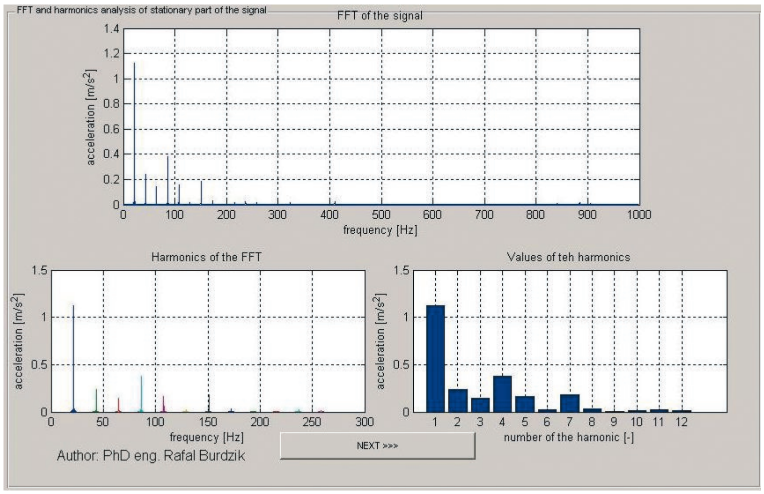


Fig. 5. Results of the FFT analysis for the stationary signal portion

of resonant frequency bands for sprung and unsprung masses has been provided in Fig. 6. The wavelet based time and frequency distribution of a signal enables accurate definition of resonant windows.

75-elemnent matrices of measures of signal characteristics were used as a multi-parameter measure of signal characteristics for an automotive vehicle's floor panel. They were established as estimators based on averaged time and frequency courses of resonant windows for sprung and unsprung masses (Fig. 7).

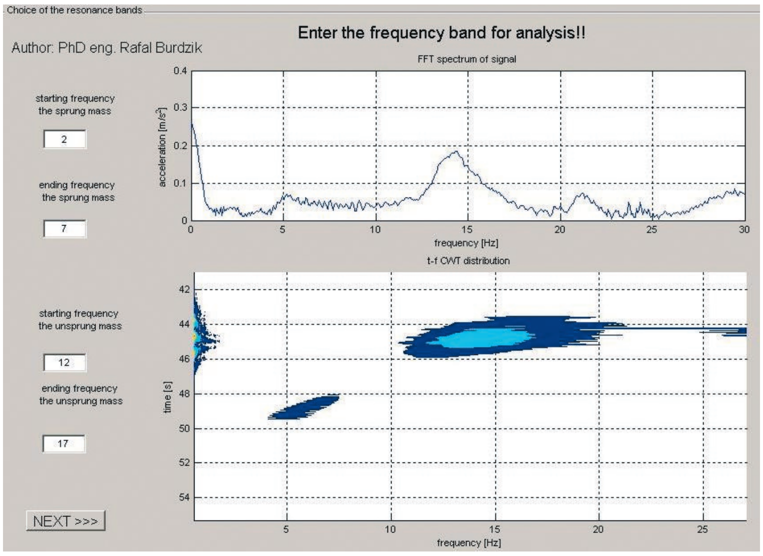


Fig. 6. Identification of resonance frequency bands – non-stationary signal portion

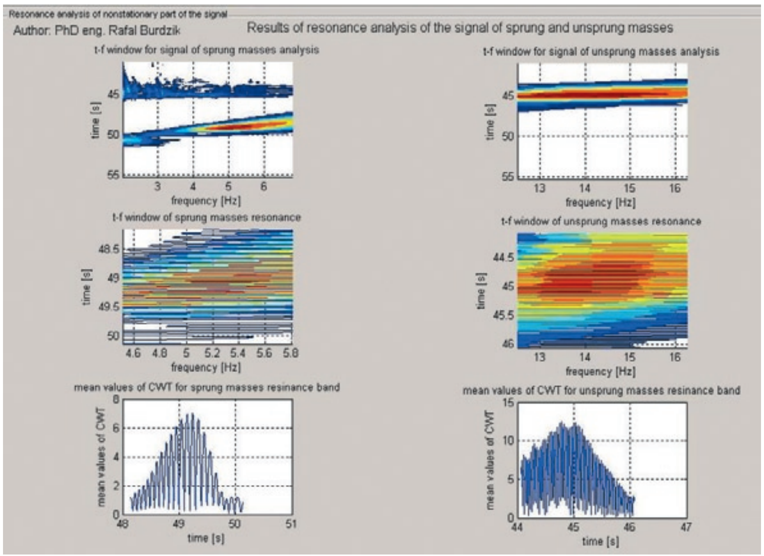


Fig. 7. Time and frequency resonance windows and averaged courses of resonance for sprung and unsprung masses

4. Table of properties of floor panel vibration

The method of multidimensional identification of vibration signal characteristics, described in previous chapter, allows to determine table of properties of an automotive vehicle's floor panel. The complicated vibration phenomena and random character of excitation forces acting on car vehicle determine to use many estimators to define vibration occurring in the car. The described method enables determining measures of signal distribution in time, frequency and time-frequency in terms of stationary and non-stationary parts of the signal.

The tables below contain a collation of the chosen estimators of vibration characteristics of an automotive vehicle's floor panel featuring built-in shock absorbers filled with working medium in 50%. These measures form 75-element table of measures of signal characteristics. From the time realization of acceleration of vibration registered during slowing of excitation, when the mechanical system goes by resonance frequencies bands of sprung and unsprung masses of the vehicle the 16 global estimators have been determined (tab. 1).

Based on the preliminary experimental research it was specified that stationary part of the vibration signal, during excitation force with constant frequency, is sensitive on changes of technical condition of car suspension. Thus for the vibration properties table were added estimators calculated on spectrum of vibration as 12th next harmonics values. The values of those estimators for the same case study (shock absorbers filled with working medium in 50%) have been presented in Table 2.

Some extra "control" estimators of identification of resonance occurring in time and frequency domains for sprung and unsprung masses of vehicle have been added to the table (tab. 3). The values can change for different technical parameters of the suspension system (masses, stiffness).

For the precise time-frequency characteristics of the resonance windows, according to the methodology described in chapter 3, the estimators of CWT (Continuous Wavelet Transform) have been determined. Time and value of the exposure on resonance vibration have been determined separately for sprung and unsprung masses. The tables below contain a collation of the chosen estimators of vibration determined from resonance distribution of CWT. Those estimators have been added to the table of properties of floor panel vibration.

Based on the previous research some extra estimators have been proposed to the table of properties of floor panel vibration. The relative (total) estimators of CWT distribution between resonances of sprung and unsprung masses have been presented in Table 6. Those are the measurements of representation of the relation of vibration characteristics of sprung and unsprung masses. Those estimators have been defined as below.

C_w – half of the sum of maximum values of amplitude of CWT of unsprung masses resonances (unsprung resonance P2P – scope range measurement):]

Table 2. Spectrum of the vibration estimators (stationary signal)

FFT estimators					
1th harm.	2nd harm.	3rd harm.	4th harm.	5th harm.	6th harm.
1,121	0,242	0,142	0,378	0,159	0,019
7th harm.	8th harm.	9th harm.	10th harm.	11th harm.	12th harm.
0,186	0,034	0,007	0,017	0,027	0,016

Table 1. Global estimators of time realization of vibration

Global estimators (amplitude, time) – resonance window					
max	skewness	kurtosis	play factor	root amplitude	standard deviation
2,951	-2,533	14,072	-30,490	0,004	1,157
shape factor	P2P	peak factor	impulsivity factor	RMS	momentum 1
-10,206	5,663	4,229	-43,166	1,339	0,000
correlaction	variance	covariance	median		
1,000	1,339	1,339	0,002		

Table 3. Estimators of resonances location

Estimators of value and location of the resonances					
sprung masses			**unsprung masses**		
max value	time	frequency	max value	time	frequency
7,511	49,142	5,078	13,909	45,072	13,542

Table 4. Collation of estimators of sprung masses resonance distribution of CWT

Estimators of resonance distribution of CWT – sprung masses window					
max	skewness	kurtosis	play factor	root amplitude	standard deviation
6,995	0,800	2,437	1,642	1,483	1,900
shape factor	P2P	peak factor	impulsivity factor	RMS	momentum 1
1,479	3,457	0,960	1,420	3,601	0,000
correlaction	variance	covariance	median	integral of average CWT	mean/max
1,000	3,610	3,610	1,703	4,883	0,698

Table 5. Collation of estimators of unsprung masses resonance distribution of CWT

Estimators of resonance distribution of CWT – unsprung masses window					
max	skewness	kurtosis	play factor	root amplitude	standard deviation
12,512	0,246	1,938	0,726	7,591	3,357
shape factor	P2P	peak factor	impulsivity factor	RMS	momentum 1
2,040	6,160	0,548	1,118	11,239	0,000
correlaction	variance	covariance	median	integral of average CWT	mean/max
1,000	11,267	11,267	5,283	11,048	0,883

Table 6. Relative dimensionless estimators of the relation of CWT vibration characteristics of sprung and unsprung masses

Dimensionless relative estimators (CWT)				
C_w	L	E_{sr}	E_{max}	E_w
6,352	0,726	7,946	19,507	4,910

$$C_w = \frac{Wz_{\max} + Wz_{\min}}{2} \tag{2}$$

where:

$Wz_{\max}$ – maximum value of the average of CWT distribution for the unsprung masses resonance window,

$Wz_{\min}$ – minimum value of the average of CWT distribution for the unsprung masses resonance window.

L – play factor of average of CWT distribution for the unsprung masses resonance window:
L – play factor of average of CWT distribution for the unsprung masses resonance window:

$$L = \frac{\overline{w}}{(\frac{1}{n}\sum |w_i|^{\frac{1}{2}})^2} \tag{3}$$

where:

w_i –average of CWT distribution for the unsprung masses resonance window,

n – number of samples of CWT distribution average values.

E_{sr} – sum of the average of CWT distribution for the sprung and unsprung masses resonance windows:

$$E_{sr} = Wz_{sr} + Wn_{sr} \tag{4}$$

where:

Wz_{sr} – mean value of CWT distribution for the unsprung masses resonance window,

Wn_{sr} – mean value of CWT distribution for the sprung masses resonance window.

E_{max} – sum of maximum values of the average of CWT distribution for the sprung and unsprung masses resonance windows:

$$E_{\max} = Wz_{\max} + Wn_{\max} \tag{5}$$

where:

$Wz_{\max}$ – maximum value of average of CWT distribution for the unsprung masses resonance window,

$Wn_{\max}$ – maximum value of average of CWT distribution for the sprung masses resonance window.

E_w – concentration coefficient of the average of CWT distribution for the resonance windows:

$$E_w = \frac{\frac{E_{max}}{E_{sr}}}{2} \quad (6)$$

For the conclusion it can be stated that the table of properties of floor panel vibration is collected from estimators determined from time realization of the vibration, spectrum and time-frequency distribution of the vibration. Exemplary structure of those table have been presented in Table 7. It represents the vibration estimators calculated on the results of the research of the real object, as passenger car with shock absorbers filled with 50% of fluid volume. The colour of the next values represents the estimators presented in tables 1–6.

Table 7. Table of properties of floor panel vibration

2,951	0,002	5,078	0,000	0,548
-2,533	1,121	13,909	1,000	1,118
14,072	0,242	45,072	3,610	11,239
-30,490	0,142	13,542	3,610	0,000
0,004	0,378	6,995	1,703	1,000
1,157	0,159	0,800	4,883	11,267
-10,206	0,019	2,437	0,698	11,267
5,663	0,186	1,642	12,512	5,283
4,229	0,034	1,483	0,246	11,048
-43,166	0,007	1,900	1,938	0,883
1,339	0,017	1,479	0,726	6,352
0,000	0,027	3,457	7,591	0,726
1,000	0,016	0,960	3,357	7,946
1,339	7,511	1,420	2,040	19,507
1,339	49,142	3,601	6,160	4,910

The proper conclusion based on the such large data collection is very difficult. Thus the paper [5,7] presents some application of neural networks as classifier or input module for the control system of vibration absorbing elements in vehicle structure. The scheme of the conception of those system have been presented in the Figure below.

5. Conclusion

Analysis and evaluation of the vibration phenomena in car vehicles are very difficult and it requires using of proper methods and mathematics algorithms. The number of physics and chemical phenomena occurring during working of many systems of vehicles which are affecting on propagation of energy in different forms [9, 10, 15].

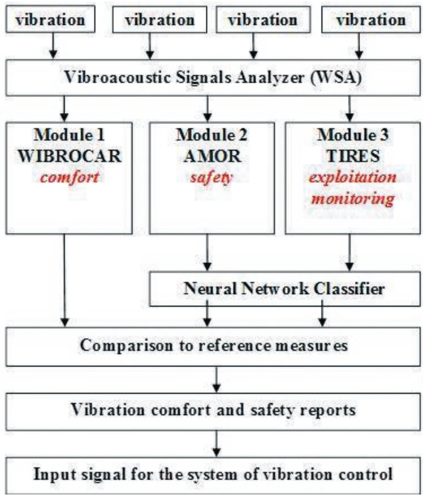


Fig. 8. Scheme of the modular conception of the monitoring and control system of vibration comfort and safety of the passenger car

Thus research on this kind of phenomena has to be conduct and the results and developed methods should be analysed for different parameters of mechanical systems working. The paper presents method verified for different exploitation parameters of the vehicle.

The method proposed and described in the article for multidimensional identification of signal characteristics in the analysis of vibration properties of an automotive vehicle's floor panel enables observation and separation of signal components in various domains. It also makes it possible to define signal measures depending on stationary and non-stationary characteristics as well as accurate time positioning of resonant frequencies. Further conclusions and assessments may rely on selected measures having the properties of state symptoms or may be achieved by means of neural algorithms to function as input databases for a neural network. The measures applied in the table of signal characteristics determine a range of properties such a dynamics, amplification, scattering, concentration, attenuation, stability etc.

The described software implementation of those method has the utilitarian character. WSA program is provided in friendly user interface. The results as table of properties of floor panel vibration could be adopted as mapping input signal to system of monitoring and control of vibration.

References

1. Belgacem W, Berry A, Masson P. Active vibration control on a quarter-car for cancellation of road noise disturbance. Journal of Sound and Vibration 2012; 331: 3240–3254.
2. Blacha L, Burdzik R, Smalcerz A, Matuła T. Effects of pressure on the kinetics of manganese evaporation from the Ot4 Alloy. Archives of Metallurgy and Materials 2013; 58(1): 197-201.
3. Borowiec M, Hunicz J, Sen A, Litak G, Koszalka G, Niewczas A. Vibration of a vehicle excited by real road profiles, Forschung im Ingenieurwessen 2010; 74: 99–109.
4. Burdzik R, Gardulski J. Frequency analysis decimation vibration signals of passenger car's suspensions. Transport Problems 2007; 2(1): 23-29.
5. Burdzik R, Peruń G, Warczek J. Possibilities of using vibration signals for the identification of pressure level in tires with application of neural networks classification. Key Engineering Materials 2014; 588: 223-231 (Online available since 2013/Oct/11 at www.scientific.net)
6. Burdzik R. Material vibration propagation in floor pan. Archives of Materials Science and Engineering 2013; 59(1): 22-27.
7. Burdzik R. Monitoring system of vibration propagation in vehicles and method of analysing vibration modes. J. Mikulski (Ed.): TST 2012, Springer, Heidelberg, CCIS 2012; 329: 406-413.

8. Chaari F, Bartelmus W, Zimroz R, Fakhfakh T, Haddar M. Gearbox vibration signal amplitude and frequency modulation. Shock and Vibration 2012; 19(4): 635-652.
9. Dzida M, Jężak S, Sumara J, Żarska M, Góralski P. High pressure physicochemical properties of biodiesel components used for spray characteristics in diesel injection systems. Fuel 2013: 111: 165-171.
10. Dzida M, Prusakiewicz P. The effect of temperature and pressure on the physicochemical properties of petroleum diesel oil and biodiesel fuel. Fuel 2008; 87: 1941–1948.
11. Engel ZW, Kowalski P. Investigation of the influence of simultaneous vibroacoustic exposures on the operator. Journal of the Theoretical and Applied Mechanics 2008; 46(4): 799–811.
12. Folęga P, Siwiec G. Numerical analysis of selected materials for flexsplines. Archives of Metallurgy and Materials 2012; 57(1): 185-191.
13. Fornalczyk A, Saternus M. Platinum recovery from used auto catalytic converters in electrorefining process. Metalurgija 2012; 52(2): 219-222.
14. Griffin M.J. Handbook of Human Vibration. Academic Press Ltd, New York, 1990.
15. Kaźmierczak-Bałata A, Bodzenta J, Wokulska K, Kucytowski J, Szperlich P, Łukasiewicz T, Hofman B. Analysis of influence of Yb concentration on thermal, elastic, optical and lattice parameters in YAG single crystal. Journal of Alloys and Compounds 2009; 473: 245-249.
16. Korzeb J, Nader M, Rózowicz J. Review and estimation of traffic generated vibration developed in proximity of Warsaw subway line. 12th International Congress on Sound and Vibration 2005: 5048-5055.
17. Kusiński J, Kac S, Kopia A, Radziszewska A, Rozmus-Górnikowska M, Major B, Major L, Marczak J, Lisiecki A. Laser modification of the materials surface layer - a review paper. Bulletin of the Polish Academy of Sciences Technical Sciences 2012; 60(4): 711-728.
18. Kuznetsov A, Mammadov M, Sultan I., Hajilarov E. Optimization of a quarter-car suspension model coupled with the driver biomechanical effects. Journal of Sound and Vibration 2011; 330: 2937–2946.
19. Lozia Z. Truck front wheels and axle beam vibrations. 5th Mini Conference on Vehicle System Dynamics, Identification and Anomalies(VSDIA 96). Budapest, Hungary 11-13 of November, 1996.
20. Lyons R.: Wprowadzenie do cyfrowego przekształcania sygnałów. Warszawa: Wydawnictwa Komunikacji i Łączności, 2006.
21. Michalski R, Wierzbicki S. An analysis of degradation of vehicles in operation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2008; 1(37): 30-32.
22. Nader M. Influence of mechanical vibration on the human body in the means of transport and its modeling. Archives of Transport – Archiwum Transportu 2000; 12(2): 33-53.
23. Oleksiak B., Blacha-Grzechnik A., Siwiec G. Application of the flotation process in the silver recovery from the wastes generated during the silvery semi-products manufacturing. Metalurgija 2012; 51(3): 298-300.
24. Oltean M, Picheral J, Lahalle E, Hamdan H, Griffaton J. Compression methods for mechanical vibration signals: Application to the plane engines. Mechanical Systems and Signal Processing 2013; 41: 313–327.
25. Pai PF. Time–frequency characterization of nonlinear normal modes and challenges in nonlinearity identification of dynamical systems. Mechanical Systems and Signal Processing 2011; 25: 2358–2374.
26. Pai P, Nguyen B, Sundaresan M. Nonlinearity identification by time-domain only signal processing. International Journal of Non-Linear Mechanics 2013; 54: 85–98.
27. Radkowski S, Gimiński R. Prognosis of gear fatigue crack using vibroacoustic signal. Acta Acustica united with Acustica 2010; 96: 81.
28. Radkowski S, Smalko Z, Pietak A, Woropay M. Use of bispectral analysis in condition monitoring of machinery. Structural Health Monitoring 2006: 627-634.
29. Toward M., Griffin M.J. The transmission of vertical vibration through seats: Influence of the characteristics of the human body. Journal of Sound and Vibration 2011; 330: 6526–6543.
30. Uhl T, Chudzikiewicz A, Karpiński J. Dynamic problems in rail vehicle design. Archives of Transport – Archiwum Transportu 2000; 12(1): 57-71.
31. Węgrzyn T, Wieszała R. Significant alloy elements in welded steel structures of car body. Archives of Materials and Metallurgy 2012; 57(1): 45-52.
32. Wilk A, Madej H, Figlus T. Analysis of the possibility to reduce vibroactivity of the gearbox housing. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2011: 42-49.
33. Zimroz R, Urbanek J, Barszcz T, Bartelmus W, Milioz F, Martin N. Measurement of instantaneous shaft speed by advanced vibration signal processing – application to wind turbine gearbox. Metrology and Measurement Systems 2011; 18(4): 701-711.

Rafał BURDZIK
Facylty of Transport
Silesian University of Technology
ul. Krasińskiego 8, 40-019 Katowice, Poland
E-mail: rafal.burdzik@polsl.pl

Article citation info:
LAI M-H. Optimal number of Minimal Repairs under a Cumulative Damage Model with Cumulative Repair Cost Limit. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 464–471.

Min-Tsai LAI

OPTIMAL NUMBER OF MINIMAL REPAIRS UNDER A CUMULATIVE DAMAGE MODEL WITH CUMULATIVE REPAIR COST LIMIT

OPTYMALNA LICZBA NAPRAW MINIMALNYCH W ŚWIETLE MODELU SUMOWANIA USZKODZEŃ PRZY OGRANICZONYM ŁĄCZNYM KOSZCIE NAPRAW

In this paper, we consider a repair number counting replacement policy under a cumulative damage model, in which the policy includes the concept of a cumulative repair cost limit. The system experiences two kinds of shocks: a type I shock causes a random amount of damage to the system leading to a serious failure when the total damage exceeds a failure level; or a type II shock causes the system into minor failure which can be corrected by minimal repair. When a minor failure occurs, the repair cost will be evaluated and minimal repair is executed if the accumulated repair cost is less than a predetermined limit L. The system is replaced anticipatively at n-th minor failure, or at the j-th minor failure (j < n) at which the accumulated repair cost exceeds a predetermined limit L, or any serious failure. In order to assess the performance of the proposed maintenance policy and to minimize the long-term expected cost per unit time, a mathematical model for the maintained system cost is derived. By minimizing that cost, the optimal number n is also verified finite and unique under certain conditions. Analyses based on numerical results are conducted to highlight the properties of the proposed maintenance policy in respect to the different parameters.*

***Keywords**: Cumulative damage model, Cumulative repair cost limit, Replacement policy, Minimal repair.*

W przedstawionym artykule omawiamy politykę wymiany systemu opartą na modelu sumowania uszkodzeń polegającą na obliczaniu liczby napraw. Polityka ta obejmuje koncepcję limitu łącznego kosztu napraw. System może być narażony na działanie dwóch rodzajów szkodliwych czynników: czynniki I-ego typu powodują losowo określony zakres uszkodzeń systemu, prowadząc do poważnej awarii, gdy łącznie uszkodzenia przekraczają poziom awarii; lub czynniki typu II-ego powodujące drobne uszkodzenia, które można skorygować poprzez minimalną naprawę. Gdy dochodzi do niewielkiego uszkodzenia, wtedy szacuje się koszt naprawy i realizuje minimalną naprawę, jeśli łączny koszt naprawy jest niższy od uprzednio ustalonego limitu L. System zostaje prewencyjnie wymieniony albo przy n-tej drobnej awarii albo przy j-tej drobnej awarii (j <n), przy której łączny koszt naprawy przekracza uprzednio ustalony limit L lub też przy jakimkolwiek poważnym uszkodzeniu. W celu oceny skuteczności proponowanej polityki obsługiwania i zminimalizowania przewidywanego długoterminowego kosztu przypadającego na jednostkę czasu, wyprowadzono model matematyczny kosztów dla obsługiwanego systemu. Poprzez minimalizację tych kosztów, określono również optymalną liczbę napraw n, która w pewnych warunkach jest liczbą skończoną i niepowtarzalną. W oparciu o wyniki numeryczne, przeprowadzono analizy mające na celu naświetlenie właściwości proponowanej polityki obsługiwania w odniesieniu do różnych parametrów.*

***Słowa kluczowe**: Model sumowania uszkodzeń, limit łącznych kosztów napraw, polityka wymiany, naprawa minimalna.*

1. Introduction.

Most production systems suffer increasing wear with usage or age and are subject to random failures resulting from this deterioration (Wang (2002)) and most of them are maintained or repairable systems. Moreover, for some systems, such as aircrafts, submarines, military systems, and nuclear systems, it is very important to maintain a system to prevent failures because it can be dangerous or disastrous. The growing importance of maintenance has generated an increasing interest in the development and implementation of optimal maintenance strategies for improving system availability, preventing the occurrence of system failures, and reducing maintenance costs of deteriorating systems. In the deteriorating system, the level of deterioration is represented by a degradation process, such as corrosion, wear out, material fatigue, and fatigue-crack-growth in engineering applications or markers of health status and quality of life data in medical settings. Cumulative damage models are often used to describe these above situations.

Cumulative damage models are a special class of mathematical models within reliability theory that describe the probability of failure of a given system under the impact of a damaging environment. The system suffers damage due to shocks and fails when the total amount of damage exceeds a failure level K, and it generates cumulative damage process. Recently, Nakagawa (2007) summarized sufficiently PM policies and their optimization problems for cumulative damage models. The replacement models where a system is replaced when the total damage exceeds a threshold level k can refer Feldman (1976), Nagakawa (1976) and Satow et al. (2000). The replacement models where a unit is replaced at a planned time T were proposed in Taylor (1975), Mizuno (1986), Nakagawa (1980, 2007), Qian *et al.* (1999) and Perry (2000). Furthermore, the replacement models where a system is replaced at shock N were proposed in Nagakawa (1984).

Nakagawa and Kijima (1989) considered a standard cumulative damage model with minimal repair at failure to obtain the optimal values T^*, N^*, and k^*, individually. Kijima and Nakagawa (1991) considered a cumulative damage shock model with imperfect PM policy. Satow and Nakagawa (1997) considered a modified cumulative damage model that the damage can be produced by shocks or increased with time at constant rate a. The optimal values T^*, N^*, and k^* are obtained individually. Qian *et al.* (1999) presented an extended cumulative damage model with two kinds of shocks: failure shock and damage shock. This model is applied to the backup of files in a database system and the optimal replacement period is obtained. Satow *et al.* (2000) considered a cumulative damage model with two types of damages that are both from external shocks and deterioration with time. The optimal threshold k^* is obtained.

Qian *et al.* (2003) considers an extended cumulative damage model with maintenance at each shock when the total damage does not exceed a failure level *K*; with minimal repair at each shock when the total damage exceeds a failure level *K*, and with replacement at time T or at failure *N*. The optimal values T^* and N^* are obtained. Qian *et al.* (2005) applied cumulative damage model for a used system with initial damage level. A unique optimal T^* or managerial level k^* which minimizes the expected cost rate are obtained. Ito and Nakagawa (2011) compared the standard cumulative damage model with two other cumulative damage models: (1) the amount of damage due to shocks is measured only at periodic time; and (2) the amount of damage increases linearly with time. This cumulative damage model can be applied to the garbage collection policies for a database system in Satow et al.(1996) and applied to obtain the optimal full and cumulative backup policies successfully for a database system in Qian et al. (1999, 2005). And, it is also applied to describe the cumulative damage of a fibrous carbon composite in Padgett(1998).

Zhao *et al.* (2012) considered a periodical replacement model that the unit is replaced at a planned time or when the total damage exceeds a failure level, whichever occurs first, and undergoes minimal repair when independent damage occurs. Furthermore, they considered a modified model that the total damage is measured at periodic times and increases approximately with time linearly. Zhao and Nakagawa (2012) considered age and periodic replacement last models with working cycles and applied this type of replacement policy to a standard cumulative damage model. Zhao *et al.* (2013) applied the notion of maintenance last to a standard cumulative damage model, in which the unit undergoes preventive maintenances before failure at a planned time *T*, at a damage level *k*, or at a shock number *N*, whichever occurs last.

With regard to repair-cost-limit policies allowing minimal repairs, Lai (2007) applied the concept of cumulative repair cost limit into replacement model that included the information of all repair costs to decide whether the system should be repaired or replaced. Following the work of Lai (2007), Chien, *et al.* (2009) extended the work of Lai (2007) by introducing the random lead time for replacement delivery. Chien, *et al.* (2010) modified the work of Chien, et al. (2009) by adding an age-dependent type of failure. Chang, *et al.* (2010) presented a model for determining the optimal number of minimal repairs before replacement. Chang, *et al.* (2013) modified the work of Chang, *et al.* (2010) by allowing an age-dependent failure type. Sheu, *et al.* (2010) presented a generalized model for determining the optimal replacement policy based on multiple factors (or more information) such as the number of minimal repairs before replacement and the cumulative repair cost limit.

In this study, we present a repair number counting replacement policy with cumulative repair cost limit where the system is subject to a cumulative damage model. The concept of cumulative repair cost limit adopts the entire repair cost history to make decision for repairing or replacing the system. The remainder of the paper is organized as follows: Section 2 presents the model formulation and optimization. In Section 3, the long-term expected cost per unit time $\overline{C}(n,L)$ is derived and the conditions characterize the optimal n^* is developed. A computational example is provided to demonstrate the above results in Section 4. Section 5 provides conclusions

2. Problem formulation

Assume that the system is subject to shocks which randomly occur according to a non-homogeneous Poisson process $\{N(t)\}_{t\geq 0}$ with intensity rate $\lambda(t)$. Whenever a shock occurs, it will be type-I shock with probability $p\ (0<p\leq 1)$ and type-II shock with probability $q\ (p+q=1)$. By using the decomposition theorem of Poisson process, it is noted that type-I and type-II shocks occur according to two non-homogeneous Poisson processes $\{N_1(t)\}_{t\geq 0}$ and $\{N_2(t)\}_{t\geq 0}$ with intensity rates $p\lambda(t)$ and $q\lambda(t)$, respectively. And, $N_1(t)$ and $N_2(t)$ denote the numbers of type-I and type-II shocks occurred during [0, *t*], respectively.

The type-I shocks whenever occur cause some damage to the system and these damages are additive. When a type-I shock occurs, a random amount D_i of damage from *i-th* type-I shock has a probability distribution $H(d)=P(D_i\leq d)$ and a finite mean μ_d, *i*=1,2,3,…. Then the accumulated damage to the the *j-th* type I shock after the installation $W_j=\sum_{i=1}^{j}D_i$ has a distribution function:

$$P(W_j\leq w)=H^{(j)}(w)=\begin{cases}1 & j=0\\ H_1*H_2*\cdots*H_j(w), & j=1,2,3,\cdots\end{cases}, \tag{1}$$

where the "*" mark is denoted the *Stieltjes* convolution of the distribution *H*(*d*) with itself. The probability of *j* type-I shocks in [0, *t*] is given by:

$$P(N_1(t)=j)=\frac{\left(m_1(t)\right)^j\exp\left(-m_1(t)\right)}{j!}=P_{1,j}(t), \tag{2}$$

where $m_1(t)=\int_0^t p\lambda(x)dx$ denote the mean number of type-I shock in [0,*t*].

If the total damage exceeds a failure level *K,* a serious failure occurs. The probability that a serious failure occurs at the *j-th* type-I shock is $H^{(j-1)}(K)-H^{(j)}(K)$. Let a random variable *Z* denote the occurrence time of the first serious failure, so the survival function of *Z* is given by:

$$\overline{F_z}(t)=P(Z>t)=P(Y_{N_1(t)}<K)=\sum_{j=0}^{\infty}P(N_1(t)=j,Y_j<K)=\sum_{j=0}^{\infty}P_{1,j}(t)H^{(j)}(K) \tag{3}$$

and the density function of Z is $f_z(t)=p\lambda(t)\sum_{j=0}^{\infty}\left(H^{(j)}(K)-H^{(j+1)}(K)\right)P_{1,j}(t)$.

Each type-II shock makes the system into minor failure. Hence, the probability of *j* minor failures in [0, *t*] is given by:

$$P(N_2(t)=j)=\frac{\left(m_2(t)\right)^j\exp(-m_2(t))}{j!}=P_{2,j}(t) \tag{4}$$

where $m_2(t)=\int_0^t q\lambda(x)dx$ denote the mean number of minor failures in [0,*t*].

Moreover, let S_{2j} (*j*=1, 2, 3, …) denote the occurrence time of the *j-th* minor failure, where $S_{20}=0$, then the distribution function of a random variable S_{2j} is given by:

$$P(S_{2j} \le t) = P(N_2(t) \ge j) = \sum_{i=j}^{\infty} P_{2,i}(t), \quad j = 1, 2, 3, ...,$$

and

$$f_{s_{2j}}(t) = \frac{d}{dt} P(S_{2j} \le t) = \frac{d}{dt} \sum_{i=j}^{\infty} \frac{(m_2(t))^i \exp(-m_2(t))}{i!} = q\lambda(t) P_{2,j-1}(t).$$

When a minor failure occurs, the repair cost due to this minor failure is evaluated. Suppose that a minimal repair cost X_i due to the *i-th* minor failure has a nonnegative independent and identical distribution function $G(x) = P(X_i \le x)$ and a finite mean μ_x, i=1,2,3,…. Then, the accumulated repair cost till to *j-th* minor failure $Y_j = \sum_{i=1}^{j} X_i$ has a distribution function:

$$P(Y_j \le y) = G^{(j)}(y) = \begin{cases} 1 & j = 0 \\ G_1 * G_2 * \cdots * G_j(y), & j = 1,2,3,\cdots \end{cases} \quad (5)$$

If the accumulated repair cost exceeds a predetermined limit *L*, then the system must be replaced at this minor failure. Let a random variable *U* denote the occurrence time when the accumulated repair cost exceeds a predetermined limit *L*, so the survival function of *U* is given by:

$$\overline{F_u}(t) = P(U > t) = P(Y_{N_2(t)} < L) = \sum_{j=0}^{\infty} P(N_2(t) = j, Y_j < L) = \sum_{j=0}^{\infty} P_{2,j}(t) G^{(j)}(L) \quad (6)$$

and the density function of *U* is $f_u(t) = q\lambda(t) \sum_{j=0}^{\infty} \left(G^{(j)}(L) - G^{(j+1)}(L) \right) P_{2,j}(t)$.

In this model, preventive maintenance policy is executed according to the following scheme. Preventive replacement is carried out at the *n-th* minor failure or at the occurrence time of one minor failure, in which the accumulated repair cost at this moment exceeds a predetermined limit *L*, and failure replacement is executed at the occurrence time of a serious failure. According to the above scheme, the replacement of the system can occur at three different situations and the probabilities of three situations will be introduced as follows.

First, if the accumulated repair cost till to (*n-1*)*-th* minor failure is less than *L* and the *n-th* minor failure precedes a serious failure, then preventive replacement is executed at the *n-th* minor failure. Therefore, the probability of situation 1 is given by:

$$\int_0^{\infty} P(Z > s_{2n},\ Y_{n-1} < L) dF(s_{2n}) = G^{(n-1)}(L) \int_0^{\infty} \overline{F_z}(t) P_{2,n-1}(t) q\lambda(t) dt. \quad (7)$$

Because the occurrences of minor and serious failures are mutually independent.

Second, if the *j-th* (j<n) minor failure occurs and the accumulated repair cost till to this failure exceeds *L*, and no serious failure has occurred, then the system will be replaced by a new one at time S_{2j}, j=1,2,3,…, n-1. Therefore, the probability of situation 2 is given by:

$$\begin{aligned} \sum_{j=1}^{n-1} P(Y_{j-1} < L < Y_j, W_1 > S_{2j}) &= \sum_{j=1}^{n-1} \left(G^{(j-1)}(L) - G^{(j)}(L) \right) \int_0^{\infty} P(W_1 > s_{2j}) dF(s_{2j}) \\ &= \sum_{j=1}^{n-1} \left(G^{(j-1)}(L) - G^{(j)}(L) \right) \int_0^{\infty} \overline{F_z}(t) P_{2,j-1}(t) q\lambda(t) dt \end{aligned} \quad (8)$$

Finally, if a serious failure occurs before time S_{2j}, j=1,2,…, n-1 and the accumulated repair cost till to this serious failure is less than *L*, then the system will be replaced at serious failure. Therefore, the probability of situation 3 is given by:

$$\sum_{j=0}^{n-1} P(N_2(z) = j, Y_j < L) = \sum_{j=0}^{n-1} G^{(j)}(L) \int_0^{\infty} P_{2,j}(t) dF_z(t) \quad (9)$$

More specifically, we also require the following assumptions:

(a1) The system is monitored continuously so that minor or serious failures can be detected instantaneously.

(a2) The times taken for minimal repair or replacement are very smaller than the mean time between failures. As a consequence, we can ignore those and treat those as being zero.

(a3) The steady state case is considered.

Finally, Replacement at *n-th* minor failure or at which the accumulated repair cost exceeds limit *L* costs C_0 and is called preventive replacement, while replacement at serious failure costs C_1 and is called as failure replacement in which $C_1 > C_0$. This problem is just to find an optimal n^* to minimize the long-term expected cost per unit time $\overline{C}(n, L)$ in the steady state case.

3. Long-term expected cost per unit time

It is well known that if a replacement is performed, a new replacement cycle will restart. Therefore, the continuous replacement cycles will constitute a renewal process. Let $E(V_1)$ and $E(R_1)$ denote the mean length of a replacement cycle and the expected total cost incurred during a replacement cycle, respectively. Using the renewal-reward theorem, we can observe that the long-term expected cost per unit time in the steady-state case is given by (Ross (1983)):

$$\overline{C}(n, L) = E(R_1) / E(V_1).$$

Under our defined preventive maintenance policy, the expected length of a replacement cycle $E(V_1)$ is given by:

$$\begin{aligned} E(V_1) &= G^{(n-1)}(L) \int_0^{\infty} t \times \overline{F_z}(t) P_{2,n-1}(t) q\lambda(t) dt + \sum_{j=1}^{n-1} \left(G^{(j-1)}(L) - G^{(j)}(L) \right) \int_0^{\infty} t \times \overline{F_z}(t) P_{2,j-1}(t) q\lambda(t) dt \\ &\quad + \sum_{j=0}^{n-1} G^{(j)}(L) \int_0^{\infty} t \times P_{2,j}(t) dF_z(t) \\ &= \sum_{j=0}^{n-1} G^{(j)}(L) \int_0^{\infty} \overline{F_z}(t) P_{2,j}(t) dt \end{aligned} \quad (10)$$

If the system is replaced preventively following the *n-th* minor failure, the total cost is $C_0 + \sum_{i=1}^{n-1} X_i$. When the system is replaced preventively following the *j-th* (j<n) minor failure because the accumulated repair cost exceeds *L*, the total cost will be $C_0 + \sum_{i=1}^{j-1} X_i,\ j = 1,2,...,n-1$. However, if the replacement is executed at serious failure, then the total cost is $C_1 + \sum_{i=1}^{N_2(Z)} X_i$. Therefore, the expected total cost $E(R_1)$ can be derived as follows:

$$\begin{aligned}E(R_1) &= G^{(n-1)}(L)\int_0^\infty E\left[C_0+\sum_{i=1}^{n-1}X_i\right]\times\overline{F_z}(t)P_{2,n-1}(t)q\lambda(t)dt\\ &+\sum_{j=1}^{n-1}\left(G^{(j-1)}(L)-G^{(j)}(L)\right)\int_0^\infty E\left[C_0+\sum_{i=1}^{j-1}X_i\right]\times\overline{F_z}(t)P_{2,j}(t)q\lambda(t)dt\\ &+\sum_{j=0}^{n-1}G^{(j)}(L)\int_0^\infty E\left[C_1+\sum_{i=1}^{N_2(t)}X_i\right]\times P_{2,j}(t)dF_z(t)\\ &=C_0+(C_1-C_0)\sum_{j=0}^{n-1}G^{(j)}(L)\int_0^\infty P_{2,j}(t)dF_z(t)+\mu_x\sum_{j=1}^{n-1}G^{(j)}(L)\int_0^\infty\overline{F_z}(t)P_{2,j-1}(t)q\lambda(t)dt\end{aligned}, \qquad (11)$$

where $\mu_x = E(X_i)$.

Combining (10) and (11), the long-term expected cost per unit time $\overline{C}(n,L)$ can be obtained as follows:

$$\overline{C}(n,L)=\frac{C_0+(C_1-C_0)\sum_{j=0}^{n-1}G^{(j)}(L)\int_0^\infty P_{2,j}(t)dF_z(t)+\mu_x\sum_{j=1}^{n-1}G^{(j)}(L)\int_0^\infty\overline{F_z}(t)P_{2,j-1}(t)q\lambda(t)dt}{\sum_{j=0}^{n-1}G^{(j)}(L)\int_0^\infty\overline{F_z}(t)P_{2,j}(t)dt} \qquad (12)$$

In the steady-state case, we want to find an optimal number n^* that minimises $\overline{C}(n,L)$ under the following assumptions:

(a1) $\lambda(t)$ is a continuous and increasing function of t with $\lambda(\infty)=\lim_{t\to\infty}\lambda(t)$, which may be infinite.

(a2) $G^{(n)}(y)$ is PF2 (a Polya frequency function of order 2).

From Lemma 3.7 in Barlow and Proschan (1975), it is known that $G^{(n)}(L)$ is decreasing in n for all $L > 0$. In addition, we can observe that $G^{(n)}(L)$ is PF2 if and only if $G^{(n)}(L)\big/G^{(n-1)}(L)$ is decreasing in n for all $L > 0$ (Gottlieb 1980, p. 749).

If an optimal n^* exists, then the inequalities $\overline{C}(n+1,L)\ge\overline{C}(n,L)$ and $\overline{C}(n,L)<\overline{C}(n-1,L)$ are both satisfied for some finite n. In the derivation of these inequalities, we can see that the inequalities $\overline{C}(n+1,L)\ge\overline{C}(n,L)$ and $\overline{C}(n,L)<\overline{C}(n-1,L)$ are equivalent to the inequalities $K(n)\ge C_0$ and $K(n-1)<C_0$, where:

$$K(n)=\begin{cases}\dfrac{(C_1-C_0)\left(G^{(n)}(L)\int_0^\infty P_{2,n}(t)dF_z(t)\right)+\mu_x\left(G^{(n)}(L)\int_0^\infty\overline{F_z}(t)P_{2,n-1}(t)q\lambda(t)dt\right)}{G^{(n)}(L)\int_0^\infty\overline{F_z}(t)P_{2,n}(t)dt}\times\left(\sum_{j=0}^{n-1}G^{(j)}(L)\int_0^\infty\overline{F_z}(t)P_{2,j}(t)dt\right) & \\ \quad -(C_1-C_0)\left(\sum_{j=0}^{n-1}G^{(j)}(L)\int_0^\infty P_{2,j}(t)dF_z(t)\right)-\mu_x\left(\sum_{j=1}^{n-1}G^{(j)}(L)\int_0^\infty\overline{F_z}(t)P_{2,j-1}(t)q\lambda(t)dt\right) & n=1,2,3,\cdots\\ 0 & n=0\end{cases} \qquad (13)$$

Therefore, if we can show that $K(n)$ is an increasing function of n and $\lim_{n\to\infty}K(n)>C_0$, then n^* *is* finite and unique. To show that $K(n)$ is an increasing function of n, the following lemma is required.

Lemma 1. Under assumptions (a1) and (a2), the following results are true:

(1) $A_n=\int_0^\infty P_{2,n}(t)dF_z(t)\Big/\int_0^\infty\overline{F_z}(t)P_{2,n}(t)dt$ is increasing in n, and $\lim_{n\to\infty}A_n=p\lambda(\infty)$.

(2) $B_n=\int_0^\infty\overline{F_z}(t)P_{2,n}(t)q\lambda(t)dt\Big/\int_0^\infty\overline{F_z}(t)P_{2,n+1}(t)dt$ is increasing in n, and $\lim_{n\to\infty}B_n=q\lambda(\infty)$.

A proof of Lemma 1 can be found in Chang *et al*. (2010). Moreover, $K(n)$ is an increasing function of n is equivalent to the condition $K(n+1)-K(n)>0$ for all n. Consequently,

$$K(n+1)-K(n)$$

$$=\left(\sum_{j=0}^{n} G^{(j)}(L)\int_0^\infty \overline{F_z}(t)P_{2,j}(t)dt\right)\times\left(\begin{array}{c}\dfrac{(C_1-C_0)\left(G^{(n+1)}(L)\int_0^\infty P_{2,n+1}(t)dF_z(t)\right)+\mu_x\left(G^{(n+1)}(L)\int_0^\infty \overline{F_z}(t)P_{2,n}(t)q\lambda(t)dt\right)}{G^{(n+1)}(L)\int_0^\infty \overline{F_z}(t)P_{2,n+1}(t)dt}\\ -\dfrac{(C_1-C_0)\left(G^{(n)}(L)\int_0^\infty P_{2,n}(t)dF_z(t)\right)+\mu_x\left(G^{(n)}(L)\int_0^\infty \overline{F_z}(t)P_{2,n-1}(t)q\lambda(t)dt\right)}{G^{(n)}(L)\int_0^\infty \overline{F_z}(t)P_{2,n}(t)dt}\end{array}\right)$$

Using *Lemma* 1, $K(n+1)-K(n)$ can be obtained as follows:

$$K(n+1)-K(n)=\sum_{j=0}^{n} G^{(j)}(L)\int_0^\infty \overline{F_z}(t)P_{2,j}(t)dt\times\left[(C_1-C_0)\times\left(A_{n+1}-A_n\right)+\mu_x\times\left(B_{n+1}-B_n\right)\right]. \quad (14)$$

Because

$\sum_{j=0}^{n} G^{(j)}(L)\int_0^\infty \overline{F_z}(t)P_{2,j}(t)dt>0$, $C_1>C_0$, $A_{n+1}>A_n$ and $B_{n+1}>B_n$, then $K(n+1)-K(n)>0$ for all *n*. Thus, we can see that $K(n)$ is increasing in *n*. In summary, the conditions for the existence and the uniqueness of an optimal value n^* are expressed in the following theorem:

Theorem 1. Under assumptions (a1) and (a2), if $\lim_{n\to\infty} K(n)>C_0$, then there exists a finite and unique n^* that minimises $\overline{C}(n,L)$ and satisfies:

$$K(n^*)\ge C_0 \text{ and } K(n^*-1)<C_0,\ n^*=1,2,3,\cdots. \quad (15)$$

Proof. The inequalities $\overline{C}(n+1,L)\ge\overline{C}(n,L)$ and $\overline{C}(n,L)<\overline{C}(n-1,L)$ imply (15). Under assumptions (a1) and (a2), we can observe that $K(n)$ is increasing in *n* from ***Lemma*** 1. Furthermore,

$$\lim_{n\to\infty} K(n)=(C_1-C_0)\left[\sum_{j=0}^{\infty} G^{(j)}(L)\int_0^\infty \overline{F_z}(t)P_{2,j}(t)p\lambda(\infty)dt-\sum_{j=0}^{\infty} G^{(j)}(L)\int_0^\infty P_{2,j}(t)dF_z(t)\right]$$
$$+\mu_x\left(\sum_{j=0}^{\infty} G^{(j)}(L)\int_0^\infty \overline{F_z}(t)P_{2,j}(t)q\lambda(\infty)dt-\sum_{j=1}^{\infty} G^{(j)}(L)\int_0^\infty \overline{F_z}(t)P_{2,j-1}(t)q\lambda(t)dt\right)$$

In equation (13), we know that $K(0)=0<C_0$. If $\lim_{n\to\infty} K(n)>C_0$, we can observe that there is a finite *n* such that equation (15) is satisfied. In addition, the optimal value n^* is unique according to the fact that $K(n)$ is increasing in *n*.

In our model, if $K=0$, then $\overline{F_z}(t)=\exp\left(-\int_0^t p\lambda(x)dx\right)$ and $\overline{C}(n,L)$ is the same as $C(n)$ in Chien, *et al.* (2010).

4. Numerical example

We consider that the intensity rate $\lambda(t)$ of arrival shocks is taking as

$$\lambda(t)=\lambda t^{\beta-1},\quad \lambda>0,\ \beta>1. \quad (16)$$

We assume that the shape parameter is set at β=2, and that $\lambda(t)=\lambda t$ is an increasing function of *t*. Let two replacement costs C_0 and C_1 be 1000 and 1500, respectively. The amount of damage from consecutive type-I shocks are independently and identically exponential random variables with finite mean μ_d=100. And, the failure level of the system is set at *K*=800. The costs for consecutive minimal repair are also independently and identically exponential random variables with finite mean μ_x=50. In addition, the cumulative repair cost limit *L* is fixed to be 500.

Because $\lambda(t)$ is strictly increasing to ∞ as $t\to\infty$, $K(n)$ is increasing in *n*. Thus, n^* is finite and unique. Using the software MAPLE, n^* and the minimum long-term expected cost per unit time $\overline{C}(n^*,L)$ are computed for various values of the parameters λ and *p* are listed in Tables 1 and 2.

From Tables 1 and 2, we have the following conclusions:

(1) As λ increases, the optimal n^* is unchanged, but the minimum $\overline{C}(n^*,L)$ increases. This situation is due to the denominator of the equation (12), i.e., the expected length of a replacement cycle decreases. A greater value of λ implies that the arriving shocks occur more frequently, so the replacement period must be shorter to prevent the occurrence of random failures.

(2) As *p* increases (i.e., (1-*p*) decreases), the optimal n^* decreases, but the minimum $\overline{C}(n^*,L)$ increases. A greater value of *p* implies that serious failures occur more easily, so the replacement period must also be shorter, i.e., the optimal n^* must be smaller, to prevent the occurrence of serious failures.

Table 1. Optimal n^* and $\overline{C}(n^*, L)$ at different λ and p, when L/μ_x =10 and K/μ_d =8

	p=0.3		p=0.4		p=0.5		p=0.6		p=0.7	
	n^*	$\overline{C}(n^*, L)$	n^*	$\overline{C}(n^*, L)$	n^*	$\overline{C}(n^*, L)$	n^*	$\overline{C}(n^*, L)$	n^*	$\overline{C}(n^*, L)$
λ=1.0	12	292.9909184	9	294.9119827	7	301.3978537	5	310.5921124	4	321.3561456
λ=1.5	12	358.8391246	9	361.1919384	7	369.1354753	5	380.3960969	4	393.5792913
λ=2.0	12	414.3517304	9	417.0685255	7	426.2409323	5	439.2435780	4	454.4662197
λ=2.5	12	463.2593182	9	466.2967872	7	476.5518497	5	491.0892494	4	508.1086802

Table 2. Optimal n^* and $\overline{C}(n^*, L)$ at different values of L/μ_x and K/μ_d when p=0.5, λ=2

	L/μ_x =6		L/μ_x =8		L/μ_x =10		L/μ_x =12	
	n^*	$\overline{C}(n^*, L)$	n^*	$\overline{C}(n^*, L)$	n^*	$\overline{C}(n^*, L)$	n^*	$\overline{C}(n^*, L)$
K/μ_d =6	6	476.5594134	6	473.0556584	6	472.1746264	6	471.9659801
K/μ_d =8	7	432.9678584	7	427.7229495	7	426.2409323	7	425.8473161
K/μ_d =10	8	406.0023083	8	398.6716537	8	396.3677583	8	395.6877385
K/μ_d =12	9	388.8043367	9	379.1945801	9	375.8645467	9	374.7818536

(3) When the ratio L/μ_x is larger, i.e. this model allows more minor failures before replacement, we can see that the optimal n^* is unchanged and the minimum $\overline{C}(n^*, L)$ decreases. For a fixed ratio K/μ_d, the decreasing magnitudes of the optimal $\overline{C}(n^*, L)$ are significantly smaller than the increment of L/μ_x.

(4) When the ratio K/μ_d is larger, i.e. this model allows more type I shocks to occur before serious failure, we can see that the optimal n^* increases but the minimum $\overline{C}(n^*, L)$ decreases. The variations in the optimal n^* and $\overline{C}(n^*, L)$ with regard to K/μ_d are significantly larger than that of L/μ_x. Therefore, the ratio K/μ_d is more important than the ratio L/μ_x when determining the optimal replacement period.

5. Conclusions

In this article, a repair number counting replacement policy based on a cumulative repair-cost limit under a standard cumulative damage model is introduced. The long-term expected cost per unit time $\overline{C}(n, L)$ in operating the system was developed which incorporating costs due to holding minimal repair and different forms of replacement state. The optimal number n^* of minimal repair, which minimizes the cost rate function under a fixed cumulative repair-cost limit L, was shown. The existence, uniqueness and structural properties were also proposed. This research verifies that under some specific conditions, the optimal number n^* of minimal repair is finite and unique under fixed L and K. This model provided a general framework for analyzing the maintenance policies for a system subject to cumulative damage models, so two previous models in the literature were the special cases of our model. We also demonstrated some numerical examples.

However, some assumptions are possible limitations in this research, such as the repair and replacement times were negligible, and the repairs are minimal and the repaired system is as bad as old. In some practical situations, it would seem to be more practical to consider the concept of imperfect repairs or multi-unit systems. Taking these realistic factors into consideration in the proposed policy is one direction for future research.

References

1. Barlow RE, Proschan F. (1975). Statistical Theory of Reliability and Life Testing Probability Models. New York: Holt, Rinehart and Winston.
2. Chang CC, Sheu SH, Chen YL. (2010). Optimal Number of Minimal Repairs before Replacement Based on a Cumulative Repair-Cost Limit Policy. Computers & Industrial Engineering, 59, 603–610.
3. Chang CC, Sheu SH, Chen YL. (2013). Optimal replacement model with age-dependent failure type based on a cumulative repair-cost limit policy. Applied Mathematical Modelling, 37, 308–317.
4. Chien YH, Sheu SH, Chang CC. (2009). Optimal Age-Replacement Time with Minimal Repair Based on Cumulative Repair Cost Limit and Random Lead Time. International Journal of Systems Science, 40, 703–715.
5. Chien YH, Chang CC, Sheu SH. (2010). Optimal Age-Replacement Model with Age-Dependent Type of Failure and Random Lead Time Based on a Cumulative Repair-Cost Limit Policy. Annals of Operations Research, 181, 723–744.

6. Feldman, R. M. (1976). Optimal replacement with Semi-Markov shock models. Journal of Applied Probability, 13, 108-117.
7. Gottlieb G. (1980). Failure Distributions of Shock Models. Journal of Applied Probability, 17, 745–752
8. Ito K. and Nakagawa, T. (2011). Comparison of Three Cumulative Damage Models. Quality Technology and Quantitative Management, 8(1), 57-66.
9. Kijima, M. and Nakagawa, T. (1991). A cumulative damage shock model with imperfect preventive maintenance. Naval Research Logistics; 38(2), 145-156.
10. Lai, M.T. (2007). A Periodical Replacement Model Based on Cumulative Repair Cost Limit. Applied Stochastic Models in Business and Industry, 26, 455–464.
11. Mizuno, N. (1986). Generalized mathematical programming for optimal replacement in a Semi-Markov shock model. Operations Research, 34, 790-795.
12. Nakagawa, T. (1976). On a replacement problem of a cumulative damage model. Operations Research, 27(4), 895-900.
13. Nakagawa, T. (1980). A summary of imperfect preventive maintenance policies with minimal repair. RAIRO Operations Research, 14, 249-255.
14. Nakagawa, T. (1984). A summary of discrete replacement policies. European Journal of Operations Research, 17, 382-392.
15. Nakagawa, T. (2007). Shock and Damage Models in Reliability Theory. London: Springer- Verlag.
16. Nakagawa, T. and Kijima, M. (1989). Replacement policies for a cumulative damage model. IEEE Transaction on Reliability, 38(3), 581-584.
17. Padgett, W.J. (1998). A multiplicative damage model for strength of fibrous composite materials. IEEE Transaction on Reliability, 47(1), 46-52.
18. Perry, D. (2000). Control limit policies in a replacement model with additive phase-type distributed damage and linear restoration. Operations Research Letters, 27, 127-134.
19. Qian, C., Nakamura, S. and Nakagawa, T. (1999). Cumulative damage model with two kinds of shocks and its application to the backup policy. Journal of the Operations Research Society of Japan, 42, 501-511.
20. Qian C.H., Ito K. and Nakagawa, T. (2005). Optimal preventive maintenance policies for a shock model with given damage level. Journal of Quality in Maintenance Engineering, 11(3), 216 - 227.
21. Qian, C., Nakamura, S. and Nakagawa, T. (2003). Replacement and minimal repair policies for a cumulative damage model with maintenance. Computers & Mathematics with Applications, 46(7), 1111-1118.
22. Ross, S. M. (1983). Stochastic Processes. New York: John Wiley & Sons.
23. Satow, T., Yasui, K. and Nakagawa, T. (1996). Optimal garbage collection policies for a database in a computer system, RAIRO-Operations Research, 30, 359-372.
24. Satow, T. and Nakagawa T. (1997). Three replacement models with two kinds of damage. Microelectronics Reliability, 37(6), 909-913.
25. Satow, T., Teramoto, K. and Nakagawa T. (2000). Optimal replacement policy for a cumulative damage model with time deterioration. Mathematical and Computer Modelling, 31(10-12), 313-319.
26. Sheu, S., Chang C.C., Chen Y.L., and Zhang, Z.G. (2010). A Periodic Replacement Model Based on Cumulative Repair-Cost Limit for a System Subjected to Shocks. IEEE Transaction on Reliability, 59, 374–382.
27. Taylor H.M. (1975). Optimal replacement under additive damage and other failure models. Naval Research Logistics, 22(1), 1-18.
28. Wang, H. (2002). A Survey of Maintenance Policies of Deteriorating Systems. European Journal of Operational Research, 139, 469–489.
29. Zhao, X., Zhang, H., Qian, C., Nakagawa, T., and Nakamura, S. (2012). Replacement models for combining additive independent damages. International Journal of Performability Engineering, 8, 91-100.
30. Zhao, X. and Nakagawa, T. (2012). Optimization problems of replacement first or last in reliability theory. European Journal of Operational Research, 223, 141-149.
31. Zhao, X., Qian, C. and Nakagawa, T. (2013). Optimal policies for cumulative damage model with maintenance last and first. Reliability Engineering and System Safety, 110, 50-59.

Min-Tsai LAI
Department of Business Administration
Southern Taiwan University of Science and Technology
Taiwan, R.O.C.
lai1962@mail.stust.edu.tw

Article citation info:
HRYNIEWICZ O, KARPIŃSKI J. Prediction of reliability – the pitfalls of using Pearson's correlation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 472–483.

Olgierd HRYNIEWICZ
Janusz KARPIŃSKI

PREDICTION OF RELIABILITY – THE PITFALLS OF USING PEARSON'S CORRELATION

PROGNOZOWANIE NIEZAWODNOŚCI – PUŁAPKI ZWIĄZANE Z UŻYWANIEM WSPÓŁCZYNNIKA KORELACJI PEARSONA

Pearson's coefficient of linear correlation r is the measure of dependence which is the most popular among practitioners. In the paper we have shown, using comprehensive computer simulations, that its application is very limited when we search for informative variables that can be used for the prediction of reliability. We have shown that Kendall's coefficient of association τ is much better for this purpose.

Keywords: *measures of dependence, Kendall's τ, Pearson's r, Spearman's ρ, prediction of reliability.*

Współczynnik korelacji liniowej r Pearsona jest najbardziej popularną wśród praktyków miarą zależności statystycznej. W artykule na podstawie wyników wyczerpujących symulacji komputerowych pokazano, że w przypadku poszukiwania zmiennych mogących służyć do prognozowania niezawodności zakres jego stosowalności jest bardzo ograniczony. Wyniki badań symulacyjnych pokazują, że temu celowi lepiej służy współczynnik asocjacji τ Kendalla.

Słowa kluczowe: *miary zależności, współczynnik τ Kendalla, współczynnik ρ korelacji r Pearsona, współczynnik korelacji ρ Spearmana, prognozowanie niezawodności.*

1. Introduction

Statistical regression models are widely used in the analysis of reliability data. In the recent overview paper by Elsayed [2] these methods have been indicated as the principal tools in areas such as reliability prediction and accelerated life tests. In a similar overview dedicated to the problem of warranty data analysis Wu [18] gives examples of the applications of regression methods in this area. As statistical data coming from long-lasting life tests are seldom available, many attempts have been made to build mathematical models for the prediction of reliability basing on easily observed (or measured) characteristics. For example, prediction models presented in the Military Handbook MIL-217F [15] link the most popular reliability characteristic, namely the hazard rate λ, with many factors describing the object itself, the condition of its usage, etc. These models are based on the statistical analysis of large sets of reliability data collected over years by organizations such as the U.S. Army. The mathematical models that are used for prediction purposes in MIL-217F and other similar documents are usually obtained using classical regression methods. Consider, for example, the prediction of the base hazard rate of a travelling wave tube. In the Notice 2 of the Military Handbook MIL-217F [16] the following formula is given for the calculation of the basic failure rate of such device $\lambda_b = 11 \cdot (1{,}00001)^P \cdot (1{,}1)^F$, where F is the operating frequency (in GHz), and P is the rated power (in Watts). When we take logarithms of both sides of this formula we arrive at a classical linear regression model that links the basic reliability characteristic with the parameters of the considered device. The parameters of the models presented in [15] and [16] are somewhat obsolete because they were computed using data collected more than twenty years ago. However, the general formulae used in MIL 217F for the prediction purposes are still used (see, e.g., the recent papers by Lee and Lee [9] or by Thaduri *et al.* [14]).

The second important area of the theory and practice of reliability in which regression models are widely used is accelerated life testing. The two most important classes of models used for the description of the accelerated life tests, namely the accelerated failure time models (AFT) and the proportional hazard models (PH), belong to the class of regression models (see [2] for a short overview). Regression models are also used in other areas of reliability and risk analysis. For example, Schneidewind [12] proposed a regression model for the prediction of risk in software engineering.

In order to build prediction models it is necessary to evaluate the strength of statistical dependence between the characteristic of interest and its best predictors. It is obvious that the values of good predictors should be strongly associated with the values of the characteristic of interest. In mathematical statistics many measures of statistical dependence exist, but Pearson's coefficient of correlation r is the most popular among practitioners. The reason of this stems from the fact that in nearly all popular software tools, such as spreadsheets or basic versions of statistical packages, Pearson's coefficient of correlation r is the main measure used for the evaluation of regression models.

Pearson's coefficient of correlation r (usually called simply "the correlation") measures the strength of *linear correlation* between random variables. In all statistical textbooks, readers are warned against the usage of this measure of dependence when the dependence between random variables is nonlinear. For example, in the case of two random variables X and $Y=X^2$ defined on the whole space of real numbers, their linear correlation coefficient will be equal to zero despite the strongest possible (deterministic) relation. In practice however, one cannot easily recognize to what extent random variables are linearly dependent, even if the type of their bivariate probability distribution is known. It is well known from the theory of mathematical statistics that such linear dependence exists when the random variables are jointly distributed according to the multivariate normal

(Gaussian) distribution. When the assumption about the multivariate normality is not fulfilled one needs to use other measures of statistical dependence, such as Kendall's coefficient of association τ or Spearman's coefficient of rank correlation ρ. Unfortunately, a general theory that explains the links between Pearson's coefficient of correlation r and nonparametric measures of dependence, such as Kendall's τ or Spearman's ρ does not exist. Therefore, the relationship between these measures of dependence is usually investigated in particular context. For example, Xu *et. al.* [19] consider the problem of the measurement of correlation in signal processing when measurements are described by contaminated normal models. A very interesting analysis is presented in the paper by Vořechovský [17] who considered the problem of the Monte Carlo simulation of interdependent random vectors.

Regression models can be built for practically all types of statistical data. However, their statistical properties as calculated by popular software or described in the majority of statistical textbooks are valid only for the data described by the normal distribution. When lifetime data are analyzed this assumption is fulfilled only in very few practical cases, as lifetimes are seldom distributed according to the normal distribution. The situation is even worse when we build a regression model for the prediction of the hazard rate λ. In this case, the probability distribution of the predicted variable is *never* distributed according to the normal distribution. Probability distributions encountered in reliability testing, such as the exponential, Weibull, gamma or log-normal distributions, are skewed, and the multivariate (bivariate in practice) normal (Gaussian) distribution *should not* be used for the modeling of statistical dependence between the characteristic of interest and its predictors. Therefore, there is a need to investigate the behavior of Pearson's correlation coefficient r when the underlying models of dependence are applicable in the context of reliability prediction. This is the main aim of this paper.

The paper has the following structure. In its second section we recall some basic information about the methods for measuring the dependence between random variables. The main aim of this section is to highlight important restrictions for the usage of the coefficient of linear correlation. The third section of the paper is devoted to the analysis of the relations between the values of the coefficient of linear correlation and the values of other popular measures of statistical dependence, such as Kendall's coefficient of associationτ or Spearman's coefficient of rank correlation ρ. Approximate formulae, based on the results of extensive Monte Carlo computer simulation experiments, which link the values of r with the values of other measures of dependence are presented in the fourth section.

2. Measuring of dependence between random variables

Let X and Y be random variables whose joint probability distribution is $H(x,y)$. In this paper we assume that these variables have continuous marginal distributions $F(x)$ and $G(y)$ with finite expected values E(X), E(Y), and variances V(X), V(Y), respectively. Many such distributions have been proposed over the last one hundred years. Sklar [13] published his famous theorem which says that any two-dimensional probability distribution function $H(x,y)$ with marginal distributions $F(x)$ and $G(y)$ is represented using a function C, called a *copula,* in the following way:

$$H(x,y) = C(F(x), G(y)) \tag{1}$$

for all $x,y \in R$.

Any function defined on a square unit $[0,1]\times[0,1]$ and such that:

$$C(0,x) = C(x,0) = 0,$$

$$C(1,x) = C(x,1) = 1,\ x \in [0,1], \text{ and}$$

$$C(b,d) - C(a,d) - C(b,c) + C(a,c) \geq 0,\ a,b,c,d \in [0,1],\ a \leq b,\ c \leq d$$

is a copula. Conversely, for any distribution functions F and G and any copula C, the function H defined by (1) is a two-dimensional distribution function with marginals F and G. Moreover, if F and G are continuous, then the copula C is unique.

Let $u=F(x)$, and $v=G(y)$. The simplest copula, the product copula $\Pi(u,v) = uv$, describes *independent* random variables. All other bivariate copulas fulfill the Fréchet-Hoeffding inequalities:

$$W(u,v) = \max(u+v-1,0) \leq C(u,v) \leq \min(u,v) = M(u,v) \tag{2}$$

The left inequality in (2) describes the case of full negative dependence between X and Y, and the right inequality in this formula describes the case of full positive dependence between X and Y.

Sklar's theorem has been generalized to the p-dimensional case, so it is applicable for any p-dimensional probability distribution. Similarly, the Fréchet-Hoeffding inequalities have been also generalized for the p-dimensional case. However, in this more general setting all mathematical formulae describing multidimensional probability distributions become very complicated, and thus have limited usage for practitioners. Therefore, in this paper we restrict ourselves only to the two-dimensional (bivariate) case.

The most popular measure of dependence between two random variables is based on the concept of the *covariance* defined for real valued random variables as:

$$Cov(X,Y) = \iint_{S_{xy}} (x - E(X))(y - E(Y)) f(x,y)\,dxdy \tag{3}$$

where S_{xy} is the area for which the bivariate probability density function $f(x,y)$ is positive. When we divide the covariance by the product of the standard deviations $\sigma(X)$, and $\sigma(Y)$ of X and Y we arrive at the famous Pearson's coefficient of linear correlation:

$$r(X,Y) = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} \tag{4}$$

described in every textbook on probability and statistics.

Let $(x_i, y_i),\ i = 1,\ldots,n$ be the observed sample of n independent pairs of observations of the random vector (X,Y). The sample version of Pearson's coefficient of linear correlation is given by the well known formula:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}\ . \tag{5}$$

It is a well known that $r(X,Y)$ describes only *linear* dependence between random variables, and thus should not be used for many bivariate probability distributions as the measure of dependence. For example, if X and Y are independent, than $r(X,Y)=0$, but the converse is not true. There exist many examples of highly dependent data for whom we observe no linear correlation ($r(X,Y)$ is equal or very close to zero). It has been proven that Pearson's coefficient of correlation fully describes the dependence structure only in the case of the bivariate normal (Gaussian) distribution. This distribution is the special case (for normal marginal distribution) of the normal copula defined as:

$$C_N(u,v) = \Phi_N\left(\Phi^{-1}(u), \Phi^{-1}(v); r\right) \tag{6}$$

where $\Phi_N(x,y;r)$ is the cumulative distribution function of the bivariate standardized normal distribution with the correlation coefficient r, and $\Phi^{-1}(x)$ is the inverse of the cdf of the univariate standardized normal distribution (the quantile function). Pearson's r may be also used as a measure of dependence for random variables that are jointly elliptically distributed. To this class of probability distributions belong the aforementioned multivariate Gaussian distribution, the multivariate t-distribution, and other distributions whose multivariate characteristic function can be represented as a certain quadratic form. However, even in the case of the elliptical distributions Pearson's r has meaning only for the distributions with finite variances.

Another popular measure of dependence is Spearman's coefficient of rank correlation. Let $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ and $y_{(1)} \le y_{(2)} \le \cdots \le y_{(n)}$ be the ordered elements of $(x_i, y_i), i = 1,\ldots,n$, and let $R_1 \le R_2 \le \cdots \le R_n$ and $S_1 \le S_2 \le \cdots \le S_n$ be the *ranks* of the original observations $x_1,\ldots x_n$ and $y_1,\ldots y_n$ in this ordering. Spearman's coefficient of rank correlation is the coefficient of linear correlation calculated for these ranks, and is given by the formula:

$$\rho_{xy} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2-1\right)}, \tag{7}$$

where:

$$d_i = R_i - S_i, i = 1,\ldots,n. \tag{8}$$

It has been proved, see Nelsen [10], that the population version of Spearman's ρ can be found for any copula using the following formula:

$$\rho(X,Y) = 12\iint_{[0,1]^2} C(u,v)dudv - 3. \tag{9}$$

Kendall's rank correlation coefficient, known as Kendall's τ, was proposed in 1938, and is based on the concept of concordant and disconcordant pairs of observations. A pair of vector observations (x_i, y_i), and (x_j, y_j) of continuous random variables (X,Y) is *concordant* if the respective ranks of the elements of both vectors agree, i.e either $R_i > R_j$ and $S_i > S_j$ or $R_i < R_j$ and $S_i < S_j$. Otherwise, this pair is *disconcordant*. The sample version of Kendall's τ is defined as:

$$\tau_{xy} = 2\frac{\text{no. of concordant pairs - no. of disconcordant pairs}}{n(n-1)}. \tag{10}$$

A convenient representation of τ has been proposed by Genest and Rivest [7] in the following form:

$$\tau_{xy} = \frac{4}{n}\sum_{i=1}^{n} V_i - 1, \tag{11}$$

where:

$$V_i = card\left\{\left(X_j, Y_j\right) : X_j < X_i, Y_j < Y_i\right\} \big/ (n-1), i = 1,\ldots,n. \tag{12}$$

The population version of Kendall's τ can be found, see Nelsen [10], for any copula using the following formula:

$$\tau(X,Y) = 4\iint_{[0,1]^2} C(u,v)dC(u,v) - 1. \tag{13}$$

Many other measures of dependence exist, described for instance in the book by Nelsen [10] or in the paper by Embrechts *et al*. [3]. Some of these measures are called the measures of *concordance*. Scarsini [11] defines a measure of concordance as a real valued measure of dependence κ between two continuous random variables X and Y whose copula C satisfies the following properties:

1. κ is defined for every pair $X; Y$ of continuous random variables.
2. $-1 \le \kappa_{X,Y} \le 1$, $\kappa_{X,X} = 1$ and $\kappa_{X,-X} = -1$.
3. $\kappa_{X,Y} = \kappa_{Y,X}$.
4. If X and Y are independent, then $\kappa_{X,Y} = 0$.
5. $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$.
6. If C and $\tilde{C}$ are copulas such that $C \le \tilde{C}$, then $\kappa_C \le \kappa_{\tilde{C}}$.
7. If $\{(X_n; Y_n)\}$ is a sequence of continuous random variables with copulas C_n, and if $\{C_n\}$ converges pointwise to C, then $\lim_{n\to\infty} \kappa_{C_n} = \kappa_C$.

Spearmans ρ and Kendall's τ are measures of concordance (the proof can be found in the book by Nelsen [10]), but Pearson's r is not (as it is shown in the paper by Embrechts et al. [3]). It does not fulfill the condition 2., and the range of possible values of r depends upon the type of marginal distributions of dependent random variables X and Y. Below, we show some important properties of Pearson's r regarding this property.

Let us consider two continuous random variables X and Y described by the probability density functions $f(x)$ and $g(y)$, respectively. Without loss of generalization let us assume that $E(X)= E(Y)=E$, and $Var(X)= Var(Y)=1$. Because Pearson's r is invariant with respect to linear transformations, transforming the original random variables to the variables defined above does not change the value of r which in this case is equal to the covariance between X and Y.

Now, let us consider the two limiting cases defined by (2). In the case of full *negative* dependence random variables X and Y are linked functionally in the following way:

$$F(x) = 1 - G(y). \tag{14}$$

where $F(x)$ and $G(x)$ are the respective cumulative probability functions of the random variables X and Y. Hence, the covariance between X and Y is given by:

$$Cov_{neg}(X,Y) = \int_{-\infty}^{\infty}(x-E)\left(\left\{G^{-1}\left[1-F(x)\right]\right\} - E\right)f(x)dx \tag{15}$$

where $G^{-}(x)$ is the inverse (the quantile function) of $G(x)$.

In the case of full *positive* dependence the link is of the form:

$$F(x) = G(y), \tag{16}$$

and a similar formula is given by:

$$Cov_{pos}(X,Y) = \int_{-\infty}^{\infty}(x-E)\left(\left\{G^{-1}\left[F(x)\right]\right\} - E\right)f(x)dx \tag{17}$$

The formulae (15) and (17) can be used for the calculation of the limiting values, r_{min} and r_{max}, of Pearson's r. From the analysis of these formulae we can derive the following properties of Pearson's r.

Property 1: When the probability distributions of X and Y have the same shape, then $r_{max}=1$.

Proof: The proof of this property if straightforward. The same shape of two probability distributions means that after appropriate transformations of scale and location we have $F(x)=G(y)$. Hence, $G^{-1}\left[F(x)\right]=x$ and $Cov_{pos}(X,Y)=Var(X)$, and thus $r(X,Y)=r_{\max}=1$.

Property 2: When probability distributions of X and Y are symmetric around zero (E=0) and have the same shape, then $r_{\min}=-1$.

Proof: For symmetric distributions, with E=0, we have $G^{-1}(-x)=-G^{-1}(x)$. Thus, for distributions with the same shape we $G^{-1}(1-F(x))=-G^{-1}(F(x))=-x$. Then, we have $Cov_{neg}(X,Y)=-Var(X)$, and thus $r(X,Y)=r_{\min}=-1$.

Property 3: When at least one of the random variables has a symmetric distribution, then $r_{\min}=-r_{\max}$.

Proof: Let Y be the random variable with a symmetric distribution, then $G^{-1}(1-F(x))=-G^{-1}(F(x))$, Hence, we have $Cov_{pos}(X,Y)=-Cov_{neg}(X,Y)$, and consequently $r_{\min}=-r_{\max}$.

With the exception of cases when Properties 1 and 2 hold, the calculation of $r_{\min}$ and $r_{\max}$ is usually difficult.

Example 1

Consider the case when both X and Y have the same exponential distribution with E=1. Because the variance in the exponential distribution is the same as the expected value we have in the considered case $r(X,Y)=Cov(X,Y)$. Then, the formula (15) takes the following form:

$$r_{\min}=Cov_{neg}(X,Y)=\int_0^{\infty}(x-1)\left(\left\{-\ln\left[1-e^{-x}\right]\right\}-1\right)e^{x}dx=1-\frac{\pi^2}{6}=-0,644934. \quad (18)$$

The integral in (18) has been evaluated using symbolic and numerical calculations provided by the mathematical package Mathematica™.

Example 2

Consider the case when X is distributed according to the exponential distribution with E=1, and Y is uniformly distributed over the interval [-0,5 , 0,5]. The maximal value of the $Cov(X,Y)$ is now:

$$Cov_{pos}(X,Y)=\int_0^{\infty}(x-1)\left(1-e^{-x}\right)e^{x}dx=\frac{1}{4}e^{-2x}\left[\left(4e^{-x}-2\right)x+1\right]\Big|_0^{\infty}=\frac{1}{4}. \quad (19)$$

Hence, $r_{\max}=\sqrt{3}/2=0,866$, and, by the Property 3, $r_{\min}=-0,866$.

When the random variables X and Y are distributed according to the reliability distributions such as the Weibull or the Log-normal, which are so popular in theory and in practice, the calculation of the minimal or maximal values of Pearson's r can be done only numerically or by simulations. However, the numerical integration can, in this case, be very difficult, as the integrated functions may adopt infinite values at zero. For this reason, the Monte Carlo simulations, described in the next section of this paper, seem to be a better way to find these values.

3. Properties of Pearson's r

It is well known that the values of Pearson's r depend upon the type of the marginal distributions of a bivariate random variable. In the previous section we have shown how the range of possible values of r depends upon the shape of these marginals. More questions, important from a practical point of view, could be asked. In this paper we will try to answer some of them, and namely:

a. How the values of r depend upon the type of marginal distributions in the case of distributions used in reliability practice?
b. Do the properties of r depend upon the type of dependence described by some popular copulas?
c. What is the relationship between the values of r and the values of other measures of dependence, such as Kendall's τ or Spearman's ρ?
d. What is the accuracy of the estimation of different measures of dependence?

For these, and many other similar questions, the answers cannot be found using analytical methods. Therefore, we have performed extensive computer simulations and analyzed samples of different size, generated from different copulas with different marginal distributions.

We have considered four types of copulas. The first one, the normal (Gaussian) copula have been already introduced, and defined by (6). The remaining three copulas belong to the family of the Archimedean copulas defined by Genest and McKay [6] in the following way:

$$C(u,v)=\varphi^{-1}\left(\varphi(u)+\varphi(v)\right), \quad (20)$$

where φ^{-1} is a pseudo-inverse of the continuous and strictly decreasing function $\varphi:[0,1]\to[0,\infty]$, called copula's generator, such that $\varphi(1)=0$. From this family we have taken the following three well known copulas:

a. Clayton copula (Clayton [1]), defined as:

$$C(x,y)=\left[F^{-\theta}(x)+G^{-\theta}-1\right]^{-1/\theta}, \theta\in\left\{(-1,\infty)\setminus\{0\}\right\}, \quad (21)$$

b. Frank copula (Frank [4]), defined as:

$$C(x,y)=-\frac{1}{\theta}\ln\left(1+\frac{\left(e^{-\theta F(x)}-1\right)\left(e^{-\theta G(y)}-1\right)}{e^{-\theta}-1}\right), \theta\in\left\{(-\infty,\infty)\setminus\{0\}\right\}, \quad (22)$$

c. Gumbel copula (Gumbel [8]), defined as:

$$C(x,y)=\exp\left(-\left[\left(-\ln F(x)\right)^{\theta}+\left(-\ln G(y)\right)^{\theta}\right]^{1/\theta}\right), \theta>0. \quad (23)$$

One of the reasons for using these particular copulas is the relative ease of the computer simulation of samples from these copulas for the given strength of dependence defined by Kendall's τ. For the normal copula, popular algorithms can be used for this purpose for the simulation of samples from a classical bivariate normal distribution, and for the remaining three copulas we used a general algorithm proposed by Genest and McKay [6] for the Archimedean copulas.

For the measure of dependence in the simulated samples we use Kendall's τ. For this measure of dependence there exist formulae that link the value of τ with the parameters of the copulas. These links depend only on the type of copula, and because of a non-parametric character of Kendall's τ do not depend upon the type of marginals. For the normal (Gaussian) copula the following relation holds:

$$\tau = \arcsin(r) / (\pi / 2). \tag{24}$$

For the chosen Archimedean copulas we have the following formulae:

a. Clayton copula

$$\tau = \frac{\theta}{\theta + 2}, \tag{25}$$

b. Frank copula

$$\tau = 1 + 4\left(\frac{1}{\theta}\int_0^\theta \frac{t}{e^t - 1} dt - 1\right)\Big/\theta, \tag{26}$$

c. Gumbel copula

$$\tau = \frac{\theta}{\theta + 1}. \tag{27}$$

We see that except for the case of Frank copula, when we have to solve for θ a very complicated equation, the dependence parameter of a given copula is straightforwardly related to the value of Kendall's τ. Such simple relationships do not exist for Spearman's ρ, so we have chosen Kendall's τ as the measure of dependence in the simulated samples.

In order to investigate the influence of the type of the marginal distribution on the value of Pearson's r we considered two cases. In the first one we assumed that both variables X and Y have the same marginal distribution: normal, exponential and Weibull (with different parameters of shape δ). In the second case, that seems to be more appropriate as regards problems of reliability prediction, we have assumed that the predictor X has the normal distribution, and Y is distributed according to different Weibull distributions (the exponential distribution included).

The properties of the considered statistics depend on the sample size n. In our simulation experiments we considered three values of n: n=500, which allow the approximation of the values of the population (theoretical) versions of the measures of dependence, n=100, which represents the case of a relatively accurate estimation of this measure, and n=20, which represents the sample size more appropriate for the analysis of reliability.

We have simulated 1000000 samples in each of the simulation experiments. Therefore, the results of the experiment are very accurate, and the impact of the randomness of the Monte Carlo methodology can be neglected.

The results of experiments have been summarized in respective tables. In this paper we present only few of them, showing the results only for some chosen values of Kendall's τ. Table 1 represents the results of one of the simulation experiments where the Clayton copula with given marginals, normal N(0,1) for X, and Weibull W(1,5) for Y, was used as the mathematical model. In this experiment samples of n=100 elements were generated for 22 different values of τ, and for each value of τ the respective value or r was estimated from the results of simulation. The consecutive columns of this table represent: the assumed value of Kendall's τ, the estimated mean value of Kendall's τ, the estimated mean value of Spearman's ρ, the estimated mean value of Pearson's r, the estimated standard deviation of Kendall's τ, the estimated standard deviation of Spearman's ρ, and the estimated standard deviation of Pearson's r, respectively.

In Table 2 we present the results of the simulation experiment when dependence is described by the normal (Gaussian) copula. Note, that in this case the random vector (X,Y) does not have a bivariate normal distribution, as its second component (Y) is distributed according to the Weibull distribution with the shape parameter δ=1,5.

Table 1. X – N(0,1), Y – Weibull (1,5), Clayton copula, n=100

TAU	TAU-EST	RHO-SP	R-PEARS	SIG-TAU	SIG-RHO	SIG-R
1	1	1	0,966172	0	0	0,0084
0,9	0,899987	0,981913	0,919786	0,015427	0,005713	0,022431
0,7	0,699994	0,868551	0,780478	0,038352	0,032555	0,043463
0,5	0,500009	0,676824	0,602098	0,054615	0,063348	0,065445
0,3	0,300006	0,430006	0,386769	0,064561	0,087601	0,085434
0,1	0,099982	0,147826	0,135595	0,0683	0,099877	0,098458
0	0,000009	0,000018	0,000000	0,067926	0,100639	0,100626
-0,1	-0,10004	-0,14776	-0,13675	0,066267	0,097949	0,099825
-0,3	-0,30002	-0,42179	-0,38989	0,061608	0,086718	0,094286
-0,5	-0,5	-0,64362	-0,59994	0,057632	0,072373	0,084198
-0,7	-0,69998	-0,81401	-0,76949	0,051404	0,054946	0,0679
-0,9	-0,89995	-0,94535	-0,907	0,03438	0,030933	0,040685
-1	-1	-1	-0,966157	0	0	0,008387

Table 2. X – N(0,1), Y – Weibull (1,5), Normal copula, n=100

TAU	TAU-EST	RHO-SP	R-PEARS	SIG-TAU	SIG-RHO	SIG-R
1	1	1	0,966172	0	0	0,0084
0,9	0,899992	0,983882	0,954148	0,012531	0,004014	0,008852
0,7	0,699955	0,876410	0,859881	0,032894	0,026777	0,022927
0,5	0,499934	0,684427	0,681280	0,049546	0,057634	0,051623
0,3	0,299910	0,433129	0,436664	0,061163	0,083980	0,080477
0,1	0,099936	0,147973	0,150245	0,067111	0,098637	0,098170
0	0,000009	0,000018	0,000000	0,067926	0,100639	0,100626
-0,1	-0,100051	-0,148130	-0,150423	0,067128	0,098671	0,098221
-0,3	-0,300033	-0,433290	-0,436767	0,061215	0,084050	0,080566
-0,5	-0,500011	-0,684503	-0,681312	0,049594	0,057687	0,051656
-0,7	-0,700006	-0,876442	-0,859885	0,032900	0,026783	0,022923
-0,9	-0,899994	-0,983882	-0,954148	0,012528	0,004013	0,008851
-1	-1	-1	-0,966157	0	0	0,008387

Table 3. X – N(0,1), Y – Weibull (2,0), Clayton copula, n=100

TAU	TAU-EST	RHO-SP	R-PEARS	SIG-TAU	SIG-RHO	SIG-R
1	1	1	0,986855	0	0	0,004596
0,9	0,899987	0,981913	0,947121	0,015431	0,005713	0,015718
0,7	0,699994	0,868551	0,816918	0,038351	0,032555	0,038158
0,5	0,500009	0,676824	0,639403	0,054615	0,063348	0,062774
0,3	0,300006	0,430006	0,415361	0,064561	0,087601	0,085563
0,1	0,099982	0,147826	0,146231	0,0683	0,099877	0,099399
0	0,000009	0,000018	-0,000009	0,067926	0,100639	0,100629
-0,1	-0,10004	-0,14776	-0,14625	0,066267	0,097949	0,098495
-0,3	-0,30002	-0,42179	-0,4108	0,061608	0,086718	0,090795
-0,5	-0,5	-0,64362	-0,62437	0,057632	0,072373	0,079514
-0,7	-0,69998	-0,81401	-0,79356	0,051405	0,054946	0,063075
-0,9	-0,89995	-0,94535	-0,929	0,034378	0,030933	0,037035
-1	-1	-1	-0,98685	0	0	0,004588

Table 4. X – N(0,1), Y – Weibull (0,5), Frank copula, n=100

TAU	TAU-EST	RHO-SP	R-PEARS	SIG-TAU	SIG-RHO	SIG-R
1	1	1	0,719517	0	0	0,050144
0,9	0,899983	0,984952	0,674247	0,010692	0,003174	0,079032
0,7	0,699959	0,881641	0,565369	0,029395	0,023669	0,084575
0,5	0,499958	0,688860	0,434077	0,047261	0,055943	0,086974
0,3	0,299965	0,434539	0,274331	0,060343	0,083666	0,093489
0,1	0,099955	0,148060	0,093864	0,067048	0,098664	0,099509
0	0,000009	0,000018	0,000017	0,067926	0,100639	0,100577
-0,1	-0,100040	-0,148174	-0,093847	0,067086	0,098728	0,099595
-0,3	-0,300027	-0,434610	-0,274362	0,060460	0,083831	0,093634
-0,5	-0,499991	-0,688871	-0,434205	0,047386	0,056102	0,087096
-0,7	-0,699970	-0,881636	-0,565504	0,029461	0,023723	0,084793
-0,9	-0,899982	-0,984951	-0,674239	0,010700	0,003183	0,079138
-1	-1	-1	-0,71946	0	0	0,050233

Table 5. X – N(0,1), Y – Exponential, Gumbel copula, n=100

TAU	TAU-EST	RHO-SP	R-PEARS	SIG-TAU	SIG-RHO	SIG-R
1	1	1	0,909318	0	0	0,017728
0,9	0,899655	0,98295	0,901015	0,013871	0,004802	0,017174
0,7	0,69967	0,870859	0,829142	0,036312	0,030757	0,026007
0,5	0,499642	0,676797	0,682615	0,053541	0,062524	0,054827
0,3	0,299587	0,428082	0,461994	0,064564	0,087684	0,090058
0,1	0,09965	0,14718	0,170754	0,068657	0,100141	0,109652
0	0,000009	0,000018	0,000011	0,067926	0,100639	0,100616

From the first two columns of Table 1 and Table 2 one can have the impression that the estimates of Kendall's τ obtained from the generated samples are unbiased. Their average values (over 1000000 simulation runs) are practically the same as their assumed values. This has been confirmed in all simulation experiments, also for small samples of n=20 elements. This could serve as the proof that the algorithms used for the generation of data from different copulas work correctly.

The comparison of the third and the fourth columns of Table 1 and Table 2 shows different relation between Kendall's τ , Spearman's ρ, and Pearson's r. This reflects the influence of the type of copula. In the case of the Clayton copula the values of ρ and r are not symmetric with respect to the case of independence, where all dependence measures should have the value of zero. However, for the normal copula this symmetry is visible. The same situation is observed, but with lower intensity, for the observed standard deviations of the considered measures of dependence.

In Table 3 we present the results of simulations from the Clayton copula, but with a different, as compared to the case presented in Table 1, marginal distribution of distribution of Y, namely the Weibull distribution with the shape parameter δ=2,0. The seed of the generator of random numbers was the same in all performed simulations, so it is possible to compare their results directly. The comparison of the second and the third columns of Table 1 and Table 3 shows that the average values of Kendall's τ and Spearman's ρ are, because of non-parametric character of these statistics, exactly the same. However, the values of Pearson's r are slightly different in the both cases. This confirms the well-known fact that the values of r depend upon the type of the marginal distributions. What seems to be important from a practical point of view is the observation that in the cases in which the marginal distributions are not very different with respect to their skewness the values of Pearson's r are not very different.

The properties of Pearson's r are completely different in the case presented in Table 4 where the data were generated from the Frank copula, and the marginal distribution of Y was highly skewed (the Weibull distribution with the shape parameter δ=0,5). The behavior of Spearman's ρ in comparison to the cases presented in Tables 1 – 3 was similar, and the differences observed could be neglected from a practical point of view. However, the behavior of Pearson's r is completely different. First of all, the absolute minimal and maximal values of r are much smaller than 1, as is the case in the bivariate normal distribution. Therefore, they may be completely misleading when this measure of dependence will be used for the analysis of strongly dependent data.

The most important difficulty with the usage of Pearson's r is its dependence upon the type of marginal distributions. One can ask, however, about the practical impact of the distributions used to the problem of reliability prediction on the range of possible values of Pearson's r. In order to investigate this problem we have assumed that the random variable X is distributed according to the normal distribution N(0,1), and the random variable Y, which in the context of reliability prediction describes the life-time, is distributed according to different Weibull distributions, the exponential distribution included. We have searched for the minimal and maximal possible values of r, defined by (15) and (17), respectively. These two values have been evaluated in the Monte Carlo experiments in which samples of 100, 500, and 1000 items each were simulated in 1000000 runs. The results of this experiment are presented in Table 6.

Table 6. Minimal and maximal values of Pearson's r. One distribution symmetric

Distrib. X	Distrib. Y	n=100		n=500		n=1000	
		r_{min}	r_{max}	r_{min}	r_{max}	r_{min}	r_{max}
N(0,1)	Weib(0,2)	-0,4407	0,4407	-0,3297	0,3298	-0,2949	0,2950
N(0,1)	Weib(0,5)	-0,7195	0,7195	-0,6864	0,6864	-0,6796	0,6796
N(0,1)	Exp	-0,9093	0,9093	-0,9045	0,9045	-0,9039	0,9039
N(0,1)	Weib(1,5)	-0,9662	0,9662	-0,9647	0,9647	-0,9646	0,9646
N(0,1)	Weib(2,0)	-0,9869	0,9869	-0,9863	0,9863	-0,9862	0,9862
N(0,1)	Weib(2,5)	0,9951	0,9951	-0,9949	0,9949	-0,9949	0,9949

In this experiment one of the variables has a symmetric distribution, so according to the Property 3 the absolute values of r_{min} and r_{max} are the same. This has been confirmed in our experiments. Moreover, it appears from Table 6 that the range of possible values of r differs from the range expected for good measures of dependence, namely $[-1,1]$ only in cases of highly skewed distributions, such as the Weibull with the parameter of shape equal or smaller than 1 or the exponential distribution. However, in the case of distributions with the increasing hazard rate the range of the possible values of r is close to $[-1,1]$. This is not unexpected as with the increasing value of the parameter of shape the Weibull distribution tends to the normal distribution for whom Pearson's r is the proper measure of dependence.

The situation becomes different when both dependent random variables have skewed distributions. In Table 7 we present the results of simulation for several such distributions.

The results presented in Table 6 and Table 7 show beyond doubt that the evaluation of the strength of dependence using Pearson's r in the case of skewed distributions may be highly misleading. In extreme cases the absolute values of r may be very small even in the case of very strong dependence. Therefore, in such cases, Pearson's r cannot be used for finding characteristics that can be used as good predictors of life-times. It is extremely important when observed life-times come from highly accelerated life tests (HALT). In these tests early failures of "weak" elements are frequently observed with the consequence of observing highly skewed life-time distributions.

The results of the simulation experiments have shown another unwanted property of Pearson's r. The estimator of r seems to be highly biased even for large sample sizes. In the Table 6 and Table 7 we see this phenomenon for the extreme values of Pearson's r. However, in practice we are more interested in the analysis of this bias for smaller grades of dependence. In Table 8 we present the comparison of the estimated values of r for different copulas, different marginal distributions, and different sample sizes.

Table 7. Minimal and maximal values of Pearson's r. Both distributions asymmetric

Distrib. X	Distrib. Y	n=100		n=500		n=1000	
		r_{min}	r_{max}	r_{min}	r_{max}	r_{min}	r_{max}
Weib(0,2)	Weib(0,2)	-0,043138	1	-0,018837	1	-0,014069	1
Weib(0,2)	Weib(0,5)	-0,104550	0,876690	-0,063580	0,804192	-0,054084	0,771048
Weib(0,5)	Exp	-0,430280	0,924765	-0,393422	0,905902	-0,386876	0,901267
Exp	Weib(1,5)	-0,773163	0,983631	-0,762072	0,982289	-0,760580	0,982088
Exp	Weib(2,0)	-0,830921	0,960043	-0,821939	0,957323	-0,820714	0,956929
Weib(1,5)	Weib(2,5)	-0,938714	0,985191	-0,935726	0,984522	-0,935335	0,984433

Table 8. Expected values of the estimator of r

Copula	X	Y	τ	n=20	n=100	n=500	n=1000
Clayton	N(0,1)	Weib(0,5)	0,8	0,652802	0,55909	0,52084	0,513599
			0,5	0,414688	0,35028	0,32491	0,32020
			-0,5	-0,44616	-0,39468	-0,37105	-0,36645
			-0,8	-0,66799	-0,60190	-0,57096	-0,56473
Frank	N(0,1)	Weib(0,5)	0,8	0,71270	0,62183	0,58206	0,574378
			0,5	0,50179	0,43408	0,40502	0,399476
			-0,5	-0,50208	-0,43421	-0,40497	-0,39943
			-0,8	-0,71270	-0,62187	-0,58203	-0,57423
Gauss	Exp	Exp	0,8	0,93886	0,94148	0,94205	0,942104
			0,5	0,66083	0,66647	0,66775	0,667828
			-0,5	-0,53688	-0,50238	-0,49300	-0,49175
			-0,8	-0,69748	-0,63947	-0,62491	-0,62299

4. Approximate relationships between the values of different measures of dependence

We will estimate the unknown relationship between Pearson's r and Kendall's τ from the simulation data using a polynomial:

$$r_a(\tau) = \sum_{i=0}^{k} w_i \tau^i , \tag{28}$$

with additional conditions $r_a(-1) = L$, $r_a(0) = 0$, and $r_a(1) = U$. When we take k=6 after some simple algebra we obtain the following regression equation:

$$r_a(\tau) = \sum_{i=0}^{4} a_i f_i(\tau) , \tag{29}$$

where a0=1, and:

$$f_0(\tau) = \tau^5 \left[2U\tau + (U-L)(1-\tau) \right] / 2 , \tag{30}$$

$$f_1(\tau) = \tau\left(1-\tau^4\right), \tag{31}$$

$$f_2(\tau) = \tau^2\left(1-\tau^4\right), \tag{32}$$

$$f_3(\tau) = \tau^3 \left(1 - \tau^2\right), \tag{33}$$

$$f_4(\tau) = \tau^4 \left(1 - \tau^2\right). \tag{34}$$

Coefficients a_1, a_2, a_3, and a_4 of (29) have been obtained for different copulas, and different marginal distributions using a standard linear regression methodology for simulated samples of n elements. They are presented in Tables 10 – 12 for n=100, and the case of the normal N(0,1) distribution for one random variable, and the Weibull(δ) distribution, where δ is the shape parameter, for the second one.

The approximate relationship between Pearson's r and Kendall's τ enables us to analyze the impact of the type of a marginal distribution on r. Figure 1 presents functions $r(\tau)$ for the Clayton copula when the marginal distribution of the first random variable X is normal N(0,1) and the marginal of the second variable Y are those represented in Table 10.

Table 10. Coefficients of the polynomial approximation. Clayton copula, n=100

Coefficient	Weibull(0,5)	Exponential	Weibull(1,5)	Weibull(2,0)	Weibull(2,5)
a_1	0,7630	1,155342	1,33729	1,420882	1,473745
a_2	-0,1367	-0,0766	-0,02122	0,029073	0,059246
a_3	-0,08325	-0,34399	-0,55061	-0,64906	-0,7292
a_4	0,206633	0,182692	0,105321	0,007765	-0,05032

Table 11. Coefficients of the polynomial approximation. Frank copula, n=100

Coefficient	Weibull(0,5)	Exponential	Weibull(1,5)	Weibull(2,0)	Weibull(2,5)
a_1	0,939852	1,285098	1,387229	1,426354	1,499399
a_2	-0,00049	-0,00041	0,000288	-0,00042	0,130106
a_3	-0,30958	-0,42447	-0,45036	-0,47068	-0,63732
a_4	0,001063	0,000884	-0,001	0,000948	-0,34511

Table 12. Coefficients of the polynomial approximation. Gauss (normal) copula, n=100

Coefficient	Weibull(0,5)	Exponential	Weibull(1,5)	Weibull(2,0)	Weibull(2,5)
a_1	1,123037	1,420301	1,509639	1,542201	1,557253
a_2	0,000281	-0,00013	-0,00031	-0,00035	0,002749
a_3	-0,44679	-0,56673	-0,60338	-0,61684	-0,63445
a_4	-0,00064	0,00031	0,000842	0,00091	-0,01628

Table 13. Coefficients of the polynomial approximation. Gumbel copula, n=100

Coefficient	Weibull(0,5)	Exponential	Weibull(1,5)	Weibull(2,0)	Weibull(2,5)
a_1	1,676811443	1,787641617	1,763614843	1,727108377	1,695654409
a_2	-1,153926307	-0,817938091	-0,642773387	-0,540317764	-0,477612328
a_3	0,378567463	0,054332151	-0,082513834	-0,133359319	-0,151628861
a_4	-0,703818956	-0,360658976	-0,153403961	-0,084205969	-0,063509321
a_5	0,847882638	0,347773779	0,107779315	0,016341698	-0,014938649

For the most skewed distribution (Weibull with the shape parameter equal to 0,5) the relationship is nearly linear. When the marginal become more symmetric this relationship becomes more non-linear, concave for the positive dependence, and convex for the negative one.

Figure 2 presents the same relationship in the case of the Frank copula. The general properties of this relationship are the same as in the case of the Clayton copula. However, in the case of the most skewed marginal the function $r(\tau)$ is slightly more non-linear than in the case of the Clayton copula.

Figure 3 presents the same relationship in the case of the Gauss (normal) copula. In the case of the Gauss (normal) copula the general properties of the function $r(\tau)$ are the same as in the case of previous copulas. However, this function is more non-linear than in the case of other copulas considered.

In the case of the Gumbel copula we have only two restrictions $r_a(1) = U$, $r_a(0) = 0$. Hence, the regression formula takes the following form:

$$r_a(\tau) = U\tau^6 + \sum_{i=1}^{5} a_i \tau^i \left(1 - \tau^{6-i}\right). \tag{35}$$

The coefficients a_1, a_2, a_3, a_4, and a_5, estimated from the simulation data are presented in Table 13.

Function $r(\tau)$ for the case of the Gumbel copula is presented on Figure 4. For all the copulas considered in this case, this function is clearly the most non-linear (concave), even in the case of the most skewed Weibull distribution. The relationship between Pearson's r and Kendall's τ is approximately linear only in the case of weak dependency between considered random variables.

The approximations given by (29) and (35) are very accurate, as their accuracy measured using the R^2 statistic is close to 1. However, the usage of a simple regression technique does not guarantee in every case that the function $r_a(\tau)$ is monotonously increasing, as it should be. Therefore, it is possible to find a better approximation solving the required optimization problem with linear constraints imposed on the values of derivatives. This can be done using specialized optimization software.

The impact of the type of copula on the relationship between Pearson's r and Kendall's τ is presented in Figures 5 and 6. In Figure 5 we present this relationship when one of the two dependent variables is symmetric, N(0,1), and the second one is highly asymmetric, such as the Weibull(0,5). In Figure 6 we present the similar relationship when the second variable is characterized by only weak asymmetry, such as the Weibull(2,5). In both Figures we present the results of 1000000 simulations of the samples of 100 elements.

From these Figures one can see that in the case of highly asymmetric distributions the type of copula plays an important role. For the Clayton copula the function $r(\tau)$ is nearly linear. For the Frank copula it is not so far from being linear. However, for the Gauss (normal) copula, and especially for the Gumbel copula $r(\tau)$ is visibly non-linear. However, in the case of weakly asymmetric distributions this role is visible to a certain rather low degree only in the case of strong negative dependency. For all considered copulas the function $r(\tau)$ is non-linear, but this non-linearity is not very strong.

One of the most important characteristics of any statistical measure is its variability, measured by its variance or standard deviation. When the value of a statistical measure is bounded, the comparison of variability of different measures is not so straightforward, as for the same data, i.e. while the data dependent in the same way, the values of the measures of dependence may be quite different. Because of the

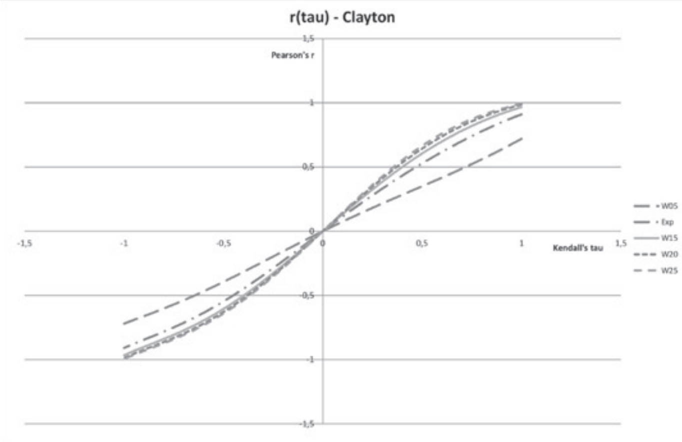


Fig. 1. Approximate relationship r(τ) – Clayton copula

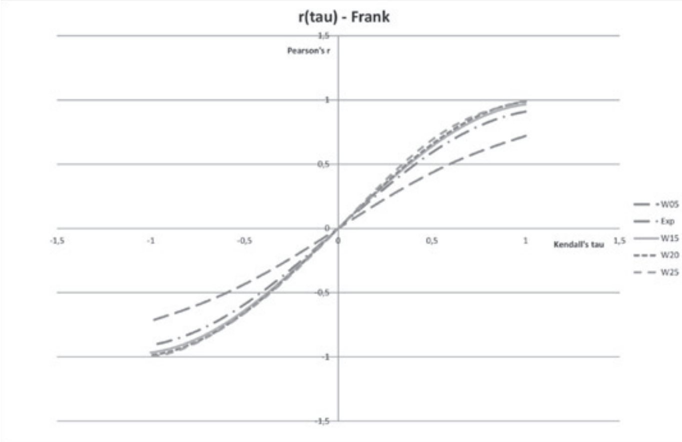


Fig. 2. Approximate relationship r(τ) – Frank copula

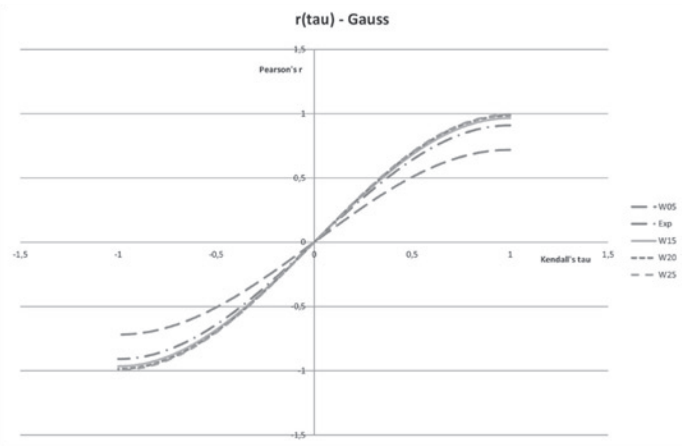


Fig. 3. Approximate relationship r(τ) – Gauss (normal) copula

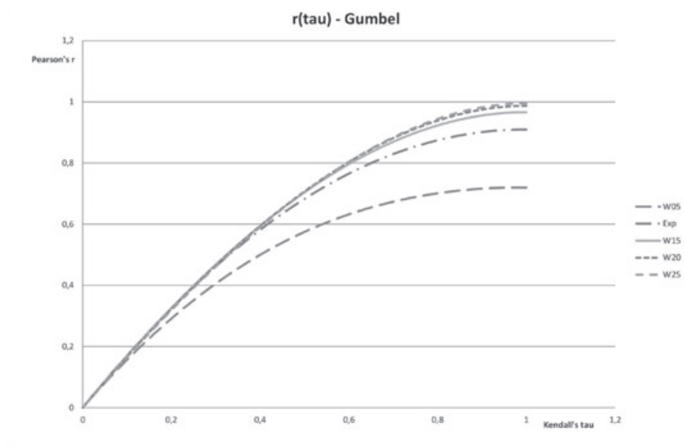


Fig. 4. Approximate relationship r(τ) – Gumbel copula

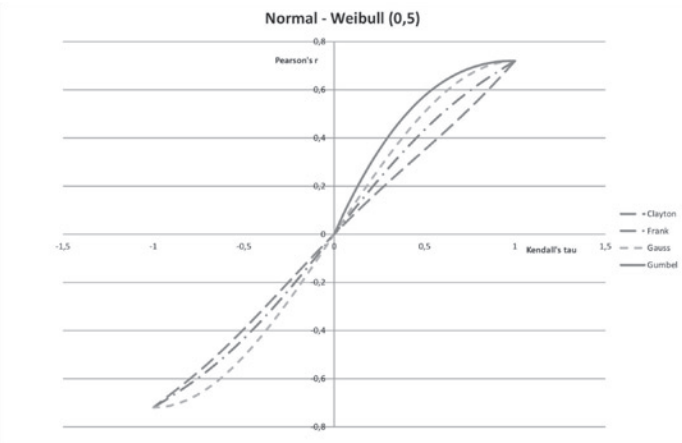


Fig. 5. Approximate relationship r(τ) – X- normal, Y – Weibull(0,5).

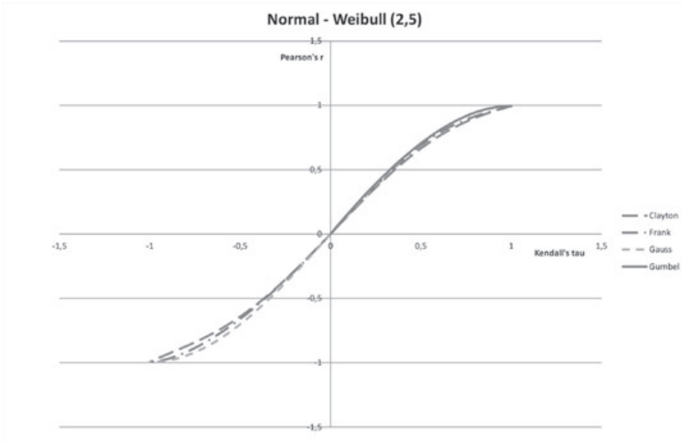


Fig. 6. Approximate relationship r(τ) – X- normal, Y – Weibull(2,5).

Table 14. Comparison of the average values of standard deviations of r and τ (n=100)

Copula	N(0,1) + Weibull (0,5)		N(0,1) + Weibull (2,5)	
	σ_τ	σ_r	σ_τ	σ_r
Clayton	0,022642	0,066985	0,047049	0,061169
Frank	0,040049	0,085868	0,040058	0,051757
Gauss (normal)	0,041423	0,071539	0,041423	0,049027
Gumbel	0,048311	0,07789	0,048311	0,064361

bounds on these values the variance of the measures whose values are closer to the bounds should be smaller. When we analyze the relationship between Pearson's *r* and Kendall's *τ* we can see that for highly skewed marginal distributions the absolute values of *τ* are greater than the values of *r* estimated from the same sample. Therefore, the observed variability of *τ* should be smaller than the observed variability of *r*. However, in the case of the more symmetric marginal distributions the values of *r* should be greater than the values of *τ*. Therefore, in

Table 15. Comparison of observed values of standard deviations of r and τ (same values of r and τ)

r=τ	σ_τ	σ_r
0,924983	0,034367	0,028979
0,91753	0,036602	0,029645
0,887326	0,045321	0,038427
0,836815	0,059091	0,061777
0,766665	0,076846	0,095548
0,677651	0,10082	0,134077
0,57147	0,118242	0,172431
0,44923	0,13792	0,206378
0,312261	0,15365	0,231068
0,162145	0,162801	0,241422
0,082468	0,163901	0,238912

Table 16. Coefficients of the polynomial approximation of ρ(τ) for samples of n=100

Coefficient	Clayton	Frank	Gauss
a_1	1,466021	1,496795	1,4891
a_2	0,069209	-0,00018	-0,00051
a_3	-0,60101	-0,48078	-0,48166
a_4	0,001862	0,000489	0,001283

Table 17. Coefficients of the polynomial approximation of ρ(τ) for samples of n=100 (Gumbel copula)

Coefficient	Gumbel
a_1	1,48496
a_2	-0,09709
a_3	-0,33526
a_4	0,14299
a_5	-0,34429

the case of similar variability of both measures of dependence the observed variability of r should be smaller than the observed values of τ. In order to verify this supposition we calculated the average values of standard deviations of the estimated values of τ and r, respectively. In Table 14 we present this comparison for two cases. In the first, the marginal distribution of X is normal, and the marginal distribution of Y is the Weibull distribution with the shape parameter equal to 0,5. This is the case of a highly skewed marginal. In the second case, the marginal distribution of X is also normal, but the marginal distribution of Y is the Weibull distribution with the shape parameter equal to 2,5. Thus, this case represents the situation when both marginals are nearly symmetric. The results presented in Table 14 have been observed for the sample size n=100, and the averages have been calculated for the sets of differently dependent samples.

From Table 14 one can see that the observed variability of Kendall's τ is smaller than the variability of Pearson's r not only, as it has been expected, in the case of highly skewed variables, but also, in contrast to our supposition, in the case of nearly symmetric variables. Therefore, one can say that the variability of Kendall's τ is smaller than the variability of Pearson's r. This finding has been confirmed in another experiment in which standard deviations of both measures of dependence have been calculated from the samples for which the numerical values of both measures were the same. In Table 15 we present the results of such an experiment where samples of n=20 elements were generated from the Gumbel copula with the normal and exponential marginal distributions.

The average value of σ_τ is in this case equal to 0,09905, and the average value of σ_r is equal to 0,13442. Thus, the results presented in Table 15 are in the perfect agreement with our previous findings.

The entire analysis presented so far shows that in the considered cases of the marginal distributions that may be used in the problems of reliability prediction non-parametric measures of dependence, such as Kendall's τ, have better properties than Pearson's coefficient of linear correlation r. This is not only because of the ranges of possible values of r which may be highly misleading for practitioners, but also because of observed smaller variability. However, the question about the choice of the non-parametric statistic that is used for measuring dependence remains open.

The relationship between the most popular measures of dependence, Kendall's τ and Spearman's ρ, , both of which are considered in this paper, have been analyzed by many authors. Some important results, and references to other important papers, can be found in the paper by Fredricks and Nelsen [5]. The authors who considered this problem were interested more by the cases of weak and moderate dependence than in the cases of strong dependence, as they are more important in the context of the problem of reliability prediction. For example, Fredricks and Nelsen [5] proved the assertion previously formulated, in different versions, by other statisticians that Kendall's τ will be about two-thirds of the value of Spearman's ρ when the sample size n is large.

The results of our simulation experiments in which we have calculated not only the values of Kendall's τ, but the values of Spearman's ρ as well, let us analyze both measures in the whole spectrum of their possible values. In order to do so, we can use the same approximation methodology as that described in this section, and to find the approximate relationship $\rho(\tau)$. This relationship does not depend upon the types of the marginal distributions, but only on the type of the copula that describes the dependence. In Table 16 we present the coefficients in the expansion according to (29).

Similar coefficients for the expansion of $\rho(\tau)$ calculated from (35) for the Gumbel copula are presented in Table 17.

The impact of the type of copula on the relationship between Spearman's ρ and Kendall's τ is presented in Figure 7.

From Figure 7 one can see that the function $\rho(\tau)$ is nearly linear, for small and moderate absolute values of τ, and slightly non-linear in the case of strong dependence. The slope of the function $\rho(\tau)$ is for small and moderate values of τ fully determined by the first coefficient a_1 which is very close do 1,5. This gives a numerical confirmation of the theoretical results mentioned above. Moreover, the influence of the type of dependence is visible only in the case of the Clayton copula and the negative dependence of considered random variables. Therefore, in the case of the considered four copulas the function $\rho(\tau)$ is nearly the same.

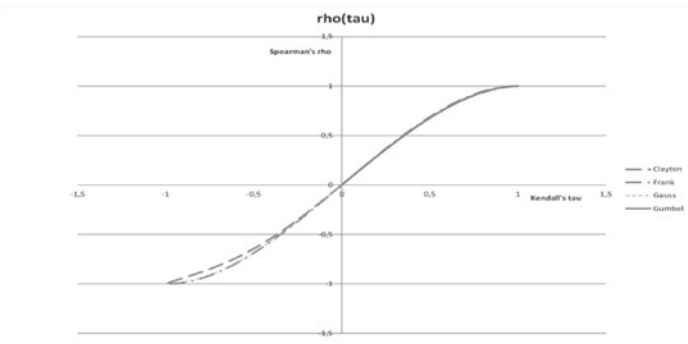


Fig. 7. Approximate relationship ρ(τ)

Because the values of ρ are greater than the respective values of τ one can think, using the same way of inference as it has been already used in this paper, that the observed variability of ρ should be smaller

Table 18. Comparison of the average values of standard deviations of ρ and τ (n=100)

Copula	σ_τ	σ_r
Clayton	0,047049	0,059035
Frank	0,040058	0,05053
Gauss (normal)	0,041423	0,05147
Gumbel	0,048311	0,060327

Table 19. Comparison of observed values of standard deviations of ρ and τ (same values of ρ and τ)

ρ=τ	σ_τ	σ_ρ
0,973269	0,018296	0,017962
0,922716	0,035034	0,045827
0,850858	0,055372	0,079796
0,760838	0,078261	0,115379
0,655911	0,101829	0,149192
0,539128	0,123897	0,178623
0,412939	0,142671	0,202589
0,279648	0,156311	0,219521
0,141375	0,16336	0,228719
0,070883	0,163837	0,230153

than the variability of τ. The results of the analysis presented in Table 18 do not confirm this claim.

In contrast to our supposition the average standard deviations of τ are visible smaller than the standard deviations of ρ. Moreover, it seems that their numerical value does not depend upon the type of the copula. Therefore, one can conclude that the empirical values of Kendall's τ are less variable than the respective values of Sperman's ρ. This is also confirmed in the results of the analysis presented in Table 19 for the case of the Gumbel copula, and the sample size equal to 20. The average value of σ_τ is in this case equal to 0,103887, and the average value of σ_r is equal to 0,146776. Thus, the results presented in Table 18 confirm our claim that Kendall's τ is, from a practical point of view, a more accurate (i.e. less variable) measure of dependence than Spearman's ρ.

5. Conclusions

Pearson's coefficient of linear correlation r is the measure of dependence most popular among practitioners despite the fact that its weaknesses have been known for more than one hundred years. In this paper we have investigated its applicability in the case of reliability prediction. In this particular practical problem the assumptions necessary for a good behavior of Pearson's r are obviously not fulfilled. However, it is not well known how the lack of the fulfillment of these assumptions influences the results of the analysis. Using some simple analytical methods and comprehensive computer simulations we have arrived at the following conclusions:

a) The observed values of Pearson's r may be completely misleading in the evaluation of the strength of dependence when the dependent variables are highly skewed, as is frequently the case in the reliability context;
b) When considered distributions are not very skewed Pearson's r can be used for the evaluation of the strength of dependence;
c) The same values of Pearson's r may describe different levels of strength of dependence depending upon the type of dependence defined by the type of the copula that describes the dependent random variables;
d) Non-parametric measures of dependence such as Spearman's ρ and Kendall's τ are better than Pearson's r when applied to the analysis of dependence of life-times;
e) Kendall's τ is better than a more popular Spearman's ρ, as its variability seems to be lower.

Therefore, in searching for the most informative variables that can be used for the prediction of reliability one should use Kendall's τ as the measure of dependence.

References

1. Clayton GG. A model for Association in Bivariate Life Tables and its Applications in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. Biometrika 1978; 65: 141 – 151.
2. Elsayed EA. Overview of Reliability Testing. IEEE Transactions on Reliability 2012; 61: 282-291.
3. Embrechts P, Lindskog F, McNeil A. Modelling Dependence with Copulas and Applications to Risk Management. In: Handbook of Heavy Tailed Distributions in Finance, S. Rachev (Ed.); 329-384, Amsterdam: Elsevier 2003 (also available as the ETHZ Report, Zurich, 2001).
4. Frank MJ. On the Simultaneous Associativity of F(x,y) and x+y-F(x,y). Æquationes Mathematicæ 1979; 19: 194 – 226.
5. Fredricks GA, Nelsen RB. On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. Journal of Statistical Planning and Inference 2007; 137: 2143-2150.
6. Genest C, MacKay J. Copules archimédiennes et families de lois bidimensionelles dont les marges sont données. Canadian Journal of Statistics 1986; 14: 145 – 159.
7. Genest C, Rivest L-P. Statistical Inference Procedures for Bivariate Archimedean Copulas. Journal of the American Statistical Association 1993; 88: 1034 – 1043.
8. Gumbel EJ. Distributions des valeurs extremês en plusieurs dimensions. Publications de l'Institut statstique de l'Universitè de Paris 1960; 9: 171 – 173.
9. Lee SW, Lee HK. Reliability prediction system based on the failure model for electronic components. Journal of Mechanical Science and Technology 2008; 22: 957-964.
10. Nelsen RB. Introduction to Copulas. New York: Springer, 2006.
11. Scarsini M. On measures of concordance. Stochastica 1984; 8: 201-218.
12. Schneidewind N. Predicting risk as a function of risk factors. Innovations in Systems Software Engineering 2005; 1: 63-70.
13. Sklar A. Fonctions de répartitions à n dimensions et leur marges. Publications de l'Institut de statistique de l'Universitè de Paris 1959; 8: 229-231.
14. Thaduri A, Verma AK, Gopika V, Gopinath R, Kumar U. Reliability prediction of semiconductor devices using modified physics of failure approach. International Journal of System Assurance Engineering and Management 2013; 4: 33-47.
15. US MIL-HDBK-217F Reliability Prediction of Electronic Equipment. National Technical Information Service; Springfield, Virginia, 1991.
16. US MIL-HDBK-217F Reliability Prediction of Electronic Equipment. Notice 2. National Technical Information Service; Springfield, Virginia, 1995.

17. Vořechovský M. Correlation control in small sample Monte Carlo type simulations II: Analysis of estimation formulas, random correlation and perfect uncorrelatedness. Probabilistic Engineering Mechanics 2012; 29: 105-120.
18. Wu S. Warranty Data Analysis: A Review. Quality and Reliability Engineering International 2012; 28: 795-805.
19. Xu W, Hou Y, Hung YS, Zou Y. (2013) A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models. Signal Processing 2013; 93: 261-276.

Olgierd HRYNIEWICZ
Systems Research Institute
Polish Academy of Sciences
and
Warsaw School of Information Technology
Newelska 6, 01-447 Warszawa, Poland
E-mail: hryniewi@ibspan.waw.pl

Janusz KARPIŃSKI
Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warszawa, Poland
E-mail: Janusz.Karpinski@ibspan.waw.pl

Article citation info:
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.

Yashun WANG
Chunhua ZHANG
Xun CHEN
Yuanyuan TAN

LIFETIME PREDICTION METHOD FOR ELECTRON MULTIPLIER BASED ON ACCELERATED DEGRADATION TEST

METODA PROGNOZOWANIA CYKLU ŻYCIA POWIELACZA ELEKTRONÓW OPARTA NA PRZYSPIESZONYCH BADANIACH DEGRADACJI

Electron multiplier (EM) is a kind of highly reliable and long-lifetime vacuum electronic device applied widely in spectrometry, space exploration and atom frequency standard. It is a critical device which might constrain the related technology. A challenge remains for researcher and engineer how to predict the life span of EM. Firstly, degradation mechanism of EM is investigated. It shows that the secondary emission ratios of each multiplier electrode reduces gradually with operating time, which results in the degradation of the key performance index of EM, i.e. the gain of electric current. So an accelerated degradation test (ADT) methodology using dual stresses is proposed to predict the life span of EM. Secondly, the ADT plan with dual stresses is designed and carried out by the corresponding test system established. Finally, the data analysis procedure is presented, and its validity is investigated by model verification. The presented method can sharply reduce testing time and cost because of using accelerated stress which can accelerate degradation process of EM. This method can also provide a new way to lifetime and reliability prediction for other products with long lifetime and high reliability.

Keywords: *electron multiplier, accelerated degradation test, lifetime prediction, reliability.*

Powielacz elektronów (EM) to elektroniczne urządzenie próżniowe o wysokiej niezawodności i długim cyklu życia, które znajduje szerokie zastosowanie w spektrometrii i badaniach przestrzeni kosmicznej, a także w atomowych wzorcach częstotliwości. Jest to urządzenie krytyczne, które może stanowić ograniczenie dla technologii, w której jest wykorzystywane. Wyzwaniem dla naukowców i inżynierów pozostaje pytanie, jak przewidzieć żywotność EM. W pierwszej kolejności w artykule zbadano mechanizm degradacji EM. Badanie pokazało, że współczynniki emisji wtórnej elektrody powielacza maleją stopniowo wraz z upływem czasu pracy, co prowadzi do degradacji kluczowego wskaźnika wydajności EM, to znaczy wzmocnienia prądu elektrycznego. W oparciu o ten fakt, zaproponowano metodę prognozowania żywotności EM zasadzającą się na metodologii przyspieszonych badań degradacji (ADT) z wykorzystaniem podwójnych naprężeń. Następnie zaprojektowano i zrealizowano plan ADT z podwójnymi naprężeniami za pomocą odpowiedniego systemu testowego. Na koniec przedstawiono procedurę analizy danych, a ich wiarygodność zbadano poprzez weryfikację modelu. Przedstawiona metoda może znacznie zredukować czas i koszty badań dzięki wykorzystaniu przyspieszonych naprężeń, które mogą przyspieszyć proces degradacji EM. Metoda ta może również umożliwić nowy sposób przewidywania niezawodności i cyklu życia produktów o długim cyklu życia i wysokiej niezawodności.

Słowa kluczowe: *powielacz elektronów, przyspieszone badanie degradacji, prognozowanie cyklu życia, niezawodność.*

1. Introduction

Electron multiplier (EM) is a kind of electronic device multiplying incident current for particles detection. It is widely applied in spectrometry, space exploration and atom frequency standard. And it is a critical device which might constrain the related technology.

The gain of EM will degrade gradually with operating time. It is regarded as a “failure” that the EM cannot multiplies incident current with a specified gain when the gain is below a threshold value. So lifetime of EM is usually defined as operating duration in use conditions before its gain degrades to the threshold. Lifetime of EM directly constrains operating lifetime of its engineering system, so it is a key problem for researchers and engineers how to predict the lifetime of EM in application.

EM is a kind of highly reliable products, and its gain degrades very slowly, so traditional accelerated life tests will provide little help, because no failures are likely to occur in a reasonable test duration, which brings a great challenge to data analysis procedure. Even if degradation tests are applied to lifetime prediction for EM, problems will still remain because decrease of EM gain is not obvious and it is impossible to obtain a good estimate within a reasonable period.

To overcome this problem, degradation data can be collected under higher levels of stress and allowing extrapolation the reliability and lifetime information at the use condition. This is called an accelerated degradation test (ADT) [5, 6]. ADT provides a new feasible approach to the problem that there is little or even a lack of failure data in application of accelerated life tests.

Researches about ADT have been widely focused on because of the advantages of ADT stated above. In statistical analysis and engineering application aspect, Nelson surveyed pertinent literature [6]. Meeker and Escobar proposed degradation models that correspond to physical-failure mechanisms and methods to estimate model parameters and confidence intervals for quantities of interest [4]. Whitmore, Park and Padgett presented degradation models based on Wiener and

Gauss stochastic process which can describe randomness of measurements well [7, 8, 10]. But these models are difficult to be widely applied for its complicated computation progress. In optimal design of test plan aspect, Yu and Tseng presented an on-line procedure for terminating an ADT [13]. Under the constraint that the total experimental cost does not exceed a pre-determined budget, Yu and Chiao studied the problem how to design an accelerated degradation test where the degradation rate follows a reciprocal Weibull distribution and a Normal distribution [11, 12]. Liao and Tseng proposed an approach to optimal design for step-stress ADT based on stochastic diffusion process [3]. Wang presented a simulation-based optimal design approach to constant stress ADT using mixed-effect degradation model to overcome problems in analytical optimal methods [9].

When ADT is applied to predict lifetime of products, the degradation mechanisms of the products at accelerated stress must accord with those at use conditions to ensure correctly extrapolating the reliability and lifetime information at accelerated stress to use condition. There always exists a limit for single accelerated stress above which degradation mechanisms will change. So if dual stresses are applied, we can not only obtain high test efficiency but also easily ensure the consistency of degradation mechanisms.

It is observed in pilot experiments that degradation speed of EM gain can be accelerated by the voltage U between electric poles of EM and by the intensity of incident current I. Therefore, a method is presented in this paper that lifetime prediction for EM is conducted based on ADT with dual stresses of U and I. Firstly, degradation mechanism of EM is investigated. It shows that the secondary emission ratios of each multiplier electrode reduces gradually with operating time, which results in the degradation of the key performance index of EM, i.e. the gain of electric current. So an accelerated degradation test (ADT) methodology using dual stresses is proposed to predict the life span of EM. Secondly, the ADT plan with dual stresses is designed and carried out by the corresponding test system established. Finally, the data analysis procedure is presented, and its validity is investigated by model verification.

2. Degradation Mechanisms of EM

2.1. Gain of EM

EM operates on the theoretic basis of electron secondary emission. The process comprises of the following three stages: (1) incident particles interacts with electrons in an emitter and a part of the electrons are stimulated to a higher energy level; (2) A part of stimulated electrons move towards the interface between the emitter and the vacuum; (3) the electrons arriving at the emitter surface whose energy is above the surface barrier are emitted into the vacuum.

A secondary emission ratio is a key performance index in a secondary emitting process of EM. It can be defined as the ratio of the number of secondary emission electrons N_2 to the number of the primary particles N_1:

$$\delta = N_2 / N_1 \tag{1}$$

The secondary emission ratio δ is a function of voltage E between electric poles of EM, i.e.:

$$\delta = a \cdot E^k \tag{2}$$

where a is a constant; k is an index determined by the structure and material of electric poles within the interval of 0.7~0.8 commonly [1].

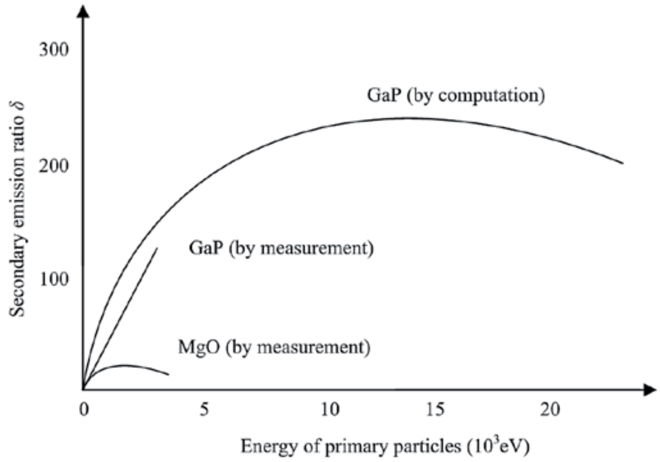


Fig. 1. The characteristics of secondary emission ratios of commonly used secondary emitters

The characteristics of the secondary emission ratios of commonly used secondary emitters are shown in Figure 1 [1]. It shows that secondary emission ratios increase firstly and then decrease with an increase of the energy of primary particles. When the energy of primary particles is low, stimulated electrons occur near the surface of an emitter. The probability of escape is large but the number of stimulated electrons is small, so δ is small. When the energy of primary particles increases, the number of stimulated electrons increases soon, but the probability of escape become small, and the overall effect is that δ increases. However, when the energy of primary particles is very large, stimulated electrons occur deeply in the emitter and the probability of escape becomes very small, so δ becomes small accordingly in general despite of large number of stimulated electrons.

If the number of electric particles is transformed to electric current, the secondary emission ratio of the ith electric pole can be defined as:

$$\delta_i = I_i / I_{i-1} \tag{3}$$

where I_{i-1} and I_i are the incident and output current of the ith electric pole respectively.

Write the incident current as I_{in}, and output current of the anode of EM as I_{out}, then:

$$I_{out} = I_{in} \cdot \alpha \cdot \delta_1 \cdot \delta_2 \cdots \delta_n \tag{4}$$

where α is the collecting efficiency of dynodes. So, the gain of EM G can be written as:

$$G = I_{out} / I_{in} = \alpha \cdot \delta_1 \cdot \delta_2 \cdots \delta_n \tag{5}$$

If α=1 and operating voltage of EM U is equally allocated to n dynodes, by combining (2) the relationship between G and U can be expressed as:

$$G = (a \cdot E^k)^n = a^n \left(\frac{U}{n+1}\right)^{kn} = A \cdot U^{kn} \tag{6}$$

where $A=a^n/(n+1)^{kn}$.

2.2. Degradation Failure of EM

The main failure mode of EM is that the gain of incident current decreases with operating time, i.e. gain degradation. The main reasons for degradation of the gain of EM include surface erosion of dynodes, small splitting crack inducing gas leakage, vaporization of material of secondary emitter, and the variation of vacuum atmosphere during operation [2].

Degradation rate of EM gain is related to the energy of incident particles [1]. Higher energy of incident particles accelerates the heating process of electric poles and then intensifies vaporization of material of secondary emitters or accelerates surface erosion of the emitters. So the degradation speed of gain become fast. Energy of incident particles is determined by intensity of incident current and voltage between dynodes. Bigger intensity of incident current and higher voltage between dynodes, higher the energy of incident particles can get. Figure 2 shows the curve of EM gain versus operating time under different intensity of incident current [9], where the corresponding operating current are respectively 1 μA, 4 μA and 10 μA. It can be seen from the figure that EM gain degrades faster if the intensity of incident current is bigger.

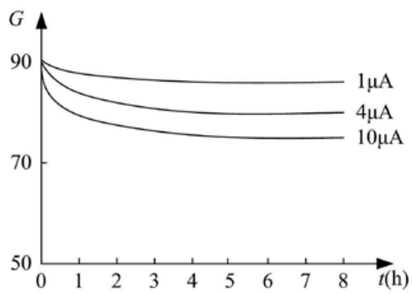


Fig. 2. Curve of EM gain versus operating time under different intensity of incident current

Therefore, degradation process of EM gain can be accelerated by increasing the intensity of incident current and voltage between dynodes in ADT for EM, which can shorten degradation lifetime of EM and reduce test cost consequently.

3. ADT Plan for EM

Voltage between dynodes U and intensity of incident current I are chosen as accelerated stresses on the basis of the analysis above. In order to reduce sample size and test cost, a fractional factorial design based on statistical design of experiment are applied. The design as shown in Table 1 is used by considering that higher accelerated stress levels can accelerate degradation rate of EM which induces higher test efficiency. The test plan comprises of 5 subtests of constant stress ADTs with dual stresses.

Table 1. Double stress ADT plan of EM

	D_1	D_2	D_3
U_1	—	—	(U_1, D_3)
U_2	—	—	(U_2, D_3)
U_3	(U_3, D_1)	(U_3, D_2)	(U_3, D_3)

According to the result of pilot experiment, the highest levels of accelerated stresses at which degradation mechanisms will not change from at use condition, i.e. voltage and intensity of incident current are taken the value of 2700V and 5×10^{-11}A respectively, i.e. U_3=2700V, D_3=5×10^{-11}A, above which degradation mechanisms will change. The lowest levels of accelerated stresses are set as the value near use condition as possible to ensure precisely extrapolation, so U_1=2500V, D_1=0.9×10^{-11}A. The intermediate levels of accelerated stresses are set as U_2=2600V, D_2=2×10^{-11}A for equal intervals.

An ADT system for EM is established according to the test design above. The system comprises of the following five parts as shown in Figure 3: ionization wire, focus electrode, power system of EM, output ammeter, temperature control system of cesium stove, vacuum system, and test tank of EM.

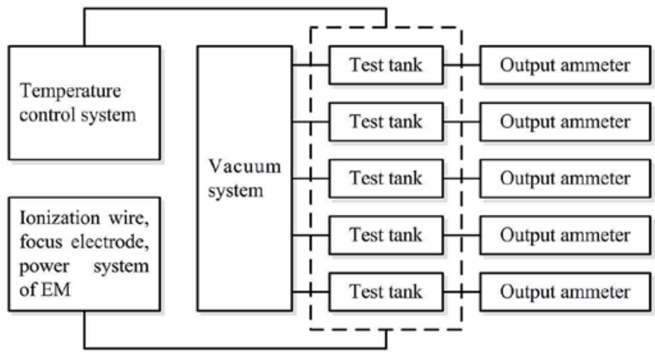


Fig. 3. ADT system of EM

Only 25 test units can be taken into the whole test under the constraint of test budget. Five units are allocated to each sub-test to meet basic requirements of statistical analysis. Performances of EM need effective inspection in ADT to obtain degradation path of test units during test process. The gain of EM is taken as the main index of performance that needs to be inspected in ADT. Equally spaced time interval is used to inspect the gain of EM for the convenience of data recording. ADT must be reasonably censored by the constraint of test time and cost. An approach of dynamic termination is applied in the ADT of EM that on-line and real-time ADT data are analyzed to obtain reasonable censored time [13]. If the relative rate of change of the asymptotic mean lifetime is smaller than a specified value, the whole test will be terminated.

4. Analysis Procedure for ADT Data of EM

4.1. ADT Data of EM

ADT for EM is conducted based on test system and test plan above. Test data as shown in Figure 4 are obtained. In the figure, y-coordinate is relative gain, i.e. the ratio of gain to its initial value, and x-coordinate is test time in hours). It shows that relative gains degrade gradually and there exist local fluctuations in each degradation curve, induced by variation of environmental factors such as temperature and vacuum. The gain of EM is sensitive to such factors, but small amplitude of fluctuation of gain has no significant effect on analysis results.

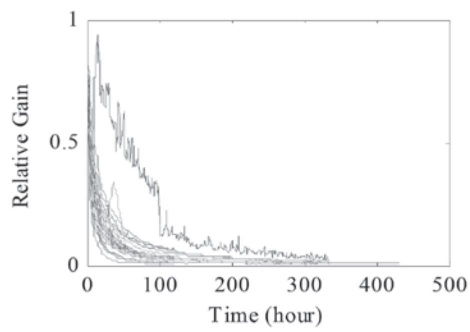


Fig. 4. ADT data of EM

4.2. Analysis Procedure

(1) The Problem

Levels of use stresses and accelerated stresses $S^{(1)}$ and $S^{(2)}$ are:

$$\begin{aligned} &S_0^{(1)} < S_1^{(1)} < \cdots < S_{l_1}^{(1)} \\ &S_0^{(2)} < S_1^{(2)} < \cdots < S_{l_2}^{(2)} \end{aligned} \tag{7}$$

where l_1 and l_2 are the numbers of levels of two stresses respectively. Use (i, j) to briefly denote a combination of S_i and S_j. The test plan is shown as Table 1 where S_i=U_i, S_j=D_j.

N test units sampled from a batch of products are allocated to each level combination of accelerated stresses with an equal number of units n, so

$$N = n \cdot L \tag{8}$$

where L is the total number of combinations of stress levels. The problem is how to analyze the degradation data obtained in the test to predict reliable lifetime of EM under use stress level.

(2) Model assumptions

A1 Theoretic degradation path under use stress level (0, 0) and accelerated stress level (i, j) can be expressed by a mixed-effect model of:

$$\begin{aligned} &G(t \mid S_i^{(1)}, S_j^{(2)}) = \exp(-\beta_{i,j,k} t^{\alpha}), \\ &t > 0, \\ &i,\ j = 0, 1, \cdots, l, \\ &k = 1, 2, \cdots, n \end{aligned} \tag{9}$$

where G is a function of relative gain; $\beta_{i,j,k}$ ($\beta_{i,j,k}$>0) is a parameter of random effects which denotes the degradation rate of No. k test unit under the combination of stress level (i, j) ; α (α>0) is a constant which denotes fixed effects. The reciprocal of $\beta_{i,j,k}$ follows Weibull distribution, i.e.

$$\beta_{i,j,k}^{-1} \sim \text{Weibull}(m, \eta_{i,j}) \tag{10}$$

where m is the shape parameter and $\eta_{i,j}$ is the scale parameter. So the observed degradation path can be expressed as:

$$H(t \mid S_i^{(1)}, S_j^{(2)}) = G(t \mid S_i^{(1)}, S_j^{(2)}) + \varepsilon_{i,j,k}(t) \tag{11}$$

where $\varepsilon_{i,j,k}$ is the measurement error which is independent from each other and follows normal distribution, i.e. $\varepsilon_{i,j,k} \sim N(0, \sigma_\varepsilon^2)$.

A2 The shape parameter m is independent from levels of the accelerated stresses so m keeps constant for different stress levels; the relationship between $\eta_{i,j}$ and the stress levels can be described by the following accelerated model:

$$\begin{aligned} &\ln \eta_{i,j} = a_0 + a_1\varphi_1(S_i^{(1)}) + a_2\varphi_2(S_j^{(2)}) + a_3\varphi_3(S_i^{(1)},\ S_j^{(2)}), \\ &i, j = 0, 1, \cdots, l \end{aligned} \tag{12}$$

where a_0, a_1, a_2, a_3 are the parameters to be estimated from test data; φ_1, φ_2, φ_3 are the known functions of accelerated stress levels. Eq. (12) can be written as a matrix formulation:

$$\mathbf{H} = \mathbf{J} \times \mathbf{a} \tag{13}$$

where $\mathbf{H}$=[ln$\eta_{0,0}$ ln$\eta_{0,1}$ ln$\eta_{1,1}$ … ln$\eta_{l,l}$]$^\mathrm{T}$, $\mathbf{a}$=[a_0 a_1 a_2 a_3]$^\mathrm{T}$, and

$$\mathbf{J} = \begin{bmatrix} 1 & \varphi_1(S_0^{(1)}) & \varphi_2(S_0^{(2)}) & \varphi_3(S_0^{(1)}, S_0^{(2)}) \\ 1 & \varphi_1(S_0^{(1)}) & \varphi_2(S_1^{(2)}) & \varphi_3(S_0^{(1)}, S_1^{(2)}) \\ 1 & \varphi_1(S_1^{(1)}) & \varphi_2(S_1^{(2)}) & \varphi_3(S_1^{(1)}, S_1^{(2)}) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \varphi_1(S_l^{(1)}) & \varphi_2(S_l^{(2)}) & \varphi_3(S_l^{(1)}, S_{l1}^{(2)}) \end{bmatrix}.$$

(3) Method for statistical analysis

Based on the assumption A1, theoretical degradation path of EMs in a dual constant stress accelerated degradation test (CSADT) is shown as Figure 5 where combinations of stress level are (0, 0), (0, 1), (1, 1) respectively.

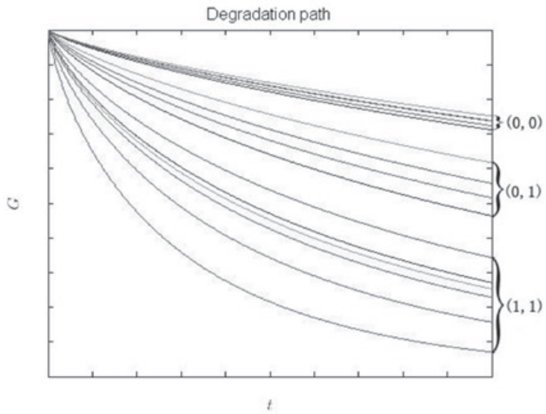


Fig. 5. Theoretical degradation path of EMs in a dual stress CSADT

The framework for solving the problem of statistical analysis consists of four major steps labeled (A)–(D) as follows:

(A) The estimation of α, $\beta_{i,j,k}$

The least-squares estimator (LSE) $\hat{\alpha}$ of $\hat{\beta}_{i,j,k}$, can be computed by minimizing:

$$SSE(\alpha, \beta_{i,j,k}) = \sum_t [H(t) - G(t)]^2 \tag{14}$$

where $H(t)$ is the observed degradation path as Eq. (11) and $G(t)$ is the theoretical degradation path as Eq. (9). And the variance of system measurement error σ_ε^2 can be estimated by:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{r - n_\nu} SSE(\alpha, \beta) \tag{15}$$

where r is total measurement times of each test units, and n_ν is the number of parameters of α, $\beta_{i,j,k}$.

(B) The estimation of $m_{i,j}$, $\eta_{i,j}$

Based on the $\hat{\beta}_{i,j,k}$ estimated in (A), the maximum likelihood estimator (MLE) $\hat{m}_{i,j}$, $\hat{\eta}_{i,j}$ of $m_{i,j}$, $\eta_{i,j}$ can be computed by Eq. (10). Because the distribution parameter m keeps constant for different stress levels as Assumption A2, m can be estimated by:

$$\hat{m} = \frac{1}{L}\sum_{i=1}^{l}\sum_{j=1}^{l}\hat{m}_{i,j} \tag{16}$$

(C) The estimation of a_0, a_1, a_2, a_3 in accelerated model

On the basis of the estimator of $\eta_{i,j}$, the estimation ($\hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{a}_3$) of (a_0, a_1, a_2, a_3) in accelerated model can be computed by:

$$\hat{\mathbf{a}} = (\mathbf{J}^{\mathrm{T}}\mathbf{J})^{-1}\mathbf{J}^{\mathrm{T}}\hat{\mathbf{H}} \tag{17}$$

where $\hat{\mathbf{H}} = [\ln\hat{\eta}_{0,0} \quad \ln\hat{\eta}_{0,1} \quad \ln\hat{\eta}_{1,1} \quad \cdots \quad \ln\hat{\eta}_{l,l}]^{\mathrm{T}}$, $\hat{\mathbf{a}} = [\hat{a}_0 \quad \hat{a}_1 \quad \hat{a}_2 \quad \hat{a}_3]^{\mathrm{T}}$.

(D) The estimation of the 100pth percentile of the EM's lifetime distribution

Let D denote the critical level for EM's degradation path under the use stress level (0, 0). The EM's lifetime τ is suitably defined as the time when the theoretical degradation path under use stress level (0, 0) crosses the critical level D for the first time, that is:

$$G[\tau \mid (0,0)] = D \tag{18}$$

From Eq. (9), τ can be expressed as:

$$\tau = \left(\frac{-\ln D}{\beta_{0,0,k}}\right)^{1/\alpha} \tag{19}$$

It can be derived by combining (10) that τ follows a weibull distribution as:

$$\tau \sim \mathrm{Weibull}(m_\tau, \eta_\tau) \tag{20}$$

Where $\hat{m}_\tau = \hat{\alpha}\hat{m}$; $\hat{\eta}_\tau = [\hat{\eta}_{0,0} \cdot (-\ln D)]^{1/\hat{\alpha}}$. Thus, the estimation of the 100pth percentile of EM's lifetime distribution can be expressed as follows:

$$\hat{\tau}_p = \hat{\eta}_\tau \cdot [-\ln(1-p)]^{1/\hat{m}_\tau} \tag{21}$$

(4) Analysis results of EM data in ADT

According to the method of data analysis stated above, results can be obtained as shown in Table 2. Analysis results show that the variance of estimation is small when accelerated model without interaction effects is applied, so estimation of a_3 is absent from Table 2.

The reliability curve of EM is shown in Figure 6 by accelerated model as Eq. (12), Eq. (20) and Eq. (21). And reliable life of EM under use stress level can be computed by the curve.

(5) Model checking

Analysis procedure above is based on the assumption of normal distribution of measurement errors, the assumption of Weibull distribution of reciprocal of degradation rate, and the assumption of normal distribution of fitting residual of accelerated model. All these assumptions are verified as follows.

Table 2. Estimation of model parameters in double stresses ADT of EM

$\hat{\alpha}$	$\hat{m}$	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{\sigma}_\varepsilon^2$
0.2747	3.904	12.29	-3.587	-0.6670	3.355×10^{-4}

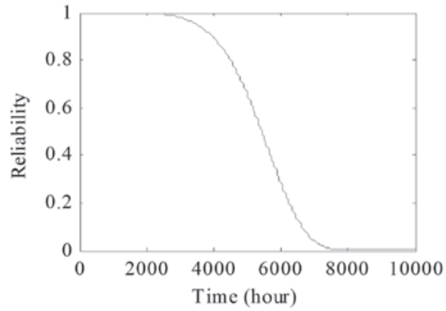


Fig. 6. Reliability curve of EM under use stress level

(A) Checking of normal distribution of measurement errors

Normal probability plot is used to validate the assumption that residuals generated in degradation path fitting process follow a normal distribution. The normal probability plot of the residuals of No.3 test unit under the stress level of (U_3, D_2) is shown in Figure 7. It can be seen from the figure that the residuals fit normal distribution well. The same conclusions can be obtained for other test units.

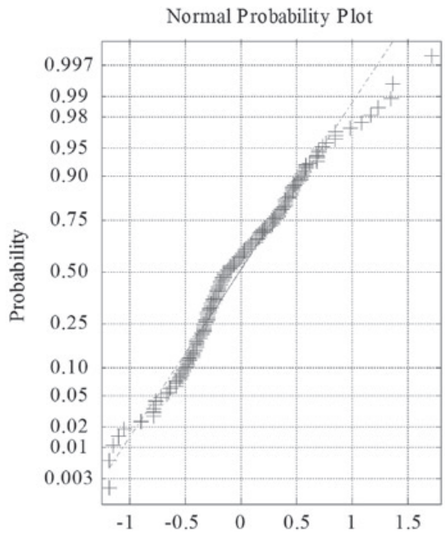


Fig. 7. Normal probability plot of residuals of No.3 test unit under a stress level of (U_3, D_2)(B) Checking of reciprocal Weibull distribution of degradation rates

(B) Checking of reciprocal Weibull distribution of degradation rates

Weibull probability plot is used to validate the assumption that reciprocal degradation rates follow a Weibull distribution. The Weibull probability plot of $\hat{\beta}_{i,j,k}^{-1}$ s under the stress level of (U_1, D_3) is shown in Figure 8. It can be seen from the figure that $\hat{\beta}_{i,j,k}^{-1}$ s fit Weibull distribution well. The same conclusions can be obtained for other combinations of stress levels.

(C) Checking of normal distribution of fitting residuals of accelerated model

Figure 9 shows the normal probability plot of residuals in accelerated model fitting. It can be seen from the figure that residuals

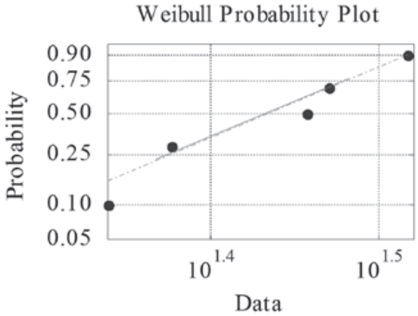


Fig. 8. Weibull probability plot of $\hat{\beta}_{i,j,k}^{-1}$ *s under a stress level of* (U_1, D_3)

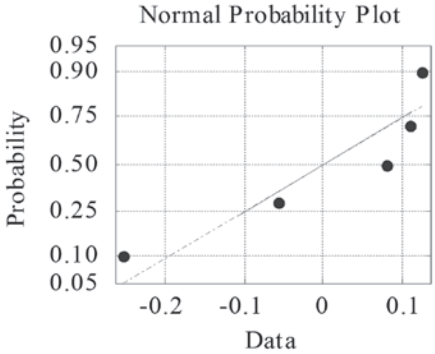


Fig. 9. Normal probability plot of fitting residuals of accelerated model

generated in accelerated model fitting process do not follow a normal distribution very good. But this will not affect the precision of point estimation of parameters in the accelerated model because the least-squares fitting method does not necessarily assume normally distributed errors when calculating parameter estimates. However, the method works best for data that does not contain a large number of random errors with extreme values. The normal distribution is one of the probability distributions in which extreme random errors are uncommon. And statistical results such as confidence and prediction bounds do require normally distributed errors for their validity. So it should be carefully to calculate interval estimation of parameters in the accelerated model and some other better methods can be applied in this case.

5. Conclusion

Electron multiplier (EM) is a kind of highly reliable and long-lifetime vacuum electronic device. It is a challenge in research of lifetime extending for EM and engineering application how to predict its operating lifetime. Analysis for degradation mechanism of EM indicates that secondary emission ratios of each multiplier electrode reduces gradually with operating time, which induces the key performance index of EM, i.e. the gain of electric current, to reduce gradually too.

Based on degradation mechanism above, an approach to lifetime prediction for EM using dual constant stresses accelerated degradation test (ADT) is presented together with the dual constant stress ADT plan, the data analysis procedure, and the model verification concerned. The applicability of accelerated degradation test to the long-term life span prediction for electron multiplier is demonstrated in this paper.

The presented methodology for EM life span prediction could reduce the testing duration and expense prominently. It may also be used as a reference for life span and reliability prediction for other similar products.

Acknowledgement*:*

The research work is financed by the National Natural Science Foundation of China in the years 2012-2014 as a research project (No. 51105367).

References

1. Chen CJ, Xu Z P. Photomultiplier. Beijing: Atomic Energy Press, 1988.
2. Kabushiki Kaisha of Editorial Committee of Hamamatsu Photonics. Photomultiplier – Base and Application. Kabushiki Kaisha of Digital Publishing Printing Institute, 2005.
3. Liao C M, Tseng S T. Optimal Design for Step-Stress Accelerated Degradation Tests. IEEE Transactions on Reliability 2006; 55(1): 59-66.
4. Meeker WQ, Escobar LA. A Review of Recent Research and Current Issues in Accelerated Testing. International Statistical Review 1993; 61(1): 147-168.
5. Meeker WQ, Escobar LA, Lu CJ. Accelerated degradation tests: Modeling and analysis. Technometrics 1998; 40(2): 89-99.
6. Nelson W. Accelerated Testing: Statistical Methods, Test Plans, and data Analysis. New York: John Wiley Press, 1990.
7. Park C, Padgett WJ. New Cumulative Damage Models for Failure Using Stochastic Processes as Initial Damage. IEEE Transactions on Reliability 2005; 54(3): 530-540.
8. Park C, Padgett WJ, Tomlinson MA. Accelerated degradation models for failure based on geometric Brownian motion and gamma processes. Lifetime Data Analysis 2005; 11: 511-527.
9. Wang YS, Zhang CH, Chen X. Simulation-based Optimal Design for Accelerated Degradation Tests with Mixed-effects Model. Chinese Journal of Mechanical Engineering 2009; 45(12): 108-114.
10. Whitmore GA, Schenkelberg F. Modelling accelerated degradation data using wiener diffusion with a scale transformation. Lifetime Data Analysis 1997; 3: 27-45.
11. Yu HF. Designing an accelerated degradation experiment with a reciprocal Weibull degradation rate. Journal of statistical planning and inference 2006; 136: 282-297.
12. Yu HF, Chiao CH. Designing an accelerated degradation experiment by optimizing the interval estimation of the mean time to failure. Journal of the Chinese Institute of Industrial Engineers 2002; 19(5): 23-33.
13. Yu HF, Tseng ST. On-line procedure for terminating an accelerated degradation test. Statistica Sinica 1998; 8: 207-220.

Yashun WANG
Chunhua ZHANG
Xun CHEN
Yuanyuan TAN
Laboratory of Science and Technology on Integrated Logistics Support
College of Mechatronics and Automation
National University of Defense Technology
Yanwachi str., 47 Changsha, 410073, P.R.China
E-mails: wangyashun@nudt.edu.cn, chzhang@nudt.edu.cn,
chenxun@nudt.edu.cn, tanyuanyuan06@nudt.edu.cn

Article citation info:
STELMASIAK Z. Uniformity of Diesel oil dosage in dual fuel engines. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2014; 16 (3): 491–495.

Zdzisław STELMASIAK

UNIFORMITY OF DIESEL OIL DOSAGE IN DUAL FUEL ENGINES

RÓWNOMIERNOŚĆ DAWKOWANIA OLEJU NAPĘDOWEGO W SILNIKACH DWUPALIWOWYCH*

Adaptation of compression ignition engines to dual fuel supply can be accomplished both in case of modern engines equipped with common rail system and older engines equipped with classic injection system (piston pumps). Due to big differences in price of gaseous and liquid fuels there is observed a natural tendency to use very small initial doses. At current level of introduction of gaseous fuels to powering of traction engines there is a need to provide alternating fuelling of the engine with the Diesel only and in dual fuel system. It requires usage of original injection systems in the dual fuel engines, what largely restricts possibility of reduction of the initial doses. In the paper are presented investigations concerning uniformity of the dosing by in-line piston pump of the P56-01 type and two types of injectors with common rail system. The investigations have shown that the P56-01 pump adjusted for nominal doses shows big non-uniformity of the dosage in area of small doses. Improved uniformity of the dosage can be attained in result of adjustment for a smaller doses, what allows reduction of the dose to about 15-20% of nominal dose and improves smoothness of engine operation. Also in case of the injectors in common rail system, reduction of the doses is limited due to worsening of uniformity of the dosing from one cycle to another, and failure of the dosing. It results from the fact, that minimal doses in dual fuel system are smaller than the ones present in case of idling speed when the engine is run on the Diesel oil only. In case of the injectors in common rail system, minimal initial doses possible to be obtained are equal to 10-15% of the nominal dose.

Keywords: *dual fuel engine, initial dose, uniformity of dosage, ignition of gas, combustion.*

Adaptacja silników o zapłonie samoczynnym do zasilania dwupaliwowego może być dokonywana zarówno dla nowoczesnych silników z systemami common rail jak i starszych silników wyposażonych w klasyczną aparaturę wtryskową (pompy tłoczkowe). Ze względu na duże różnice cen paliw gazowych i ciekłych występuje naturalna tendencja do stosowania bardzo małych dawek inicjujących. Na obecnym poziomie wprowadzania paliw gazowych do zasilania silników trakcyjnych istnieje konieczność zachowania przemiennego zasilania silnika samym olejem napędowym i w systemie dwupaliwowym. Wymaga to zastosowania w silnikach dwupaliwowych oryginalnej aparatury wtryskowej co ogranicza w znacznym stopniu możliwość zmniejszania dawek inicjujących. W pracy przedstawiono badania równomierności dawkowania tłoczkowej pompy rzędowej P56-01 oraz dwóch typów wtryskiwaczy układu common rail. Badania pokazały, że pompa P56-01 wyregulowana dla dawek znamionowych wykazuje dużą nierównomierność dawkowania w zakresie małych dawek. Poprawę równomierności dawkowania można uzyskać przez regulacją pompy dla dawek mniejszych, co pozwala zmniejszyć dawkę do około 15-20% dawki znamionowej i poprawia równomierność pracy silnika. Również w przypadku wtryskiwaczy common rail zmniejszanie dawek jest ograniczone z powodu pogorszenia równomierności dawkowania z cyklu na cykl i zaniku dawkowania. Wynika to z faktu, że minimalne dawki w systemie dwupaliwowym są mniejsze od występujących dla wolnych obrotów przy zasilaniu samym olejem napędowym. W przypadku wtryskiwaczy common rail minimalne dawki inicjujące jakie można uzyskać wynoszą 10-15% dawki znamionowej.

Słowa kluczowe: *silnik dwupaliwowy, dawka inicjująca, równomierność dawkowania, zapłon gazu, spalanie.*

1. Introduction

Dual fuel supply of compression ignition engines with CNG gas and Diesel oil results from many benefits offered by this type of fuelling. To the most important can be included:

- possibility to maintain the engine power output at unchanged level [6, 8],
- high engine efficiency [1, 3, 4, 6, 8, 12],
- possibility of alternate fuel supply – in dual fuel system or with Diesel oil only,
- possibility of combustion of gas-air mixtures in wide range of change of the excess air ratio [2, 8, 12],
- lower operational costs of the engine.

The last from the above mentioned factors is the most decisive when decision about gaseous fuelling of the engine is to be taken. It results from big difference in price of gaseous and liquid fuel, being worldwide on a similar level for decades. Generation of unit engine work in case of gaseous fuelling is two times cheaper than in case of fuelling with Diesel oil. Due to the above, in adaptation of the engines to dual fuel supply is seen a tendency to maximal reduction of liquid fuel consumption. It results in natural tendency to usage of small initial doses.

Size of the doses depends on whether the engine is to be run alternatively on Diesel oil and in dual fuel system, or serial injection system is to be used, or the engine is to be specially prepared to small doses. Depending on type of the adaptation, one uses the following initial doses, specified as percentage portion to energy of unit dose of Diesel oil in nominal conditions Q_{jzn}:

- $15 \div 25\% Q_{jzn}$ – when serial injection system and piston pumps are used [1, 6, 12],
- $10 \div 15\% Q_{jzn}$ – when serial high pressure systems of common rail type are used [13],
- about $5\% Q_{jzn}$ – when additional pumps and injectors of initial dose are used [8, 11],

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

- 1,0÷1,5%Q_{jzn} – when special apparatus for unit dose injected to ignition pre-chambers is used [9, 11].

At present prevalence of CNG filling stations network, issue of alternate engine operation in dual fuel system and in traditional system is very important, because it allows for continuous engine operation in case of shortage of the gas or failure of gas supply system, what in many applications is of a fundamental meaning.

Significant economic benefits coming from utilization of gaseous engines result in profitable adaptations of both modern engines with common rail systems and an older engines with classic injection systems [5, 7, 10].

In case of adaptation of compression ignition engines equipped with classic injection systems and remained traditional fuelling with Diesel oil only, serial injection pump and injectors are left, as a rule, in the engine. In multi-cylinder engines, uniformity of the dosing by the pump is adjusted at nominal doses, limiting size of maximal and minimal dose. At partial engine loads, size of the dose is changed simultaneously for all cylinders by change of position of toothed bar of the pump. Worsening of uniformity of the dosage by the pump in such conditions is not significant due to lower combustion pressures in the cylinders.

From application in dual fuel engine point of view, injection pump should fulfill the following conditions:

- should enable injection of possibly small initial dose,
- the sections should inject similar doses for nominal load,
- the sections should inject the same dose in such range of change of the rotational speeds, for which fuelling with the gas is anticipated.

In the paper is presented issue connected with uniformity of the dosing of Diesel oil when small doses are used in classic piston pumps (on example of the P56-01 pump) and in modern systems of common rail type.

2. Analysis of test results

Uniformity of the dosing by the P56-01 pump was investigated on the test bench, measuring average doses injected by individual sections, 1-6, for 1000 successive injections, with various adjustments of the toothed bar (what corresponded to various values of average dose calculated for all sections) and different rotational speeds of the pump.

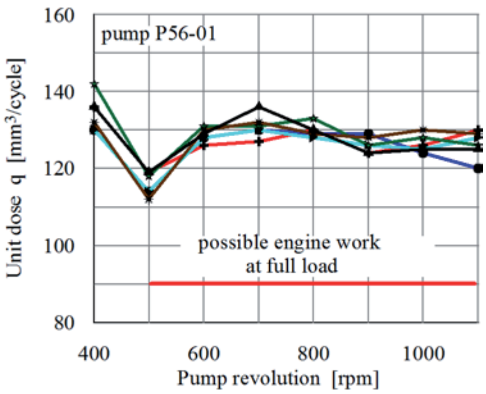


Fig. 1. Dosing of the sections, 1-6, of the P56-01 piston pump adjusted at nominal doses: maximal displacement of the toothed bar

Investigations of the P56-01 pump used in the SW680 engine have shown, that the pump adjusted at nominal dose and full displacement of the toothed bar assures satisfactory uniformity of the dosing in range of rotational speeds of 600÷1100 rpm (Fig. 1 – individual lines in the Fig. 1-3 correspond to values of the doses injected by individual sections), at which the engine can operate at full load. In range of rotational speeds specified above, a characteristic of the dosing is nearly flat, while change of size of the dose for an individual section doesn't exceed 13%.

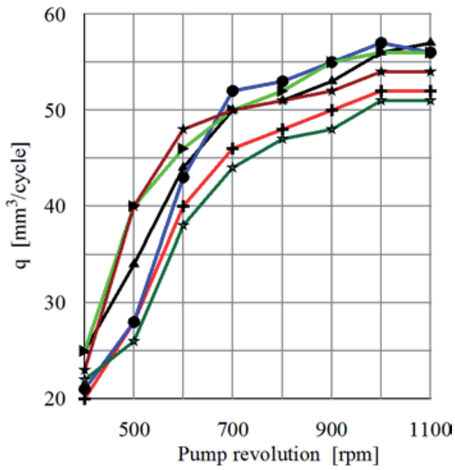
Together with reduction of the dose (Fig. 2) occurs a significant worsening of uniformity of the dosing, especially at lower rotational speeds. Doses injected by the sections decrease distinctly together with reduction of the rotational speed (especially clearly visible for rotational speed of 500 rpm), what may cause failure of the dosing at the lowest speeds. It results from increased portion of leaks in the sections in total volume of the fuel injected by the injectors. Such phenomena increase when the doses are smaller than 30 mm^3/cycle, Fig. 2c

If the dose is further diminished to average value of about 20 mm^3/cycle, failure of the dosing occurs at a higher rotational speed (Fig. 2d). It results in non-uniform engine operation and increased emission of the THC. The phenomena described here disable further reduction of the initial dose.

On the base of the test results presented in the Fig. 2 it is possible to ascertain that the P56-01 pump adjusted according to parameters specified by the producer doesn't comply with any condition required for dual fuel supply.

Improved dosing by the pump can be achieved by adjustment of dosing uniformity for a smaller dose. For instance, after performed adjustment for the dose 40 mm^3/cycle at 600 rpm, one obtained a considerable improvement of uniformity of the dosing at doses of 20÷25 mm^3/cycle, Fig. 3a. It enabled substantial improvement of engine operation when the engine was powered with the gas. The SW 680 en-

a)



b)

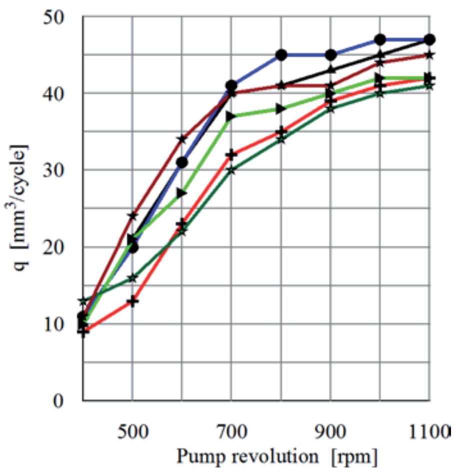


Fig. 2. Non-uniformity of the dosing by the sections, 1 – 6, of the P56-01 piston-type injection pump with relief valves – DV 86g1: a) adjustment of the toothed bar for about 48 mm^3/cycle of average dose for all sections at 700 rpm, b) adjustment of the toothed bar for average dose of about 35 mm^3/cycle at 700 rpm

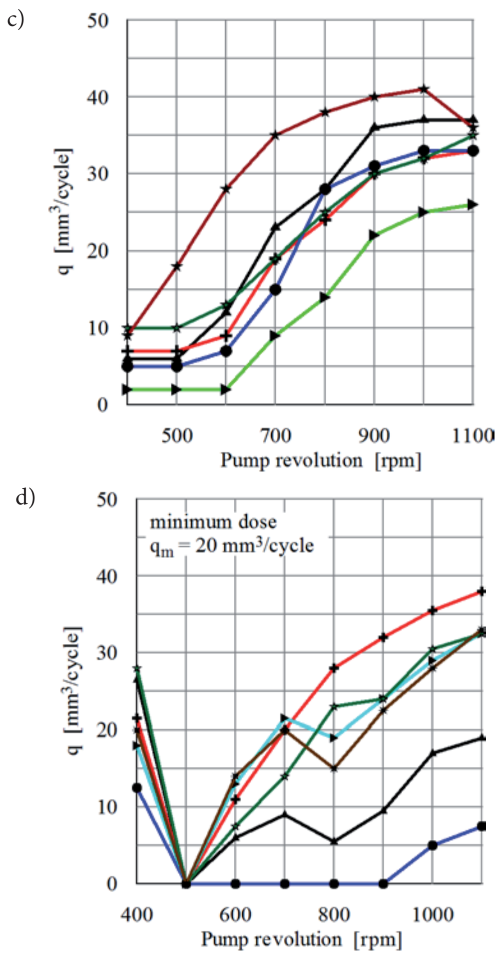


Fig. 2. (continued) Non-uniformity of the dosing by the sections, 1–6, of the P56-01 piston-type injection pump with relief valves – DV 86g1: c) adjustment of the toothed bar for average dose of about 25 mm³/cycle at 700 rpm, d) adjustment of the toothed bar for minimal dose

gine showed smooth operation in case of various rotational speeds and loads, both in steady states and during rapid changes of performance parameters.

Simultaneously, however, uniformity of the dosing by the pump was worsened after shift of the toothed bar to position of nominal dose (Fig. 3b), what corresponds to fuelling with the Diesel oil only. Differences in dosing by the pump for the nominal dose are presented in the Table 1.

Worsening of the dosing by the pump at a bigger doses should not create any serious problem in operation of dual fuel engines, because engine operation on the Diesel oil only should be considered as emergency fuelling (lacking of the gas or failure of the installation). Frequency of its usage should decrease together with development of CNG filling stations network. It is necessary therefore to suppose, that worsening of uniformity of the dosing at a bigger doses should not have any effect on operational durability of the engine due to small percentage of run on pure Diesel oil in total time of engine operation.

Proposed adjustment of the dosing by the P56-01 pump at a smaller doses enables substitution of liquid fuel in nominal conditions at the level of about 85%. However, due to possibility of overheating of the injectors by minimal initial dose, one proposes to increase it to about 20%Q_{jzn}. Comprehensive investigations performed on the SW 680 and SB3.1 engines by the author in the years 1995-2005 showed that use of such doses in conditions of long lasting engine operation at maximal loads doesn't result in any distinct symptoms of spray nozzles wear and overheating of the injectors. Such adjustment can be used, therefore, in traction conditions of engine operation.

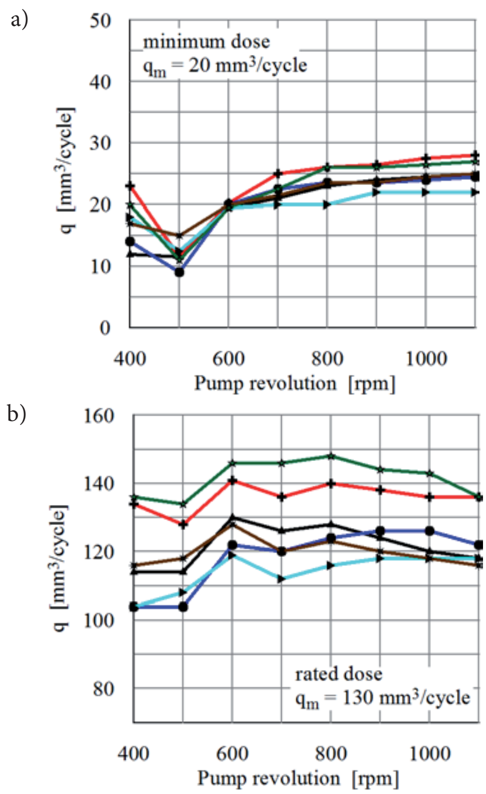


Fig. 3. Dosing by the sections, 1–6, of the P56-01 pump after adjustment of uniformity of the dosage for the dose of 40 mm³/cycle at 600 rpm: a) adjustment of the toothed bar for dose of about 20 mm³/cycle at 600 rpm, b) adjustment of the toothed bar for about 130 mm³/cycle at 600 rpm

Table 1. Differences in dosing of a section of the P56-01 pump after adjustment of uniformity of the dosing for the dose of 40 mm³/cycle at 600 rpm. Position of the toothed bar corresponding to nominal dose of 130 mm³/cycle

Pump revolution	q_{max}-q_{min}	(q_{max}-q_{min})/q_m
[rpm]	[mm³/cycle]	[%]
400	30	25,4
500	30	25,5
600	27	20,6
700	26	20,5
800	22	16,9
900	26	20,2
1000	25	19,7
1100	18	14,5

Use of common rail system with maintained condition of periodical operation of dual fuel engine on the Diesel oil only doesn't solve all problems connected with minimal initial dose. Indeed, use of factory injectors enables reduction of the dose to percentage of energetic portion of 10÷15%, but further reduction of the dose is connected with risk of non-uniformity, or failure of the dosing.

Investigations performed on two types of electromagnetic injectors used in an engines to Mercedes and Isuzu cars have shown, that at small doses a nonlinearity on characteristics of the dosage is present (Fig. 4), while at too small doses is seen a failure of the dosing. It disables decrease of the dose to a value required for robust ignition of the gas only. It denotes that when factory injectors are used, it isn't possible to obtain a high substitution of liquid fuel by the gas. This problem especially concerns partial engine loads, because dose of the

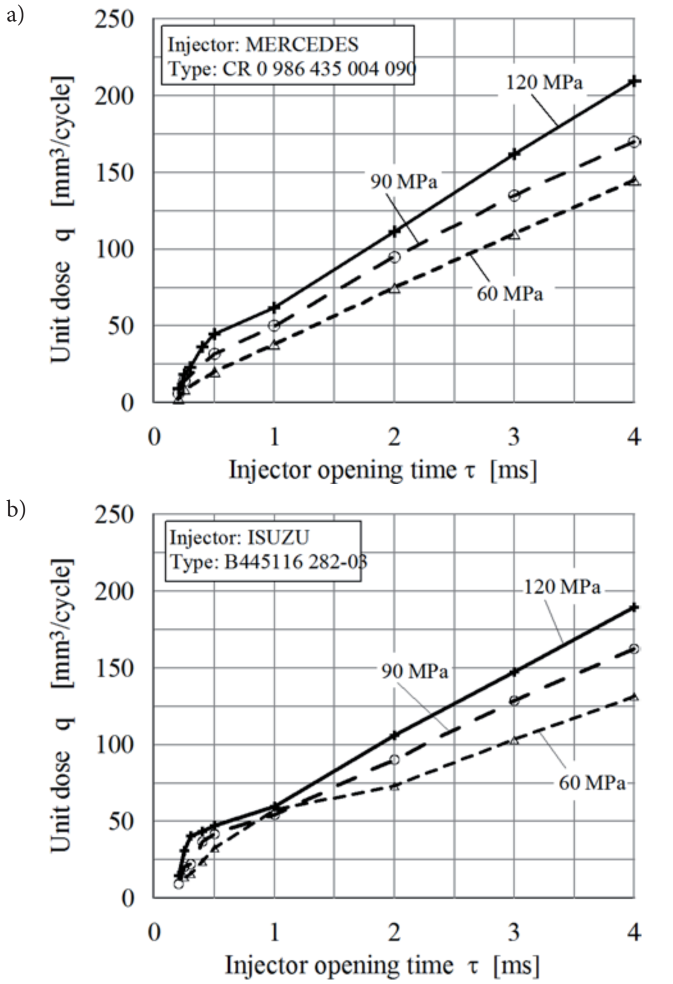


Fig. 4. Dosing by the injectors in common rail system in function of opening time of the injector for different pressures in fuel rail: a) injector from Mercedes engine, b) injector from ISUZU engine

Diesel oil remains constant due to hazard of interruptions in the dosing, and energy of the gas decreases. In traction engines such situation leads to substantial reduction of portion of the gas in operational consumption of the fuels.

3. Summary

On the base of performed investigations it is possible to make the following observations concerning fuel supply systems of dual fuel engines:

- at actual stage of introduction of the gases to fuelling of traction engines it is necessary to maintain versatility of powering of dual fuel engine, what denotes engine operation on pure Diesel oil only with maintained full engine performance;
- in case of liquid fuel powering, it is necessary to maintain correct characteristics of the injection, in range of injection angle and quality of atomization, in complete field of engine operation determined by change of engine rotational speed and load;
- in case of dual fuel operation, size of minimal dose of the liquid fuel should be selected to assure repeatable injection from one cycle to another, and proper atomization of the fuel in all conditions of engine operation;
- size of the initial dose should take into account thermal load of the injectors in nominal operational conditions;
- fulfillment of traditional and dual fuel supply conditions is possible by serial injection systems of the most engines regardless of injection systems used;
- in big stationary engines one should use a dual installations to injection of the main dose at traditional fuelling and to initial dose at dual fuel supply;
- in a high-speed stationary low-power engines, the injection systems used to fuelling with the Diesel oil should be left in the engine, and in case of dual fuel supply, size of minimal dose should be adapted to its injection capacity, taking into consideration repeatability and reliability of injections;
- in piston-type injection pumps, change of size of the dose occurs together with growth of rotational speed at constant position of the toothed bar; due to it, maintaining of constant initial dose requires changed position of the toothed bar together with change of rotational speed, what is difficult due to sensitivity of the adjustment;
- in small engines with classic injection systems it is recommended to use a fixed position of the control bar, at dual fuel supply, while size of the dose should be selected for less advantageous conditions of injectors' operation;
- adjustment of uniformity of the dosing by the pump should be performed for small doses, what improves quality of the dosing at dual fuel supply; occurring in such case worsening of dosage's quality when fuelling with pure Diesel oil should not have any effect on durability and increase of operational costs of the engine, because such type of engine fuelling should be considered as emergency mode;
- in high-pressure injection systems of common rail type at dual fuel supply, one should reduce injection pressure, what improves characteristics of the dosage and extends control time of opening of the injector.

References

1. Alla GHA, Soliman HA, Badr OA, Rabbo MFA. Effect of pilot fuel quantity on the performance of dual fuel engine. Energy Conversion & Management 2000; 41: 559-572.
2. Badr O, Karim GA. An examination of the flame spread limits in a dual fuel engine. Applied Thermal Engineering 1999; 19: 1071÷1080.
3. Barroso P, Ribas X, Domingues J, De Sella E, Garcia JM. Study of dual-fuel (diesel+natural gas) particle mater and CO_2 emissions of a heavy-duty diesel engine during transient operation. Combustion Engines; 2/2013 (153) ISSN 0138-0346: 28-39.
4. Beroun S, Martins J. The Development of gas (CNG, LPG and H_2) Engines for Buses and Trucks and their Emission and Cycle Variability Characteristics. SAE Paper 2001; 01-0144: 1-9.
5. Clark NN, Atkinson Chr M, Atkinson RJ, McDaniel T, Park T. Optimized Emission Reduction Strategies for Dual Fuel Compression Ignition Engines Running on natural Gas and Diesel. http://www.cemr.wvu.edu 2002: 1-6.
6. Friedeman Z.: Gasmotoren. Würzburg: Vogel Buchverlag, 2001.
7. Galyshev U, Magidovich I, Novichkov M, Fomin N. The Development and Improvement of Dual-Fuel Engines for Low-Sized Motor Ship. Journal of KONES Internal Combustion Engines 2000; 7 (1-2): 108-114.
8. Gebert K, Beck J, Barkhimer RL, Wong HCh. Strategies to Improve Combustion and Emission Characteristics of Dual-Fuel Pilot Ignited Natural Gas Engines. SAE Paper 1997; 971712: 79-87.

9. Jarg Chr, Sutkowski M. The Wärtsila 32GD engine for heavy gases. Combustion Engines 2009; 2/2009(137): 3-11.
10. Kowalewicz A. Adaptacja silnika wysokoprężnego do zasilania gazem naturalnym. Czasopismo Techniczne Wydawnictwo Politechniki Krakowskiej 2008; 7-M/2008: 67-78.
11. Schiffgens H, Brandt D, Dier L, Rieck K, Glauber R. Die Entwicklung des neuen MAN B&W Diesel-Gas-Motors 32/40 DG. MTZ Motortechnische Zeitschrift 1997; 58 (1997) 10: 584-590.
12. Stelmasiak Z. Studium procesu spalania gazu w dwupaliwowym silniku o zapłonie samoczynnym zasilanym gazem ziemnym i olejem napędowym. Bielsko-Biała: Wydawnictwo ATH w Bielsku-Białej, Rozprawy naukowe Nr 5, 2003.
13. Stelmasiak Z, Larisch J, Gilowski T, Matyjasik M. The optimization of combustion process in a dual fuel engine with Common Rail and gas injection systems. Combustion Engines 2007; 2007-SC2: 347-359.

Zdzisław STELMASIAK
Internal Combustion Engines an Automobiles Branch
Technical University of Bielsko-Biala
ul. Willowa 2, 43-309 Bielsko-Biala, Poland
E-mail: zstelmasiak@ath.bielsko.pl

INFORMATION FOR AUTHORS

Terms and Conditions of Publication:

- The quarterly „Maintenance and Reliability" publishes original papers written 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.

Fees:

- Pursuant to a resolution of the Board of PNTTE, as of 2009 the publication fee for one text is 600 zloty + VAT.
- Coloured graphical elements in the submitted text require agreement from the Editor and are charged extra.

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 graphical files in the *.tif, *.jpg or *.cdr (v. X3) formats.
- A manuscript should include (in accordance with the enclosed correct manuscript format: *.pdf, *.doc):
- 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.

Editor contact info: (Submissions should be sent to the Editor's address)

Editorial Office of „Eksploatacja i Niezawodność - 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 zloty and includes delivery costs.
Subscribers receive any additional special issues published during their year of subscription free of charge.

Orders

Subscription orders along with authorization to issue a VAT invoice without receiver's signature should be sent to the Editor's address.

Note

In accordance with the requirements of citation databases, proper citation of publications appearing in our Quarterly should include the full name of the journal in Polish and English without Polish diacritical marks, i.e.,

Eksploatacja i Niezawodnosc – Maintenance and Reliability.

No text or photograph published in „Maintenance and Reliability" can be reproduced without the Editor's written consent.

Wydawca:
Polskie Naukowo Techniczne
Towarzystwo Eksploatacyjne
Warszawa

Publisher:
Polish Maintenance Society
Warsaw

Członek:
Europejskiej Federacji
Narodowych Towarzystw
Eksploatacyjnych

Member of:
European Federation of National
Maintenance Societies

Patronat naukowy:
Polska Akademia Nauk
Oddział Lublin

Scientific Supervision:
Polish Academy of Sciences
Branch in Lublin