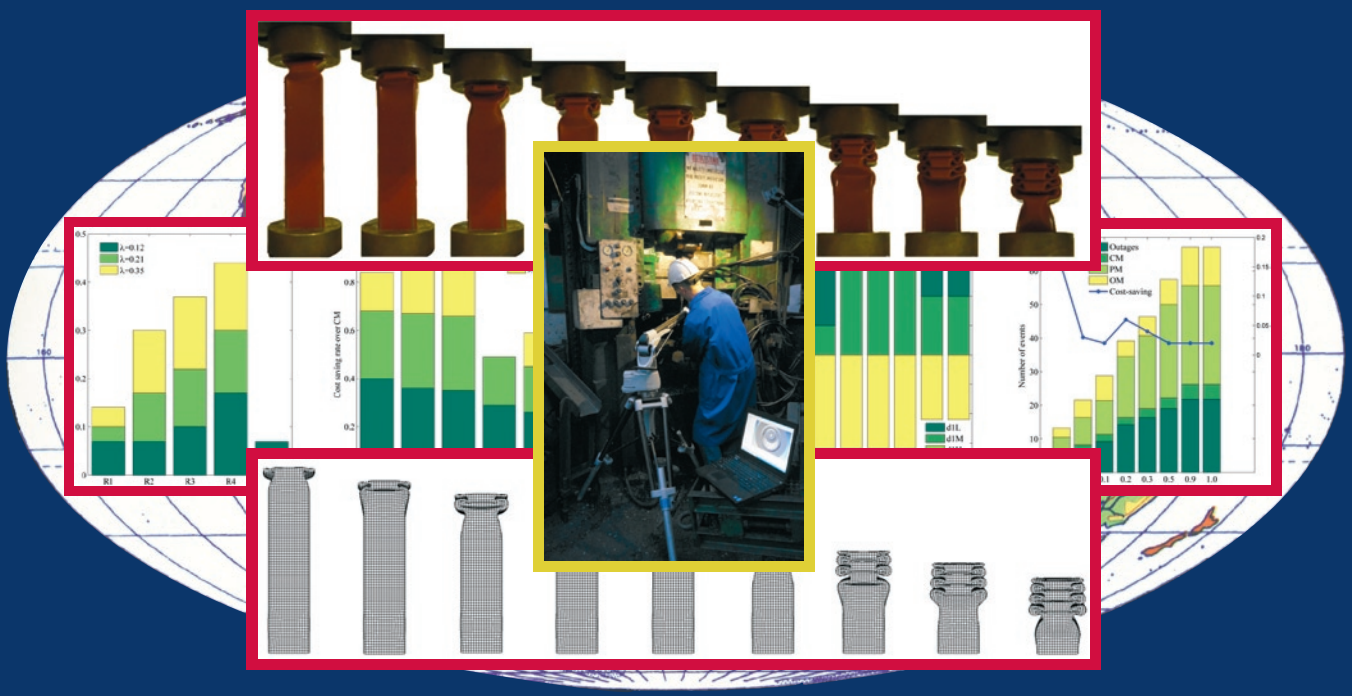
Vol. 18. No 2, 2016

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

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 0,983

Editorial staff:	Dariusz Mazurkiewicz, PhD, DSc (Eng), Associate Professor (Editor-in-Chief, Secretary of the Scientific Board) Tomasz Klepka, PhD, DSc (Eng), Associate Professor (Deputy Editor-in-Chief) Teresa Błachnio-Krolopp, MSc (Eng) (Editorial secretary) Andrzej Koma (Typesetting and text makeup) Krzysztof Olszewski, PhD (Eng) (Webmaster)
Publisher:	Polish Maintenance Society, Warsaw
Scientific patronage:	Polish Academy of Sciences Branch in Lublin
Address for correspondence:	"Eksploatacja i Niezawodność" – Editorial Office ul. Nadbystrzycka 36, 20-618 Lublin, Poland e-mail: office@ein.org.pl http://www.ein.org.pl/
Circulation:	550 copies

Science and Technology

BOŠNJAK S, ARSIĆ M, SAVIĆEVIĆ S, MILOJEVIĆ G, ARSIĆ D. **Fracture analysis of the pulley of a bucket wheel boom hoist system**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 155–163, http://dx.doi.org/10.17531/ein.2016.2.1.

This paper presents the results of the pulley fracture analysis. Experimental investigations confirmed that the chemical composition and basic mechanical properties of the pulley material, except the impact energy at a temperature of –20°C, meet the requirements of the corresponding standard. The impact energy value at the temperature of –20°C is for ≈45% lower than the prescribed value which has considerable influence on the appearance of the brittle fracture, especially having in mind the fact that the bucket wheel excavators operate at low temperatures. Metallographic examinations as well as magnetic particle inspections indicated that initial cracks in the welded joints occurred during the manufacture of the pulleys. Characteristic levels of the rope load cycle are obtained by using in-house software which includes the dynamic effects of the resistance-to-excavation. The FEA results pointed out that in the representative load cases the combinations of the mean stress and the alternating stress in the pulley critical zone lie considerably below the limit line of the modified Goodman's diagram. The conclusion, based on the presented results, is that the fracture of the pulley appeared as the result of the 'manufacturing-in' defects.

BOŠNJAK S, ARSIĆ M, SAVIĆEVIĆ S, MILOJEVIĆ G, ARSIĆ D. **Analiza pęknięć koła pasowego układu wciągarki wysięgnika koła czerpakowego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 155–163, http://dx.doi.org/10.17531/ein.2016.2.1.

Artykuł przedstawia wyniki analizy pęknięć koła pasowego. Badania doświadczalne potwierdziły, że skład chemiczny oraz podstawowe właściwości mechaniczne materiału, z którego zostało wykonane koło pasowe, za wyjątkiem energii udaru w temperaturze –20°C, były zgodne z odpowiednią normą. Wartość energii udaru w temperaturze –20°C była o ≈45% niższa od wartości zalecanej, co ma znaczący wpływ na występowanie pękania kruchego, zwłaszcza gdy weźmie się pod uwagę fakt, że koparki kołowe są przeznaczone do pracy w niskich temperaturach. Badania metalograficzne oraz badania magnetyczno-proszkowe wykazały, że pęknięcie pierwotne w połączeniu spawanym pojawiło się już w fazie produkcji koła pasowego. Charakterystyczne poziomy cyklu obciążenia liny uzyskano stosując własne oprogramowanie, które uwzględnia dynamiczne oddziaływanie odporności na urabianie. Wyniki MES pokazały, że w przypadku obciążeń reprezentatywnych, wartości średniego naprężenia w funkcji naprężenia zmiennego w strefie krytycznej koła pasowego były znacznie niższe niż wartości graniczne wyznaczone na podstawie zmodyfikowanego wykresu Goodmana. Na podstawie otrzymanych wyników stwierdzono, że pęknięcie koła pasowego powstało wskutek wad produkcyjnych.

KOPECKI T, MAZUREK P, LIS T. **The effect of the type of elements used to stiffen thin-walled skins of load-bearing aircraft structures on their operating properties**. Experimental tests and numerical analysis. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 164–170, http://dx.doi.org/10.17531/ein.2016.2.2.

The paper presents results of a study on thin-walled structures modelling representative fragments of aircraft fuselages subjected to bending and torsion. The type of the considered load and deformation corresponds to the state of such structures under in-flight conditions. The subject of the study were structures made of composite materials. Adopted assumptions include admissibility of post-buckling deformation in the operating load regime. Results of experimental studies are presented together with nonlinear numerical analyses carried out with the use of the finite elements method applied to a number of variant structures provided with various types of skin stiffening elements. Operating properties of the examined structures have been compared on the grounds of adopted criteria.

KOPECKI T, MAZUREK P, LIS T. **Wpływ rodzajów usztywnień pokryć cienkościennych struktur nośnych statków powietrznych na ich właściwości eksploatacyjne. Badania eksperymentalne i analiza numeryczna**. Experimental tests and numerical analysis. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 164–170, http://dx.doi.org/10.17531/ein.2016.2.2.

Praca prezentuje wyniki badań ustrojów cienkościennych, stanowiących modele reprezentatywnych fragmentów struktur lotniczych, poddawanych zginaniu oraz skręcaniu. Rodzaj obciążenia oraz deformacji odpowiada stanowi struktury w warunkach eksploatacji. Przedmiotem rozważań były ustroje wykonane z kompozytów. Przyjęto założenie o dopuszczalności deformacji zakrytycznych dla obciążeń eksploatacyjnych. Przedstawiono wyniki badań eksperymentalnych i nieliniowych analiz numerycznych w ujęciu metody elementów skończonych szeregu wariantów ustrojów, zawierających różne rodzaje usztywnień pokryć. Dokonano porównania właściwości eksploatacyjnych badanych ustrojów, w oparciu o przyjęte kryteria.

NI X, ZHAO J, SONG W, GUO C, LI H. **Nonlinear degradation modeling and maintenance policy for a two-stage degradation system based on cumulative damage model**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 171–180, http://dx.doi.org/10.17531/ein.2016.2.3.

This paper attempts to take into account a two-stage degradation system which degradation rate is non-stationary and change over time. The system degradation is thought to be caused by shocks, and system degradation model is established based on cumulative damage model. The nonlinear degradation process is expressed by different shock damage and shock counting. And shock damage and shock counting are assumed to be Gamma distribution and non-homogeneous Poisson process, respectively. On the basis of these, system reliability model and nonlinear degradation model are given. In order to optimal maintenance policy for considered system, adaptive maintenance policy and time-dependent maintenance policy are studied, and mean maintenance cost rate is established to evaluate the maintenance policies. Numerical examples are given to analyze the influences of degradation model parameters and find optimal maintenance policy for considered system.

NI X, ZHAO J, SONG W, GUO C, LI H. **System charakteryzujący się dwuetapowym procesem degradacji: nieliniowe modelowanie degradacji oraz wyznaczanie strategii eksploatacji systemu na podstawie modelu sumowania uszkodzeń**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 171–180, http://dx.doi.org/10.17531/ein.2016.2.3.

W przedstawionym artykule badano system, w którym proces degradacji zachodzi dwuetapowo, a szybkość degradacji jest zmienna w czasie. Przyjęto, że do degradacji systemu dochodzi w wyniku wstrząsów. Model degradacji systemu oparto na modelu sumowania uszkodzeń. Nieliniowy proces degradacji określono jako taki, w którym uszkodzenie powodowane wstrząsem oraz częstotliwość wstrząsów są wartościami zmiennymi. Przyjęto, że uszkodzenie powodowane wstrząsem ma rozkład gamma a częstotliwość wstrząsów jest niejednorodnym procesem Poissona. Na tej podstawie utworzono model niezawodności systemu oraz model degradacji nieliniowej. W celu opracowania optymalnej strategii eksploatacji dla rozpatrywanego systemu, rozważono dwa typy strategii utrzymania ruchu: strategię adaptacyjną oraz strategię czasowo-zależną. Strategie te oceniano określając średni poziom kosztów eksploatacji. Przykłady numeryczne posłużyły do analizy wpływu parametrów modelu degradacji oraz pozwoliły określić optymalną strategię utrzymania dla rozpatrywanego systemu.

PRZYBYŁEK P, SIODŁA K. **Application of capacitive sensor for measuring water content in electro-insulating liquids**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 181–185, http://dx.doi.org/10.17531/ein.2016.2.4.

The article discusses the problem of water content measurement in electro-insulating liquids using capacitive sensors. The article describes coefficients affecting reliability of the water content measurement. The authors discussed issues connected with water saturation limit in electro-insulating liquids. The authors also proposed a method which allows determining coefficients by means of which it is possible to calculate the water saturation limit in electro-insulating liquid as a function of temperature. Determining the coefficients allows proper calculating of the water content in ppm by weight by means of relative water saturation of the investigated liquid, what was measured with a capacitive probe. Propositions included in the article improve reliability of the method to determine water content in electro-insulating liquids and thus contribute to breakdown-free operation of electric power equipment insulated with these liquids.

PRZYBYŁEK P, SIODŁA K. **Zastosowanie czujnika pojemnościowego do pomiaru zawartości wody w cieczach elektroizolacyjnych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 181–185, http://dx.doi.org/10.17531/ein.2016.2.4.

W artykule omówiono problematykę pomiaru zawartości wody w cieczach elektroizolacyjnych przy wykorzystaniu czujników pojemnościowych. Opisano czynniki wpływające na wiarygodność pomiaru zawartości wody. Autorzy pracy omówili zagadnienia związane z granicznym nasyceniem cieczy elektroizolacyjnych wodą. W artykule zaproponowana została metoda umożliwiająca wyznaczenie współczynników, za pomocą których możliwe jest obliczenie granicznego nasycenia cieczy elektroizolacyjnej wodą w funkcji temperatury. Wyznaczenie współczynników umożliwia poprawne obliczenie zawartości wody w ppm wagowo za pomocą zmierzonego sondą pojemnościową względnego nasycenia badanej cieczy wodą. Propozycje zawarte w artykule poprawiają niezawodność metody wyznaczania zawartości wody w cieczach elektroizolacyjnych, a przez to przyczyniają się do bezawaryjnej eksploatacji urządzeń elektroenergetycznych izolowanych tymi cieczami.

ESTRADA Q, SZWEDOWICZ D, MAJEWSKI T, MARTINEZ E, RODRIGUEZ-MENDEZ A. **Effect of quadrilateral discontinuity size on the energy absorption of structural steel profiles**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 186–193, http://dx.doi.org/10.17531/ein.2016.2.5.

In this paper the effect of discontinuity size on energy absorption performance of steel square profiles is reported. The analysis consists of finite element simulations and experimental results of the compression strength of steel profiles with discontinuities. The discontinuities were placed at the mid span of the profiles in two walls opposite to each other. Square, rectangular and diamond initiators were evaluated at different scales. The numerical results determined the size intervals that present a good energy absorption performance in each case. Energy absorption capabilities were increased up to 12.54% with respect to a structure without discontinuities. Additionally, the peak load value (Pmax) was decreased 25.97% with the implementation of a diamond initiator. For structures with discontinuities with major axis close to the profile width, a buckling effect was observed. Finally, it was observed that the size of the initiators contributes to reduce the peak load (Pmax) value.

ESTRADA Q, SZWEDOWICZ D, MAJEWSKI T, MARTINEZ E, RODRIGUEZ-MENDEZ A. **Wpływ wielkości czworobocznych nieciągłości na pochłanianie energii w stalowych profilach strukturalnych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 186–193, http://dx.doi.org/10.17531/ein.2016.2.5.

W pracy przedstawiono analizę wpływu rozmiaru nieciągłości na pochłanianie energii przez stalowe profile o przekroju kwadratowym. Analiza przedstawia wyniki symulacji elementami skończonymi próby ściskania profili stalowych z nieciągłościami oraz porównanie z danymi eksperymentalnymi. Nieciągłości zostały usytuowane w środku profilu w dwóch przeciwległych ścianach. W pracy zostały przebadane nieciągłości o formach kwadratowych, prostokątnych i rombowych dla różnych wymiarów. Stwierdzono wzrost o 12,54% możliwości pochłaniania energii w porównaniu dla struktur bez nieciągłości. Dodatkowo, w przypadku nieciągłości rombowych stwierdzono spadek wartości siły maksymalnej (Pmax) o 25,97%. Zaobserwowano występowanie efektu wyboczenia dla nieciągłości rombowej gdy wymiar jej osi zbliża się do szerokości profilu. Zaobserwowano, ze rozmiar nieciągłości wpływa na redukcje wartości maksymalnego obciążenia oraz w tym samym czasie na obniżenie pochłanianej energii.

GRONOSTAJSKI Z, HAWRYLUK M, KASZUBA M, ZIEMBA J. **Application of a measuring arm with an integrated laser scanner in the analysis of the shape changes of forging instrumentation during production**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 194–200, http://dx.doi.org/10.17531/ein.2016.2.6.

In the article, the authors present an innovative approach consisting in an analysis of the wear (shape changes) of one of the forging tools directly during production, with the use of a laser scanner, without the necessity of their disassembly from the forging unit. The tests consisted in direct measurements of the shape changes of cyclically sampled forgings during the forging process (every 1000 item), and next, based on the proceeding wear, an indirect analysis was performed of the shape change of the impression of the selected tool, i.e. a filler. At the time of the short technological intervals in the process, direct measurements were performed of the tool itself, with the purpose of verifying the results of the forgings' measurement in relation to the actual changes in the tool. The performed analyses showed a good agreement of the geometrical properties of the surfaces (of the selected forgings representing the proceeding wear of the tool) and the geometrical defect of the working impression of the tool, based on the direct measurements during the production process. The obtained results allow for a fast analysis of the forging tool life with respect to the quality and the quantity (of material defect), which, in consequence, leads to significant economical savings. The proposed method makes it possible to make decisions on the time period of the tool operation based on the tools' actual wear, instead of – as has been the case in forging plants up till now – after the given maximal number of forgings has been made or a premature tool damage has been observed.

GRONOSTAJSKI Z, HAWRYLUK M, KASZUBA M, ZIEMBA J. **Zastosowanie ramienia pomiarowego ze zintegrowanym skanerem laserowym do analizy zmian kształtu oprzyrządowania kuźniczego w trakcie produkcji**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 194–200, http://dx.doi.org/10.17531/ein.2016.2.6.

W artykule autorzy przedstawili innowacyjne podejście polegające na analizie zużywania się (zmian kształtu) jednego z narzędzi kuźniczych bezpośrednio w trakcie produkcji przy wykorzystaniu skanera laserowego, bez konieczności ich demontażu z agregatu kuźniczego. Badania polegały na bezpośrednich pomiarach zmian kształtu cyklicznie pobieranych odkuwek podczas procesu kucia (co 1000szt), a następnie na podstawie postępującego zużywania dokonywana była w sposób pośrednia analiza zmian kształtu wykroju wybranego narzędzia - wypełniacza. Natomiast w momencie krótkich przerw technologicznych w procesie przeprowadzano bezpośrednie pomiary samego (analizowanego) narzędzia w celu weryfikacji wyników pomiaru odkuwek w stosunku do rzeczywistych zmian narzędzia. Przeprowadzone analizy wykazały dużą zgodność cech geometrycznych powierzchni (wybieranych odkuwek odzwierciedlających zużywanie się narzędzia), a ubytkiem geometrycznym wykroju roboczego narzędzia na podstawie bezpośrednich pomiarów podczas produkcji. Uzyskane rezultaty pozwoliły na dokonanie szybkiej analizy trwałości narzędzia kuźniczego pod względem jakościowym i ilościowym (ubytku materiału), co w konsekwencji prowadzi do znacznych oszczędności. Zaproponowana przez autorów metoda pozwala na podejmowanie decyzji o czasie eksploatowania narzędzi na podstawie ich rzeczywistego zużycia, a nie, jak to ma miejsce obecnie w kuźniach, po określonej maksymalnej ilości wykonanych odkuwek lub zaobserwowanego w tym okresie przedwczesnego uszkodzenia narzędzia.

ROMAŃSKI L, BIENIEK J, KOMARNICKI P, DĘBOWSKI M, DETYNA J. **Operational tests of a dual-rotor mini wind turbine**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 201–209, http://dx.doi.org/10.17531/ein.2016.2.7.

The article presents the results of wind-tunnel tests and field studies of a mini dual-rotor wind turbine. The first stage involved testing of an open-circuit wind tunnel built with the aim of performing laboratory tests. The coefficient of uneven air stream distribution at a rated speed was 1.7%, while the index of turbulence intensity in the entire measurement range was between 1.2 and 1.8%. The mini wind turbine was equipped with rotors with new design blades. Compared to the blade designs used in mini wind turbines available on the market, the blades used in the present study were characterized by an efficiency of 0.28. The results of performance tests in the wind tunnel were evaluated statistically using Pearson correlation coefficients and Spearman's rank. We examined the relationship between a dependent variable (power P) and independent variables (average air stream speed V, incidence angle of the blades of the first rotor α1, incidence angle of the blades of the second rotor α2, and the distance between the rotors l). The analysis showed, as expected, that the strongest correlation was between power and speed of the air stream. While incidence angles of the two rotors also affected the turbine's power, no such effect was observed for changes in the distance between the rotors. Field tests confirmed the findings and observations made in the wind tunnel.

ROMAŃSKI L, BIENIEK J, KOMARNICKI P, DĘBOWSKI M, DETYNA J. **Badania eksploatacyjne dwuwirnikowej mini elektrowni wiatrowej**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 201–209, http://dx.doi.org/10.17531/ein.2016.2.7.

W artykule przedstawiono wyniki badań dwuwirnikowej mini elektrowni wiatrowej przeprowadzone w tunelu aerodynamicznym oraz w terenie. W pierwszym etapie testowano zbudowany w celu przeprowadzenia badań laboratoryjnych tunel aerodynamiczny o konstrukcji otwartej. Wyznaczony współczynnik nierównomierności strugi powietrza przy prędkości nominalnej wynosił 1,7%, natomiast wskaźnik intensywności turbulencji w całym zakresie pomiarowym zawierał się w granicach 1,2-1,8%. Budując mini elektrownię wiatrową wyposażono ją w wirniki w których zastosowano nową konstrukcję łopat. Zastosowane łopaty w porównaniu do zbliżonej konstrukcji łopat stosowanych w mini elektrowniach dostępnych na rynku charakteryzowały się sprawnością wynoszącą 0,28. Po wykonanych badaniach eksploatacyjnych w tunelu aerodynamicznym uzyskane wyniki poddano ocenie statystycznej z wykorzystaniem współczynników korelacji liniowej Pearsona oraz rangi Spearmana. Zbadano zależności między zmienną zależną (moc P) oraz zmiennymi niezależnymi (średnie prędkości strugi powietrza V, kąt zaklinowania łopat pierwszego wirnika α1, kąt zaklinowania łopat drugiego wirnika α2, odległości pomiędzy wirnikami l). Na podstawie analizy, zgodnie z oczekiwaniem, stwierdzono, że najsilniejsza korelacja występuje w odniesieniu do prędkości strugi powietrza. Wpływ na moc mają także kąty zaklinowania na obu wirnikach, natomiast nie stwierdzono takiego wpływu w przypadku zmian odległości pomiędzy wirnikami turbiny. Badania w terenie potwierdziły ustalenia i spostrzeżenia poczynione w tunelu aerodynamicznym.

SHEN Q, QIU J, LIU G, LV K. **Intermittent fault's parameter framework and stochastic petri net based formalization model**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 210–217, http://dx.doi.org/10.17531/ein.2016.2.8.

The intermittent fault widely exists in many products and brings high safety risk and maintenance cost. At present there are some different opinions on the notion of intermittent fault and there is no comprehensive parameter framework for fully describing intermittent fault. Also the formalization model which can mathematically describe intermittent fault hasn't been constructed. In this paper, the conception of intermittent

SHEN Q, QIU J, LIU G, LV K. **Model parametryczny niezdatności przejściowej oraz model formalny oparty na stochastycznej sieci Petriego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 210–217, http://dx.doi.org/10.17531/ein.2016.2.8.

Niezdatność przejściowa charakteryzuje wiele produktów i pociąga za sobą wysokie zagrożenie bezpieczeństwa oraz wysokie koszty eksploatacji. Obecnie istnieje wiele poglądów na temat pojęcia niezdatności przejściowej; nie stworzono jednak kompleksowego modelu parametrycznego pozwalającego w pełni opisać zjawisko niezdatności przejściowej. Nie skonstuowano także modelu formalnego, za pomocą którego można by opisać niezdatność

fault is discussed. A new definition of intermittent fault is put forward. Then the intermittent fault's parameter framework is presented. After that, the Stochastic Petri Net (SPN) based formalization model for intermittent fault is constructed. Finally an application of the SPN formalization model is shown. The parameters for intermittent fault are computed based on the proposed model and a case study is presented. The result shows the validity of the model. The model could assist the further research such as intermittent fault diagnosis and prognostic of remaining life.

przejściową w kategoriach matematycznych. W pracy omówiono koncepcję niezdatności przejściowej. Zaproponowano nową definicję tego pojęcia a następnie przedstawiono model parametryczny niezdatności przejściowej. Skonstruowano także model formalny niezdatności przejściowej oparty na stochastycznej sieci Petriego (SPN). Wreszcie, pokazano zastosowanie formalizacji SPN. Na podstawie zaproponowanego modelu obliczono parametry dla niezdatności przejściowej. Przedstawiono także studium przypadku. Otrzymane wyniki potwierdzają wiarygodność modelu. Opracowany model może być pomocny w dalszych badaniach dotyczących problemów, takich jak diagnozowanie niezdatności przejściowej czy prognozowanie pozostałego okresu użytkowania produktu.

KOZIELSKI M, SIKORA M, WRÓBEL Ł. **Decision support and maintenance system for natural hazards, processes and equipment monitoring**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 218–228, http://dx.doi.org/10.17531/ein.2016.2.9.

This paper presents the DISESOR integrated decision support system and its applications. The system integrates data from different monitoring and dispatching systems and contains such modules as data preparation and cleaning, analytical, prediction and expert system. Architecture of the system is presented in the paper and a special focus is put on the presentation of two issues: data integration and cleaning, and creation of prediction model. The work contains also two case studies presenting the examples of the system application.

KOZIELSKI M, SIKORA M, WRÓBEL Ł. **System wspomagania decyzji dla monitorowania zagrożeń naturalnych, procesów i urządzeń**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 218–228, http://dx.doi.org/10.17531/ein.2016.2.9.

W pracy przedstawiono zintegrowany system wspomagania decyzji DISESOR oraz jego zastosowania. System pozwala na integrację danych pochodzących z różnych systemów monitorowania i systemów dyspozytorskich. Struktura systemu DISESOR składa się z modułów realizujących: przygotowanie i czyszczenie danych, analizę danych, zadania predykcyjne oraz zadania systemu ekspertowego. W pracy przedstawiono architekturę systemu DISESOR, a szczególny nacisk został położony na zagadnienia związane z integracją i czyszczeniem danych oraz tworzeniem modeli predykcyjnych. Działanie systemu przedstawione zostało na dwóch przykładach analizy dla danych rzeczywistych.

LU J-M, LUNDTEIGEN MA, LIU Y, WU X-Y. **Flexible truncation method for the reliability assessment of phased mission systems with repairable components**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 229–236, http://dx.doi.org/10.17531/ein.2016.2.10.

Phased-mission systems (PMS) are the system in which the component stresses and the system configuration may change over time. Real-world PMS usually consist of a large number of repetitive phases and repairable components. Existing approaches for the reliability analysis of this kind of PMS tend to suffer from the problem of state explosion or binary-decision-diagram (BDD) explosion. This paper presents a truncation method based on the BDD and Markov chains to solve the scaling issue. In our approach, the truncation mitigates the BDD explosion and broadens the applicability of the BDD & Markov method. Different from the classic truncations, our truncation limit is flexible, which ensures that ensure the truncation error is lower than the predefined threshold. The advantages of the proposed method are illustrated through two practical PMS which are challenging to classic non-simulation approaches.

LU J-M, LUNDTEIGEN MA, LIU Y, WU X-Y. **Zastosowanie metody elastycznego obcięcia do oceny niezawodności systemów o zadaniach okresowych z elementami naprawialnymi**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 229–236, http://dx.doi.org/10.17531/ein.2016.2.10.

Systemy o zadaniach okresowych (phased mission systems, PMS) to takie systemy, w których naprężenia elementów składowych oraz konfiguracja systemu mogą z czasem ulegać zmianie. W warunkach rzeczywistych, PMS zazwyczaj charakteryzują się dużą liczbą powtarzalnych faz zadaniowych i składają się z wielu naprawialnych elementów. Istniejące metody analizy niezawodności tego typu systemów niestety posiadają ograniczenia związane z problemem eksplozji stanów lub eksplozji diagramów binarnych decyzji (binary decision diagram, BDD) Praca przedstawia metodę obcinania opartą na BDD oraz łańcuchach Markowa, która pozwala rozwiązać wspomniane problemy złożoności obliczeniowej. W proponowanym podejściu, obcięcie minimalizuje eksplozję BDD zwiększając możliwości zastosowania metody opartej na BDD oraz łańcuchach Markowa. W odróżnieniu od klasycznego obcinania, w opracowanej przez nas metodzie granica obcięcia jest elastyczna co pozwala zredukować błąd obcięcia poniżej wcześniej określonego progu. Zalety proponowanej metody zilustrowano na przykładzie dwóch stosowanych w praktyce systemów PMS, które stanowią wyzwanie dla klasycznych metod niesymulacyjnych.

TOPIĆ D, ŠLJIVAC D, STOJKOV M. **Reliability model of different wind power plant configuration using sequential Monte Carlo simulation**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 237–244, http://dx.doi.org/10.17531/ein.2016.2.11.

Paper presents an enhanced model for calculation of reliability indices for different wind power plants configuration concepts used over past two decades. The autoregressive – moving average (ARMA) model is used combined with the sequential Monte Carlo simulation in order to predict expected energy not served (EENS) more accurately during the failure. Statistical database of LWK (Land Wirtschafts Kammer) is used for determining different wind power plant configuration types component reliability (performance) used for calculating influence of individual wind power plant configuration concepts on expected energy not served. Furthermore, a comparison of the distribution of EENS of different wind power plants configuration concepts have been presented, as well as the influence of the predominantly mechanical and electrical components failures on both EENS and failure rates.

TOPIĆ D, ŠLJIVAC D, STOJKOV M. **Model niezawodności różnych konfiguracji zestawu elektrowni wiatrowej oparty na sekwencyjnej symulacji Monte Carlo**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 237–244, http://dx.doi.org/10.17531/ein.2016.2.11.

W pracy przedstawiono udoskonalony model służący do obliczania wskaźników niezawodności dla różnych koncepcji konfiguracji zestawów elektrowni wiatrowych jakie stosowano w ostatnich dwóch dziesięcioleciach. Wykorzystano autoregresyjny model średniej ruchomej (ARMA), który w połączeniu z symulacją sekwencyjną Monte Carlo pozwala z większą dokładnością przewidzieć oczekiwaną wartość energii niedostarczonej (EENS) podczas awarii. Baza statystyczna LWK (Land Wirtschafts Kammer) posłużyła autorom do określania niezawodności (wydajności) części składowych elektrowni wiatrowych przy różnych typach konfiguracji zestawu. Otrzymane wartości wykorzystano do obliczenia wpływu poszczególnych koncepcji konfiguracji zestawu elektrowni wiatrowej na oczekiwaną wartość energii niedostarczonej. Ponadto, przedstawiono porównanie rozkładu EENS dla różnych koncepcji konfiguracji zestawu elektrowni wiatrowej jak również omówiono wpływ uszkodzeń części mechanicznych i elektrycznych elektrowni na EENS oraz awaryjność.

QIAN X, WU Y. **An electricity price-dependent control-limit policy for condition-based maintenance optimization for power generating unit**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 245–253, http://dx.doi.org/10.17531/ein.2016.2.12.

For the control-limit policy of condition-based maintenance (CBM), it usually focuses on the internal condition of the equipment while neglecting the un-constant external conditions. However, the electricity price-dependent downtime cost have influence on the cost-effectiveness of control-limit policy for a generating unit in a power system. To make a linkage between CBM and the non-constant cost model, an electricity price-dependent control-limit policy (EPCLP) is proposed to accommodate the time-dependent downtime costs. For the proposed EPCLP, preventive maintenance control-limits is much flexible to be adjusted to different electricity price levels, and the maintenance cost reduction can be achieved among the planning horizon as a result. The optimal control-limits and maintenance costs for different downtime-cost ratios, reliabilities, covariate processes and electricity price scenarios are analysed

QIAN X, WU Y. **Zastosowanie strategii uzależniającej termin przeglądu od ceny prądu elektrycznego do optymalizacji utrzymania ruchu agregatu prądotwórczego z uwzględnieniem jego stanu technicznego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 245–253, http://dx.doi.org/10.17531/ein.2016.2.12.

Stosując strategie utrzymania ruchu uwzględniające bieżący stan techniczny obiektu (condition based maintenance, CBM) oparte na pojęciu progu konserwacji koniecznej (control limit), najczęściej przywiązuje się wagę do stanu samego sprzętu, ignorując przy tym niestałe warunki zewnętrzne. Należy jednak pamiętać, że w przypadku agregatów prądotwórczych wchodzących w skład układów elektroenergetycznych, koszty przestoju zależne od ceny energii elektrycznej mają wpływ na opłacalność stosowania strategii progu konserwacji koniecznej. Aby powiązać CBM z modelem kosztów niestałych, zaproponowano strategię progu konserwacji koniecznej, w której wysokość progu uzależniona jest od ceny prądu elektrycznego (electricity price-dependent control-limit policy, EPCLP). Przyjęcie takiej strategii pozwala uwzględnić koszty przestojów zależne od czasu. W

to compare the performances between the proposed policy and the constant control-limit policy. Through the sensitivity analysis, the application scope of the proposed policy is evaluated.

EPCLP, progi czasowe konserwacji zapobiegawczej są bardzo elastyczne, co pozwala na ich regulację zgodnie z aktualną ceną energii elektrycznej. Strategia umożliwia redukcję kosztów w danym horyzoncie planowania. W celu porównania proponowanej strategii ze strategią stałego progu konserwacji koniecznej, w pracy przeanalizowano optymalne progi czasowe konserwacji koniecznej oraz koszty utrzymania ruchu dla różnych stosunków przestoju do kosztu, różnych wartości niezawodności, różnych procesów kowariantnych oraz różnych scenariuszy zmian cen energii elektrycznej. Zakres zastosowania proponowanej strategii oceniano za pomocą analizy czułości.

TCHÓRZEWSKA-CIEŚLAK B, PIETRUCHA-URBANIK K, URBANIK M. **Analysis of the gas network failure and failure prediction using the Monte Carlo simulation method**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 254–259, http://dx.doi.org/10.17531/ein.2016.2.13.

The scope of the article includes the analysis of the gas network failure based on a material obtained from field tests covering the years 2004-2014, conducted on the gas network of 120 thousand city, allowing to specify the failure rate of the gas network with division into material, pressure and pipelines diameter and indicate the main causes of failure on gas networks. On the base of the results of this analysis the Monte Carlo method to predict failures in gas pipe network has been presented.

TCHÓRZEWSKA-CIEŚLAK B, PIETRUCHA-URBANIK K, URBANIK M. **Analiza awaryjności sieci gazowych oraz prognozowanie awarii z zastosowaniem symulacyjnej metody Monte Carlo**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 254–259, http://dx.doi.org/10.17531/ein.2016.2.13.

Artykuł swoim zakresem obejmuje analizę awaryjności sieci gazowej na podstawie uzyskanego materiału z badań eksploatacyjnych obejmujących lata 2004-2014 prowadzonych na terenie Zakładu Gazowniczego w 120 tys. mieście, co pozwoliło na podanie intensywności uszkodzeń sieci gazowych z podziałem na materiał, ciśnienie i średnice rurociągów oraz podanie głównych przyczyn powstawania awarii na sieciach gazowych. Na podstawie wyników analizy zaprezentowano zastosowanie metody Monte Carlo do prognozowania awarii sieci gazowych.

LIU H, JIANG W, HULIO Z, WANG Q. **Estimation of system reliability by using the PLS-regression based corrected response surface method**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 260–270, http://dx.doi.org/10.17531/ein.2016.2.14.

A new computational method, referred as PLS-regression (PLSR) based corrected response surface method, has been developed for predicting the reliability of structural and mechanical systems subjecting to random loads, material properties, and geometry. The method involves a Corrected-Response Surface Model (C-RSM) based on the Partial Least Squares Regression Method (PLSRM) combined with some correction factors, and Monte Carlo Simulation (MCS), which is named as the Corrected-Partial Least Squares Regression-Response Surface Method (C-PLSRRSM). In order to develop an accurate surrogate model for the region determining the reliability of the system, a proper coefficient is presented to determine the sampling region of the input random variables. Due to a small number of original function evaluations, the proposed method is effective, particularly when a response evaluation entails costly finite-element, mesh-free, or other numerical analysis. Three numerical examples involving reliability problems of two structural systems and a mechanical system illustrate the method developed. Results indicate that the proposed method provides accurate and computationally efficient estimates of reliability. The proposed correction method, the PLSR based corrected response surface (C-PLSR-RS), can be the accurate surrogate model for calculating system reliabilities, especially for the implicit performance functions.

LIU H, JIANG W, HULIO Z, WANG Q. **Ocena niezawodności systemu z wykorzystaniem poprawionej metody powierzchni odpowiedzi opartej na regresji cząstkowych najmniejszych kwadratów**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 260–270, http://dx.doi.org/10.17531/ein.2016.2.14.

Nowa metoda obliczeniowa o nazwie "poprawiona metoda powierzchni odpowiedzi oparta na regresji PLS" (C-PLSRRSM) została opracowana dla potrzeb przewidywania niezawodności systemów konstrukcyjnych i mechanicznych poddanych obciążeniom losowym oraz charakteryzujących się losową geometrią oraz losowymi właściwościami materiałowymi. W metodzie uwzględniono pewne czynniki korekcyjne oraz symulację Monte Carlo. W celu opracowania odpowiedniego modelu zastępczego dla regionu stanowiącego o niezawodności systemu, przedstawiono współczynnik, który pozwala określić obszar pobierania próbek wejściowych zmiennych losowych. Ze względu na niewielką liczbę ocen funkcji początkowych, proponowana metoda jest skuteczna zwłaszcza wtedy, gdy ocena odpowiedzi wymaga kosztownej analizy numerycznej metodą elementów skończonych czy metodą automatycznie generowanej siatki (free mesh). Opracowaną metodę zilustrowano za pomocą trzech przykładów numerycznych dotyczących niezawodności dwóch systemów konstrukcyjnych oraz jednego układu mechanicznego. Wyniki wskazują, że proponowana metoda zapewnia dokładne i wydajne obliczeniowo oszacowanie niezawodności. Proponowana metoda C-PLSR-RS może stanowić trafny model zastępczy do obliczania niezawodności systemu, zwłaszcza w przypadku uwikłanych funkcji stanu granicznego.

JAWORSKI J, KLUZ R, TRZEPIECIŃSKI T. **Operational tests of wear dynamics of drills made of low-alloy high-speed HS2-5-1 steel**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 271–277, http://dx.doi.org/10.17531/ein.2016.2.15.

To determine the effect of drill wear on the value of the axial force and cutting torque, a series of durability tests of drills with a diameter of 10 mm made of high-speed steel HS2-5-1 were carried out. The investigations were conducted during the durability period and at constant values of cutting parameters. The tests were carried out while drilling holes in samples made of C45 steel and EN-GJS-500-7 cast iron. The dynamics of wear on all parts of the drill was also determined. It has been shown that while drilling with different values of cutting parameters, there is a loss of machinability for different values of the wear indicators. While drilling with high cutting speeds and with small feeds, there is a loss of cutting ability in the area of accelerated wear. The application of TiN coating does not change the controlled wear locations. TiN coating only reduces the intensity of wear on the tool flank, which increases the durability of the drill.

JAWORSKI J, KLUZ R, TRZEPIECIŃSKI T. **Badania eksploatacyjne dynamiki zużycia wierteł z niskostopowej stali szybkotnącej HS2-5-1**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 271–277, http://dx.doi.org/10.17531/ein.2016.2.15.

W celu określenia wpływu zużycia wierteł na wartość siły osiowej oraz momentu skrawania przeprowadzono serię badań trwałościowych wierteł o średnicy 10 mm wykonanych ze stali szybkotnącej HS2-5-1, przy stałych parametrach skrawania, w czasie jednego przyjętego okresu trwałości. Badania prowadzono podczas obróbki otworów na próbkach ze stali C45 oraz z żeliwa EN-GJS-500-7. Określono również dynamikę zużycia na wszystkich częściach skrawających wiertła. Wykazano, że podczas eksploatacji wierteł z różnymi parametrami skrawania, utrata ich skrawności następuje dla różnych wartości wskaźników zużycia. Podczas wiercenia z dużymi prędkościami skrawania i małymi posuwami, utrata skrawności wiertła następuje w obszarze przyspieszonego ich zużycia. Naniesienie powłoki TiN nie zmienia kontrolowanych miejsc zużycia a tylko zmniejsza intensywność zużycia na powierzchni przyłożenia, co powoduje wzrost trwałości wiertła.

IWANEK M, KOWALSKA B, HAWRYLUK E, KONDRACIUK K. **Distance and time of water effluence on soil surface after failure of buried water pipe. Laboratory investigations and statistical analysis**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 278–284, http://dx.doi.org/10.17531/ein.2016.2.16.

One solution to limit the inconvenience caused by suffosion processes following pipe breakages is retaining so-called protection zones near the pipes, the utilization of which would be handled by the system operator. Due to the fact that to determine the size of such zones is a challenging task, the analysis should be performed gradually, based on successive field studies, laboratory and numerical research. The present article is the outcome of the first stage of the laboratory research eventually aiming at the determination of the protection zone around a potential leakage in a water supply pipe. The first stage of the investigations was devoted to (1) the assessment of an average

IWANEK M, KOWALSKA B, HAWRYLUK E, KONDRACIUK K. **Odległość i czas wypływu wody na powierzchnię terenu po awarii podziemnego wodociągu. Badania laboratoryjne i analizy statystyczne**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 278–284, http://dx.doi.org/10.17531/ein.2016.2.16.

Jedną z propozycji ograniczenia uciążliwości spowodowanych zjawiskami sufozyjnymi po awarii wodociągu jest zachowanie w pobliżu przewodów tzw. stref ochronnych, o zagospodarowaniu których decydowałby eksploatator sieci. Określenie wymiarów takich stref jest bardzo trudnym zadaniem, dlatego stosowne analizy powinny odbywać się stopniowo, na bazie kolejnych etapów badań terenowych, laboratoryjnych i numerycznych. W ramach niniejszej pracy przedstawiono wyniki pierwszego etapu badań laboratoryjnych, których ostatecznym celem jest wyznaczenie strefy ochronnej wokół ewentualnej nieszczelności rury wodociągowej. Pierwszy etap badań objął określenie przeciętnej odległości wypływu

distance between the place of water effluence on the soil surface and the place of the water failure for 4 different areas of leak and 11 values of hydraulic pressure head in the pipe, (2) the initiatory assessment of the protection zone dimensions for analysed soil conditions, (3) the analysis of dependence between the time of water effluence on the soil surface after a failure of a buried water pipe and the leak area as well as the hydraulic pressure head in the pipe. The scope of the works comprises laboratory study and statistical analysis. The research was carried out preserving geometrical and kinematic similarity. The obtained results should be considered initial, oriented towards further stages of laboratory research comprising dynamic similarity.

wody na powierzchnię terenu od miejsca awarii podziemnego wodociągu dla 4 różnych powierzchni nieszczelności przewodu oraz 11 wysokości ciśnień w przewodzie, wstępne oszacowanie wielkości strefy ochronnej dla analizowanych warunków gruntowych oraz analizę zależności między czasem wypływu wody na powierzchnię terenu po awarii podziemnego wodociągu a powierzchnią nieszczelności i wysokością ciśnienia w przewodzie. Zakres pracy obejmował badania laboratoryjne i analizy statystyczne. Badania przeprowadzono z zachowaniem podobieństwa geometrycznego i kinematycznego. Uzyskane wyniki należy więc traktować jako wstępne, ukierunkowujące dalsze etapy badań laboratoryjnych, uwzględniające również podobieństwo dynamiczne.

KOSOBUDZKI M, STAŃCO M. **The experimental identification of torsional angle on a load-carrying truck frame during static and dynamic tests**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 285–290, http://dx.doi.org/10.17531/ein.2016.2.17.

The underframe of a truck is one of the most loaded parts of a vehicle. It is a spatial unit and it must be strong enough to withstand random loading within many years of maintenance. The most severe form of deformation is in torsion. So, frame side members are often made from elements with channel sections, rigid for bending and flexible for torsion. Authors have conducted the research of 6x6 high mobility wheeled vehicle assigned to 20-feet container. Their load-carrying structure is made from two separate underframes: longitudinal and auxiliary connected with bolted joints. The goal of the research was to check if the torsional angle of deformation of the underframe during static and dynamic tests is within an acceptable range. The static test was carried out for the main underframe first to assess the characteristic of torsional stiffness without the auxiliary frame. After connecting both frames together the measure was conducted again. In the experiment the diagonal wheels were lifted up and the resulting displacement of the ends of the frame side members was recorded. Simultaneously the strain at chosen points of the underframe was measured with a system of turned half bridge strain gauges. After calibrating the measuring system a second part of experiment was conducted within proving ground tests when the vehicle was fully loaded. The collected strain data at chosen points allowed for calculating the resultant displacement of the ends of the frame side members in function of sort of road and to indicate the influence of auxiliary frame on increasing the torsional stiffness of the underframe.

KOSOBUDZKI M, STAŃCO M. **Identyfikacja eksperymentalna kąta skręcenia ustroju nośnego pojazdu podczas testu statycznego i dynamicznego**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 285–290, http://dx.doi.org/10.17531/ein.2016.2.17.

Ustrój nośny pojazdu jest jednym z jego najbardziej obciążonych zespołów konstrukcyjnych. Jest to zespół o złożonej budowie przestrzennej, który musi być wystarczająco wytrzymały by wytrzymać zmienne obciążenia przez wiele lat eksploatacji pojazdu. Najbardziej obciążające są te obciążenia które wywołują skręcanie ustroju nośnego. Stąd ustrój nośny składa się najczęściej z podłużnic połączonych poprzeczami co w efekcie zapewnia dużą sztywność na zginanie i podatność na skręcanie. W artykule przedstawiono badania podwozia pojazdu kołowego wysokiej mobilności 6x6 przeznaczonego do połączenia z kontenerem 20-stopowym. Ustrój nośny pojazdu składa się z ramy głównej połączonej za pomocą połączeń podatnych z ramą pośrednią. Celem badań było sprawdzenie czy kąt skręcenia ustroju nośnego pojazdu w badaniach statycznych i dynamicznych nie wywołuje naprężeń wykraczających poza zakres dopuszczalny. Test statyczny został przeprowadzony najpierw tylko do ramy głównej w celu wyznaczenia jej sztywności skrętnej. Następnie ramy zostały połączone i wyznaczenie sztywności zostało powtórzone. W ramach testu koła znajdujące się w pojeździe po przekątnej zostały podniesione aż do utraty kontaktu z podłożem. Równocześnie rejestrowano przemieszczenie końców podłużnic ramy i odkształcenia w wybranych punktach, w których naklejono tensometry. Po skalibrowaniu układu pomiarowego przeprowadzono szereg testów przebiegowych z pojazdem całkowicie obciążonym ładunkiem. Zarejestrowane wartości odkształceń wykorzystano do wyznaczenia odkształcenia wypadkowego końców podłużnic ramy w funkcji rodzaju drogi oraz wpływu zamocowania kontenera na wypadkową sztywność skrętną ustroju pojazdu.

WANG H, DENG G, LI Q, KANG Q. **Research on bispectrum analysis of secondary feature for vehicle exterior noise based on nonnegative tucker3 decomposition**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 291–298, http://dx.doi.org/10.17531/ein.2016.2.18.

Nowadays, analysis of vehicle exterior noise has become more and more difficult for NVH (noise vibration and harshness) engineer to find out the fault among the exhaust system when some significant features are masked by the jamming signals, especially in the case of the vibration noise associating to the bodywork. New method is necessary to be explored and applied to decompose a high-order tensor and extract the useful features (also known as secondary features in this paper). Nonnegative Tucker3 decomposition (NTD) is proposed and applied into secondary feature extraction for its high efficiency of decomposition and well property of physical architecture, which serves as fault diagnosis of exhaust system for an automobile car. Furthermore, updating algorithm conjugating with Newton-Gaussian gradient decent is utilized to solve the problem of overfitting, which occurs abnormally on traditional iterative method of NTD. Extensive experimen results show the bispectrum of secondary features can not only exceedingly interpret the state of vehicle exterior noise, but also be benefit to observe the abnormal frequency of some important features masked before. Meanwhile, the overwhelming performance of NTD algorithm is verified more effective under the same condition, comparing with other traditional methods both at the deviation of successive relative error and the computation time.

WANG H, DENG G, LI Q, KANG Q. **Badania nad analizą bispektrum cech drugorzędnych hałasu zewnętrznego pojazdów w oparciu o nieujemną dekompozycję Tuckera3**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 291–298, http://dx.doi.org/10.17531/ein.2016.2.18.

Obecnie inżynierowie NVH (zajmujący się problematyką hałasu, drgań i uciążliwości akustycznych) napotykają na coraz większe trudności przy analizie hałasu zewnętrznego pojazdów wynikające z faktu, że istotne cechy związane z nieprawidłowościami układu wydechowego są maskowane przez sygnały zakłócające, szczególnie hałas wibracyjny związany z pracą nadwozia. Niezbędna jest zatem nowa metoda, która pozwoli rozkładać tensory wysokiego rzędu i wyodrębniać przydatne cechy (zwane w tym artykule także cechami drugorzędnymi). Do ekstrakcji cech drugorzędnych wykorzystano w prezentowanej pracy metodę nieujemnej faktoryzacji tensorów znaną także jako nieujemna dekompozycja Tuckera 3 (NTD) , która cechuje się wysoką efektywnością dekompozycji i może być wykorzystywana w diagnostyce uszkodzeń układu wydechowego samochodów. Problem nadmiernego dopasowania, który występuje w tradycyjnej metodzie iteracyjnej NTD rozwiązano przy pomocy algorytmu aktualizacyjnego sprzężonego z gradientem prostym Newtona-Gaussa. Wyniki doświadczeń pokazują, że bispektrum cech drugorzędnych nie tylko pozwala doskonale interpretować stan hałasu zewnętrznego pojazdu, ale również umożliwia wykrywanie wcześniej maskowanych nieprawidłowych częstotliwości odpowiadających niektórym ważnym cechom. Badania potwierdzają, że algorytmu NTD jest bardziej efektywny, w tych samych warunkach, w porównaniu z innymi tradycyjnymi metodami zarówno w zakresie odchyleń błędu względnego jak i czasu obliczeń.

ZUBER N, BAJRIĆ R. **Application of artificial neural networks and principal component analysis on vibration signals for automated fault classification of roller element bearings**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 299–306, http://dx.doi.org/10.17531/ein.2016.2.19.

The article addresses the implementation of feature based artificial neural networks and vibration analysis for automated roller element bearings faults identification purpose. Vibration features used as inputs for supervised artificial neural networks were chosen based on principal component analysis as one of the possible methods of data dimension reduction. Experimental work has been conducted on a specially designed test rig and on a drive of the Ganz port crane in port of Novi Sad, Serbia. Different scalar vibration features derived from time and frequency domain were used as inputs to fault classifiers. Several types of roller elements bearings faults, at different levels of loads were tested: discrete faults on inner and outer race and

ZUBER N, BAJRIĆ R. **Zastosowanie sztucznych sieci neuronowych oraz analiz głównych składowych sygnału drgań do automatycznej klasyfikacji uszkodzeń łożysk tocznych**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 299–306, http://dx.doi.org/10.17531/ein.2016.2.19.

Artykuł omawia zastosowanie sztucznych sieci neuronowych opartych na cechach oraz analizy drgań do celów automatycznej identyfikacji uszkodzeń łożysk tocznych. Cechy drgań mające posłużyć jako dane wejściowe do nadzorowanych sztucznych sieci neuronowych wybrano na podstawie analizy głównych składowych, która stanowi jedną z metod zmniejszania rozmiaru zbioru danych statystycznych. Badania prowadzono na specjalnie do tego celu zaprojektowanym stanowisku badawczym oraz na układzie napędu żurawia portowego firmy Ganz w porcie Novi Sad w Serbii. Jako wejścia klasyfikatorów uszkodzeń wykorzystano różne skalarne cechy drgań określone w dziedzinie czasu i częstotliwości. Badano kilka typów uszkodzeń łożysk tocznych przy różnych poziomach obciążenia: uszkodzenia dyskretne w obrębie pierścienia wewnętrznego i zewnętrznego łożyska

looseness. It is demonstrated that proposed set of input features enables reliable roller element bearing fault identification and better performance of applied artificial neural networks.

PAN Z, JING F, SUN Q. **Lifetime distribution and associated inference of systems with multiple degradation measurements based on gamma processes**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 307–313, http://dx.doi.org/10.17531/ein.2016.2.20.

With development of science and technology, many engineering systems take on high reliable characteristic and usually have complex structure and failure mechanisms, with their reliability being evaluated by multiple degradation measurements. In certain physical situations, the degradation of these performance characteristics would be always positive and strictly increasing. Therefore, the gamma process is usually considered as a degradation process due to its independent and non-negative increments properties. In this paper, we suppose that a system has multiple dependent performance characteristics and that their degradation can be modeled by gamma processes. For such a multivariate degradation involving three or more performance characteristics, we propose to use a multivariate Birnbaum-Saunders distribution and its marginal distributions to approximate the reliability function and give the corresponding lifetime distribution. And then, the inferential method for the model parameters is developed. Finally, for an illustration of the proposed model and method, a simulated example is discussed and some computational results are presented.

oraz nadmierny luz. Wykazano, że proponowany zbiór cech wejściowych umożliwia niezawodną identyfikację uszkodzeń łożysk tocznych oraz zapewnia lepszą wydajność zastosowanych sztucznych sieci neuronowych.

PAN Z, JING F, SUN Q. **Wyznaczanie rozkładu czasów życia oraz wnioskowanie dla systemów wymagających pomiarów współistniejących degradacji w oparciu o procesy gamma**. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 307–313, http://dx.doi.org/10.17531/ein.2016.2.20.

Wraz z rozwojem nauki i techniki, powstaje coraz więcej systemów inżynieryjnych o wysokich parametrach niezawodnościowych, które zwykle charakteryzują się złożoną strukturą i złożonymi mechanizmami uszkodzeń. Ocena niezawodności w przypadku takich systemów wymaga pomiarów współwystępujących procesów degradacji . W pewnych sytuacjach fizycznych, degradacja właściwości użytkowych systemu będzie zawsze dodatnia oraz ściśle rosnąca. Proces degradacji jest zwykle procesem gamma, który charakteryzują niezależne i nieujemne przyrosty. W niniejszej pracy, założono, że system ma wiele zależnych charakterystyk pracy oraz że ich degradację można modelować procesem gamma. W przypadkach takiej wielowymiarowej degradacji obejmującej trzy lub więcej charakterystyk pracy zaproponowano zastosowanie rozkładu Birnbauma-Saundersa (uwzględniającego wiele zmiennych) oraz jego rozkładów brzegowych do aproksymacji funkcji niezawodności oraz określania odpowiadającego jej rozkładu czasu pracy. Opracowano metodę wnioskowania dla parametrów modelu. Wreszcie, dla zilustrowania proponowanego modelu oraz metody, omówiono przykład symulacyjny oraz przedstawiono niektóre wyniki obliczeniowe.

Article citation info:
BOŠNJAK S, ARSIĆ M, SAVIĆEVIĆ S, MILOJEVIĆ G, ARSIĆ D. Fracture analysis of the pulley of a bucket wheel boom hoist system. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 155–163, http://dx.doi.org/10.17531/ein.2016.2.1.

Srđan BOŠNJAK
Miodrag ARSIĆ
Sreten SAVIĆEVIĆ
Goran MILOJEVIĆ
Dušan ARSIĆ

FRACTURE ANALYSIS OF THE PULLEY OF A BUCKET WHEEL BOOM HOIST SYSTEM

ANALIZA PĘKNIĘĆ KOŁA PASOWEGO UKŁADU WCIĄGARKI WYSIĘGNIKA KOŁA CZERPAKOWEGO

This paper presents the results of the pulley fracture analysis. Experimental investigations confirmed that the chemical composition and basic mechanical properties of the pulley material, except the impact energy at a temperature of –20°C, meet the requirements of the corresponding standard. The impact energy value at the temperature of –20°C is for ≈45% lower than the prescribed value which has considerable influence on the appearance of the brittle fracture, especially having in mind the fact that the bucket wheel excavators operate at low temperatures. Metallographic examinations as well as magnetic particle inspections indicated that initial cracks in the welded joints occurred during the manufacture of the pulleys. Characteristic levels of the rope load cycle are obtained by using in-house software which includes the dynamic effects of the resistance-to-excavation. The FEA results pointed out that in the representative load cases the combinations of the mean stress and the alternating stress in the pulley critical zone lie considerably below the limit line of the modified Goodman's diagram. The conclusion, based on the presented results, is that the fracture of the pulley appeared as the result of the 'manufacturing-in' defects.

Keywords: *bucket wheel excavator, pulley fracture, experimental investigations, FE stress analyses.*

Artykuł przedstawia wyniki analizy pęknięć koła pasowego. Badania doświadczalne potwierdziły, że skład chemiczny oraz podstawowe właściwości mechaniczne materiału, z którego zostało wykonane koło pasowe, za wyjątkiem energii udaru w temperaturze –20°C, były zgodne z odpowiednią normą. Wartość energii udaru w temperaturze –20°C była o ≈45% niższa od wartości zalecanej, co ma znaczący wpływ na występowanie pękania kruchego, zwłaszcza gdy weźmie się pod uwagę fakt, że koparki kołowe są przeznaczone do pracy w niskich temperaturach. Badania metalograficzne oraz badania magnetyczno-proszkowe wykazały, że pęknięcie pierwotne w połączeniu spawanym pojawiło się już w fazie produkcji koła pasowego. Charakterystyczne poziomy cyklu obciążenia liny uzyskano stosując własne oprogramowanie, które uwzględnia dynamiczne oddziaływanie odporności na urabianie. Wyniki MES pokazały, że w przypadku obciążeń reprezentatywnych, wartości średniego naprężenia w funkcji naprężenia zmiennego w strefie krytycznej koła pasowego były znacznie niższe niż wartości graniczne wyznaczone na podstawie zmodyfikowanego wykresu Goodmana. Na podstawie otrzymanych wyników stwierdzono, że pęknięcie koła pasowego powstało wskutek wad produkcyjnych.

Słowa kluczowe: *koparka kołowa, pęknięcie koła pasowego, badania doświadczalne, analiza naprężeń metodą elementów skończonych.*

1. Introduction

The bucket wheel boom (BWB) of the bucket wheel excavator (BWE) SRs 1300 (Fig. 1) is hung by two stays hinged to the trolley with the pulley block (the so-called "moving pulley block"). Changing of the BWB inclination angle is realized by shifting the moving pulley block.

The BWB hoist system is the vital part of the BWE. Failures of its components may lead to catastrophic consequences as described in [1, 3, 34]. Even in cases where the direct failure effects are not so drastic, the indirect financial losses are high [9, 10, 13]. In-service fracture of one pulley of the fixed pulley block (Fig. 2) is a typical example of a failure in which the direct material loss (≈3,000 €) is far less than the indirect financial loss (714,000 €) caused by the system downtime during the execution of very complex operations such as: temporary

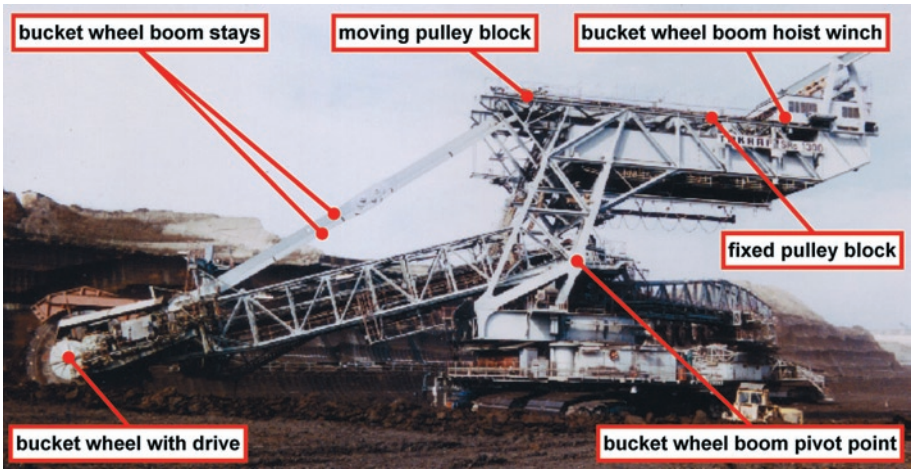


Fig. 1. BWE SRs 1300: total weight 2303 t; theoretical capacity 4500 m³/h

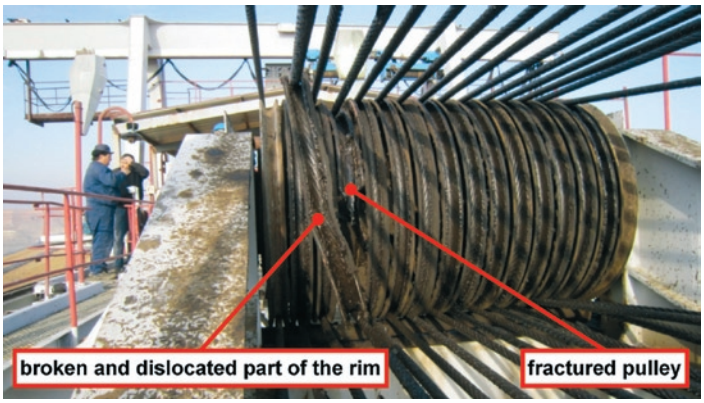


Fig. 2. Fracture of the pulley of the fixed pulley bloc

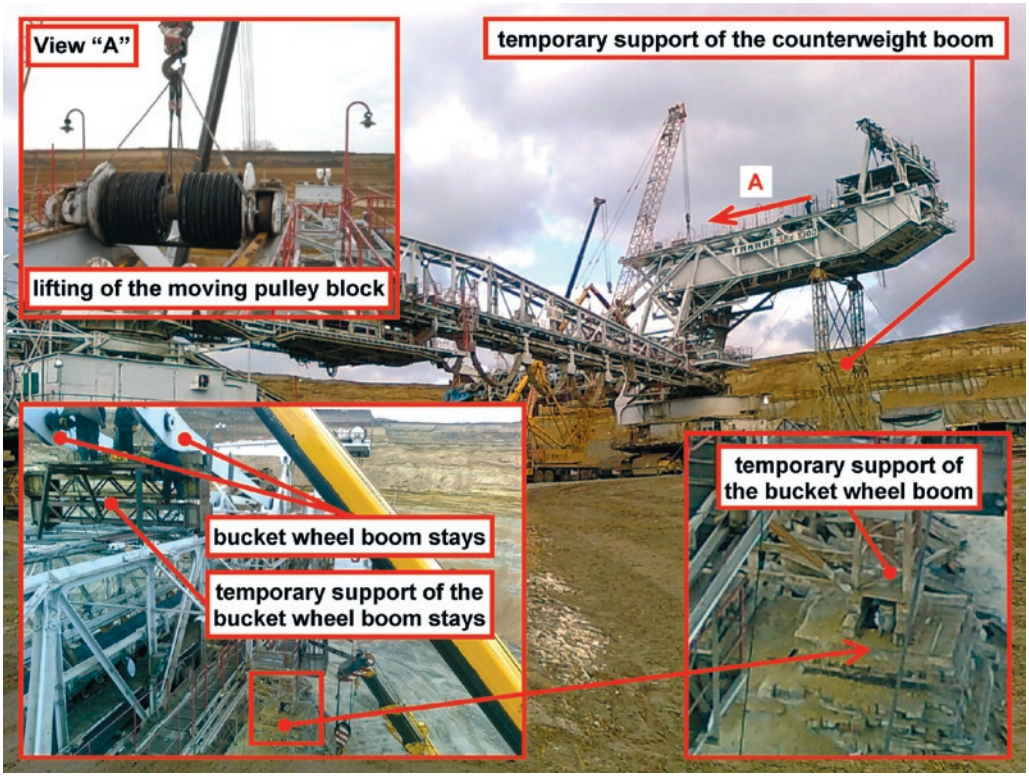


Fig. 3. Details of the BWE temporary supports and the dismantling of the BWB hoist system

supporting of the machine, dismantling of the BWB hoist system, testing and repair of the failed pulleys, testing of the rope and reassembling the BWB hoist system (Fig. 3, Table 1).

The goals of the study presented in the paper were to: (1) Develop a method of identifying pulley working loads, taking into account the dynamic nature of the external loads caused by the resistance-to-excavation; (2) Establish the procedure and determine the cause of pulley fracture; (3); To give the expert judgment: repair or redesign the pulleys.

The following sections will present details of the carried out experimental and numerical researches and the conclusions arrived at therein. The investigation results are important because: (a) pulleys are vital parts of the rope mechanisms; (b) same or similar problems could arise in rope hoisting mechanisms of not only various types of mining machines [40] but also of a wide class of construction machines and cranes. Besides, research results indicate the importance of the non-destructive testing (NDT) of welded joints of the BWE vital structural parts.

Table 1. Specification of costs due to the overburden system downtime caused by one pulley failure

Nomenclature	Cost in €
DT and NDT testing before, during and after the pulley repair	6,000
Engagement of workers and machines	36,000
14–days system downtime (14 x 24 h x 2,000 €/h)	672,000
Total	714,000

2. Fracture description

During BWE exploitation a failure of the welded joints of the spokes and rim occurred, which led to the plastic deformation and fracture of the rim (Figs 4a, b). Apart from that, plastic deformations of the spokes are observed (Fig. 4c) as well as fractures of their welded joints with the hub.

3. Experimental investigations

3.1. Destructive testing

According to the design documentation, the pulleys were supposed to be made from steel quality grade St 37-3 (according to the code [11]). Experimental examinations are performed on samples taken from the damaged pulley (Fig. 5). Results of the chemical analysis, tensile and impact tests are presented in Tables 2–4. Average macrohardness is 129 HB [17].

Metallographic examinations are carried out on the replicas [28]

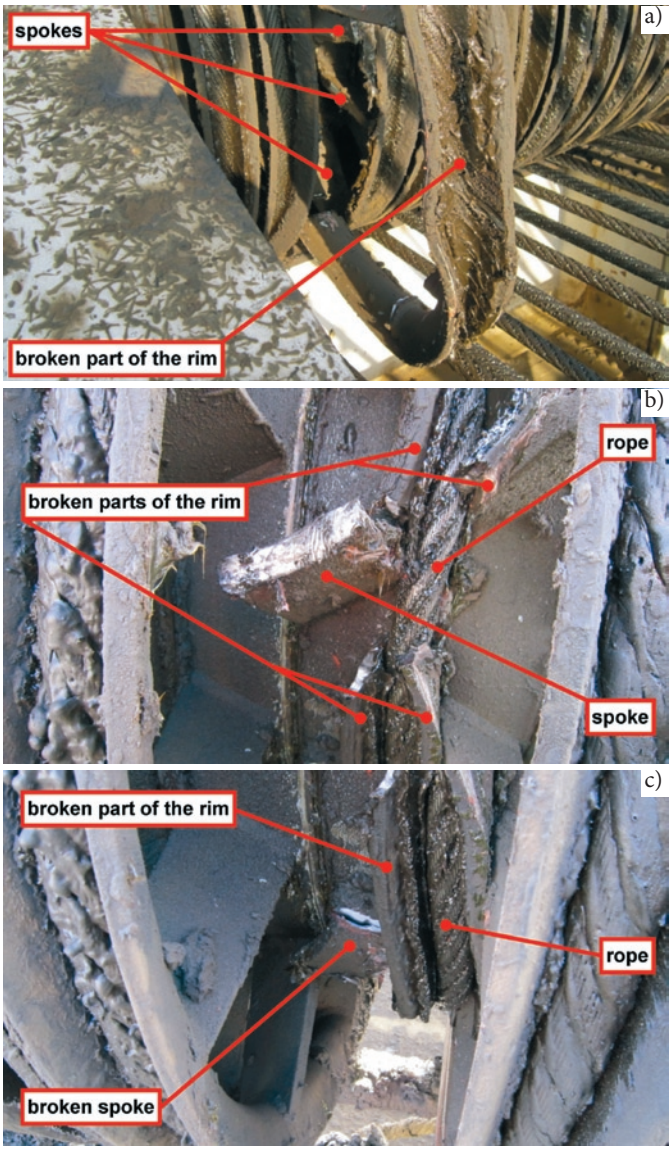


Fig. 4. Details of the fractured pulley: (a) front view; (b) back view; (c) view from below

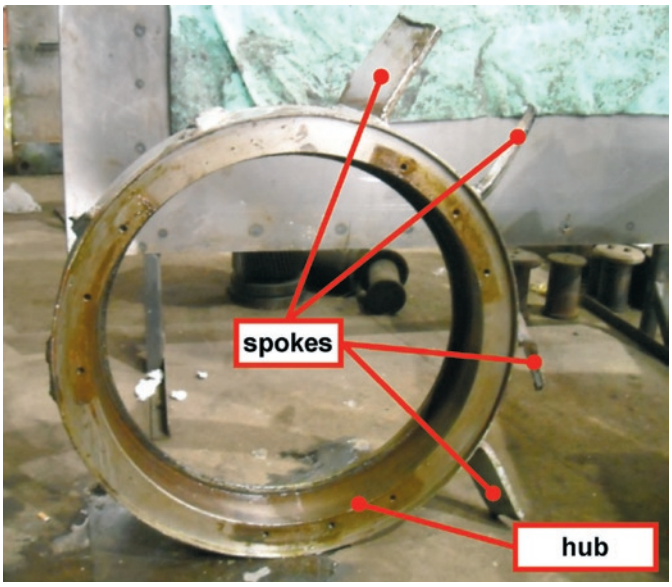


Fig. 5. A part of the damaged pulley used for sampling

Table 2. Chemical analysis (wt.%) of the pulley material and chemical composition of St 37-3 [11]

Material	C	Si	Mn	S	P	Al
Sample	0.159	0.188	0.625	0.018	0.016	0.004
St 37-3	max. 0.19	–	–	max. 0.050	max. 0.050	–

Table 3. Tension test [14] results of the pulley material and tensile properties of St 37-3 [11]

Specimen	σ_{YS} (MPa)	σ_{UTS} (MPa)	Elongation A_{50} (%)	Contraction Z (%)
1	278	434	44.5	46.9
2	283	433	38.2	44.4
3	281	435	40.5	44.4
St 37-3	min. 235	360–510	min. 24	–

Table 4. Impact energy test [15] results of the pulley material and impact energy of St 37-3 [11]

Temperature	Specimen	Impact energy $KV_{300/2}$ (J)	Average (J)
–20°C	1	14.7	14.7
	2	12.7	
	3	16.7	
	St 37-3	min. 27	–
0°C	4	39,2	40.5
	5	42.2	
	6	40.2	

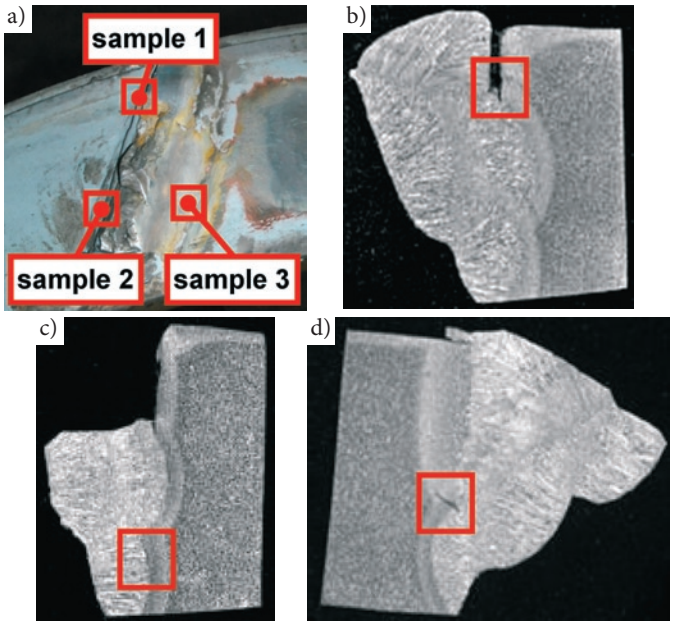


Fig. 6. Sampling zones (a) and samples: 1 – (b), 2 – (c), 3 – (d)

taken from the welded joint of the spoke and hub (Fig. 6). Grain sizes were determined using the standard [19], and the content of the non-metallic inclusions according to the code [20] (Figs. 7, 8).

3.2. Non-destructive testing

After dismantling the fixed and moving pulley blocks (Fig. 3) the magnetic particle inspection (MPI) of the fillet welds was carried out according to the code [16]. Crack indications (Fig. 9, Table 5) were

Fig. 5. A part of the damaged pulley used for sampling

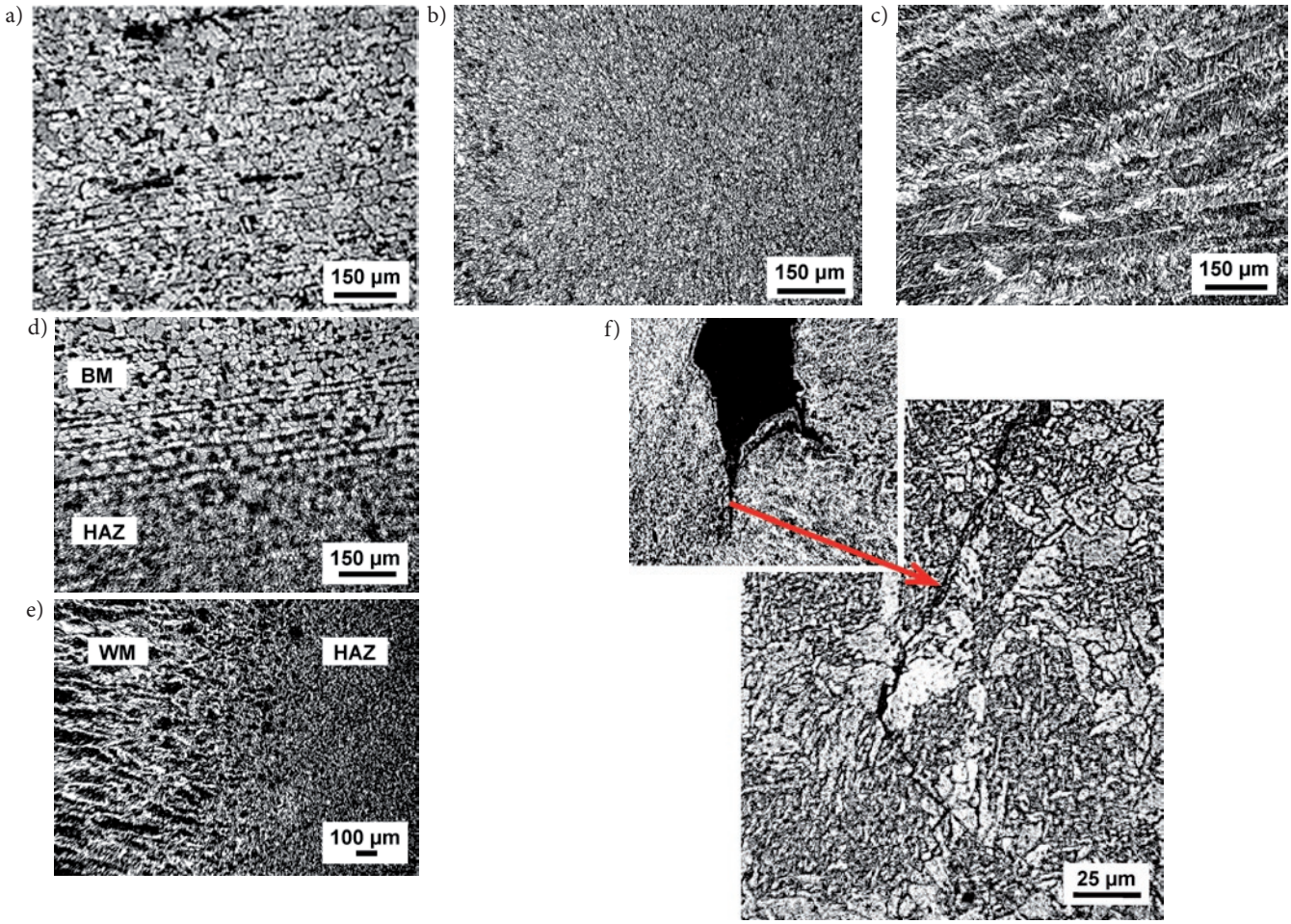


Fig. 7. Sample 1 (etched with 3% nital): (a) strip-type ferrite-pearlite microstructure of the BM, microhardness of 140–144 HV1 [27]; (b) fine-grained ferrite-pearlite microstructure in HAZ, microhardness of about 159 HV1; (c) casting dendritic microstructure of the WM, microhardness of 172–197 HV1; (d) microstructure on the verge of the BM and HAZ; (e) microstructure on the verge of the WM and HAZ; (f) tip of the crack (depth ≈4 mm, width ≈0.5 mm) and its propagation

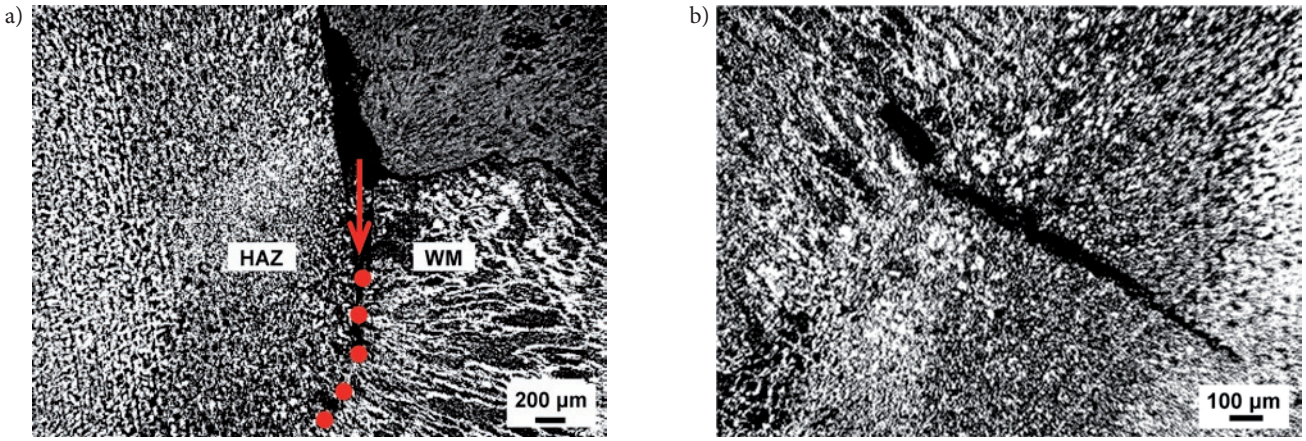


Fig. 8. (a) Crack propagation alongside the verge of the WM and HAZ, sample 2 (framed detail in Fig. 6c); macrocrack (total length of ≈1.5 mm) initiated in WM and its propagation through HAZ and BM, sample 3 (framed detail in Fig. 6d)

observed on all pulleys. They are considerably more pronounced on the welded joints of the spokes and rim than the welded joints of the spokes and hub.

4. Pulley stress analyses

The load analysis of the BWE structure is very complex due to its changeable geometry configuration (Fig. 10).

The forces in the rope of the BWB hoist system, shown in Fig. 11, are determined using standard [12]. The intensity of the cutting force is calculated based on the parameters of the BW drive, adopting [12] that the total cutting force is realized on one bucket only (Fig. 10). For load case (LC) H [12] the intensity of the cutting force (U_{nom}=298.5 kN) is calculated based on the nominal torque of the BW drive motor, whilst in LC HZ [12] its intensity (U_{max}=376.4 kN) is calculated based on the maximum torque of the clutch. In both LCs, rope force reaches its maximum for α_{BWB}=3°15' (Fig. 11). It can be

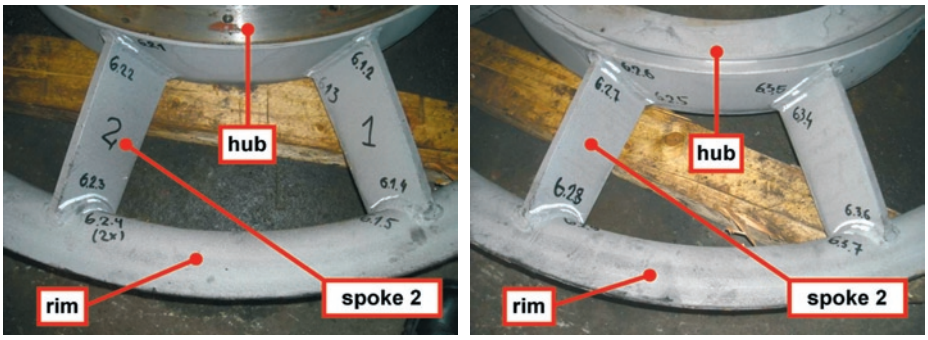


Fig. 9. Typical MPI indications – pulley 6, spoke 2: (a) left side; (b) right side

Table 5. The MPI crack indications on spoke 2 (pulley 6).

Designation (a.b.c*)	6.2.1	6.2.2	6.2.3	6.2.4	6.2.5	6.2.6	6.2.7	6.2.8	6.2.9
Figure	9a				9b				
Length (mm)	15	30	80	20;30	25	30	40	100	55

a = ordinal of a pulley; b=ordinal of a spoke; c=ordinal of an indication.

noted that usage of the procedure prescribed in [12] leads to the loss of one of the key properties of the BWEs' working process – the dynamic character of the external load caused by resistance-to-excavation [2, 4, 6–8, 22–24, 29, 30, 32, 33, 35–39]. By extracting the static influence of the cutting force from the curves shown in Fig. 11 and introducing its dynamic influence determined in the manner presented in [2, 4], a more realistic character of changing of the rope force during the excavation process is obtained (Fig. 12, Table 6).

The stress state analyses are done by applying the finite element method (FEM). The 3D model of the pulley and rope (Fig. 13) was discretized by 10-node tetrahedron elements in order to create the FEM model (317,066 nodes, 185,852 elements, Fig. 14). Calculations are carried out for the maximum value of the angle between the legs of the rope $\alpha_{R,max}$=5° (Fig. 13b). Interaction between the

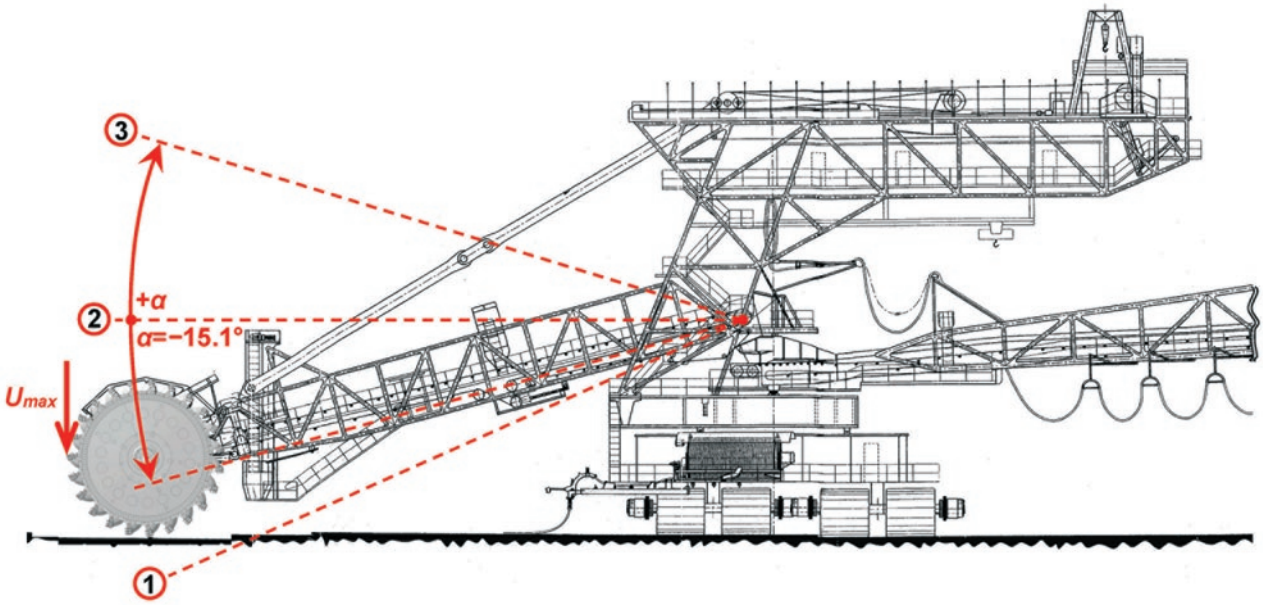


Fig. 10. Characteristic BWB positions: 1 – low, α_{BWB}=–24°2'22"; 2 – horizontal; 3 – high, α_{BWB}=16°34'48" (U–cutting force)

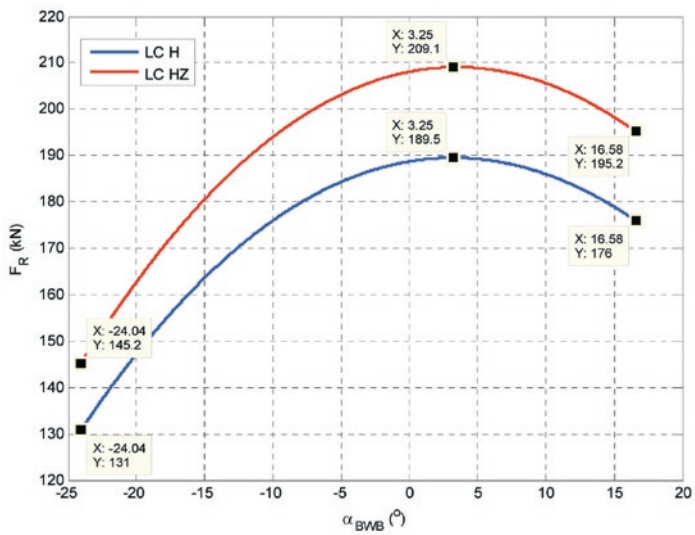


Fig. 11. The dependence of the rope force (F_R) on the BWB inclination angle (α_{BWB})

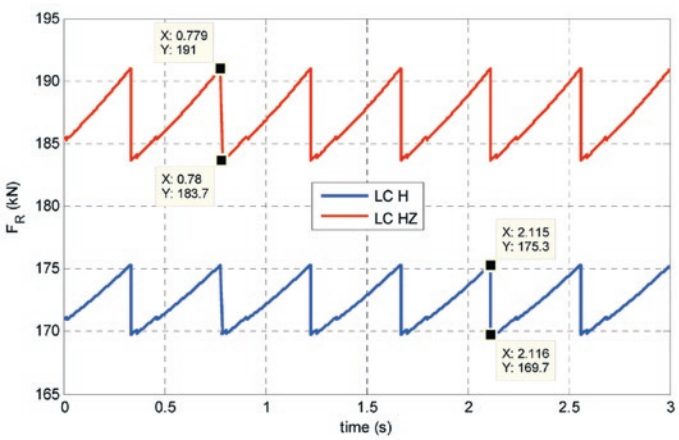


Fig. 12. Simulation of the rope force during the excavation process (α_{BWB}=3°15')

Table 6. Characteristic levels of the rope load cycle.

Load case	Rope force (kN)			
	maximum (F_{Rmax})	minimum (F_{Rmin})	mean (F_{Rm})	amplitude (F_{Ra})
H	175.3	169.7	172.5	2.8
HZ	191.0	183.7	187.35	3.65

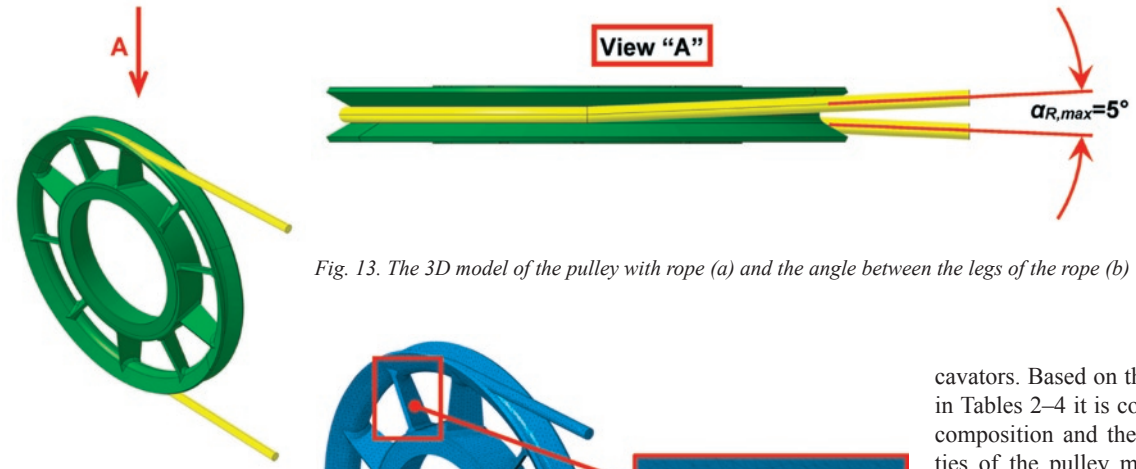


Fig. 13. The 3D model of the pulley with rope (a) and the angle between the legs of the rope (b)

rope and the pulley was simulated by contact connection. Maximum values of von Mises stresses are obtained in the T-fillet welded connections of the rim (Fig. 15). The characteristic stress levels of the pulley load cycle during excavation for both considered LCs are presented in Table 7. For the maximum rope force (209.1 kN, Fig. 11) obtained by using code [12], the maximum von Mises stress value in the T-fillet welded connection of the rim is $\sigma_{vM\,\max,[12]} = 186$ MPa.

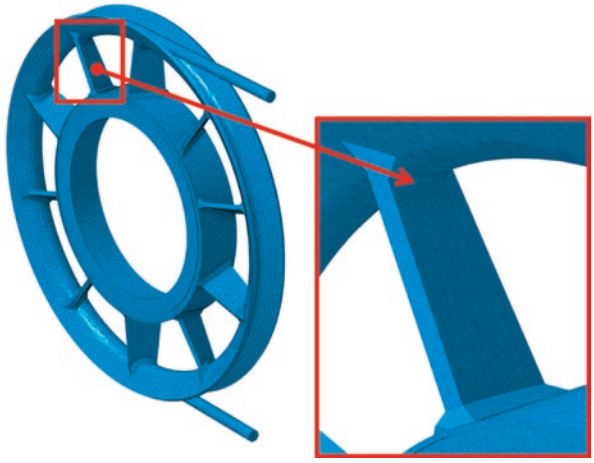


Fig. 14. The FEM model

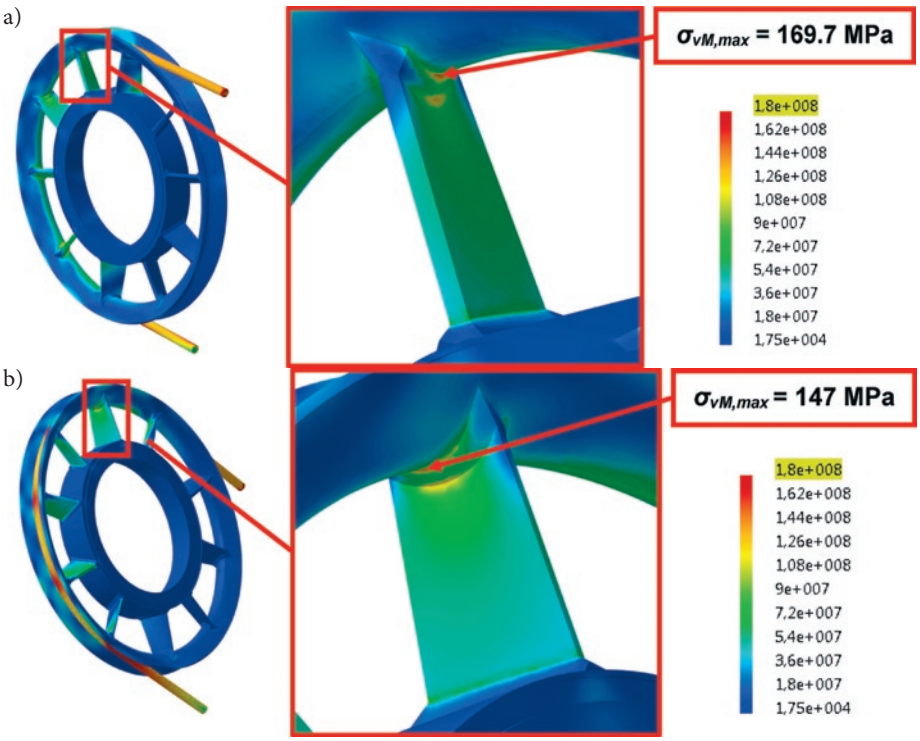


Fig. 15. The von Mises stress field obtained for F_R=191 kN (the maximum rope force in LC HZ, Fig. 12)

5. Discussion

In order to make a decision on whether to repair or redesign the pulleys, it was necessary to conduct a complex procedure whose basic stages are shown in Fig. 16.

In the considered case, both pulley design and the material were adequately selected – steel quality grade RSt 37-3 is commonly used for manufacturing the pulley blocks for the bucket wheel excavators. Based on the testing results presented in Tables 2–4 it is conclusive that the chemical composition and the basic mechanical properties of the pulley material, except the impact energy value at the temperature of –20°C, meet the requirements of standard [11] prescribed for steel grade RSt 37-3. Namely, the impact energy value at the temperature of –20°C is for ≈45% lower (Table 4) than the value listed in [11] which has considerable influence on the appearance of the brittle fracture, especially having in mind the fact that the BWEs operate at low temperatures.

Micrographic testing indicates the notably structural heterogeneity of the welds. BM has the strip-type ferrite-pearlite microstructure with non-metallic inclusions of both oxide and sulfide type (Fig. 7a). The microstructure in HAZ (Fig. 7b) is fine-grained ferrite-pearlite with fine-grained oxide type non-metallic inclusions. WM has the casting dendritic microstructure (Fig. 7c) because the appropriate heat treatment of the pulley welded structure was not carried out. Under fatigue loading the non-metallic inclusions in BM and HAZ may cause the appearance of initial cracks, whilst the dendritic microstructure of the WM indicates the tendency towards the brittle fracture. Poor manufacturing practice led to multiple welding defects – incomplete welding (Fig. 6b). Those defects significantly accelerate premature crack initiation by playing a role, from the welding point of view, as the local HAZ based weak link.

The considerably more pronounced presence of the MPI crack indications at welded joints of the spokes and rim is the consequence of their geometry being more complex than the geometry of the welded joints of the spokes and hub.

In accordance with the recommendations [25, 26] regarding fatigue safety evaluation, it is adopted that the fatigue limit of the critical welded joint is S_e=45 MPa. The tensile strength of the weld metal (σ_c) is determined by the following expression [31]:

Table 7. Characteristic levels of the pulley stress cycle.

Load case	Stress value (MPa)			
	maximum (σ_{max})	minimum (σ_{min})	mean (σ_m)	amplitude (σ_a)
H	155.8	150.8	153.3	2.5
HZ	169.7	163.3	166.5	3.2

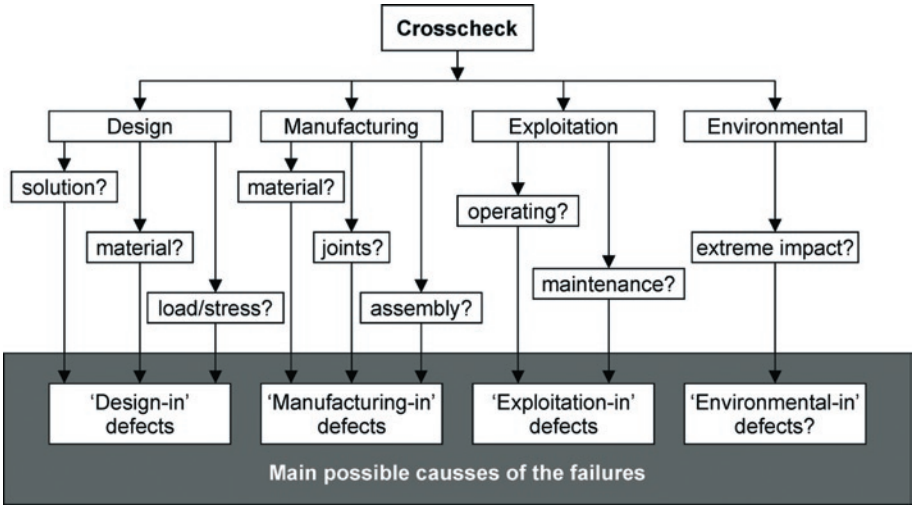


Fig. 16. The procedure of determining the causes of pulley fracture

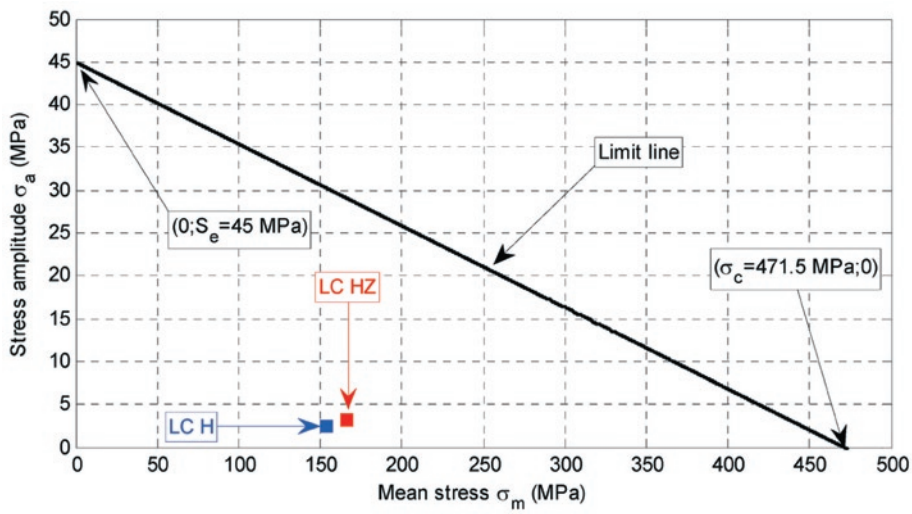


Fig. 17. The modified Goodman diagram

$$\sigma_c = 0.5\sigma_w + 0.5\sigma_{UTS\min} = 0.5\times 510 + 0.5\times 433 = 471.5 \text{ MPa}, \quad (1)$$

where σ_w=510 MPa is the ultimate tensile strength of deposited metal (electrode BÖHLER FOX EV 50 [18]), while σ_{UTSmin}=433 MPa is the minimum ultimate tensile strength value of base metal obtained by tensile testing (Table 3).

It is obvious (Fig. 17) that in both representative LCs the combinations of the mean stress and the alternating stress in the critical zone (Table 7) lie considerably below the limit line connecting the fatigue limit S_e and the tensile strength of the weld metal σ_c. Besides that, in the case of the maximum rope force obtained by using code [12], the maximum von Mises stress value ($\sigma_{vM\,\max,[12]}$) is 2.5 times lower than the tensile strength of the weld metal (σ_c, Eq. (1)).

Based on the presented results, it was concluded that the pulley fracture was caused by the 'manufacturing-in' defects. That is why it was decided to carry out the repairs of the pulley spokes' welded joints, without modifying their design.

6. Conclusion

Perennial exploitation of the BWEs in harsh working conditions leads to a gradual degradation of their subsystems. Despite rigorous controls during designing, manufacturing and assembling, compliant with the relevant standards, failure occurrence is almost inevitable during BWEs exploitation. Their causes could be of the different nature [5, 21] which is determined by using the procedure presented in Section 5.

Load analysis of the fractured pulley was carried out by using the original procedure which includes the dynamic effects of the resistance-to-excavation, unlike the procedure prescribed by code DIN 22261-2 [12]. Results of the FEAs indicate that the considered pulley is designed in full accordance with its function and working loads.

The considerably lower impact toughness at the temperature of –20°C points to the failures in the steelmaking technology. Metallographic examinations as well as MPIs indicate that the initial cracks in the welded joints occurred during the manufacture of the pulleys. Apart from that, the above mentioned cracks were located in the zones of maximum calculation stresses, which inevitably led to fracture. Therefore, the considered pulley fracture appeared as the result of 'manufacturing-in' defects [5, 21] which is why repairs of the spokes' welded joints of each of the pulleys were performed, without changing the design solution. This way, the downtime of the complete surface mining system, and indirect material losses were drastically reduced.

The presented investigation results underline the importance of the NDT of the vital structural parts' welded joints, both during production and the BWE's exploitation, especially in the zones of high calculated stress values. Finally, to the designers and manufacturers of the BWEs, the above mentioned investigation results present an indicator of the necessity to increase the extent of controls during the BWEs' manufacturing and assembling, prescribed by relevant standards, especially when it comes to the sub-systems whose failures can cause serious material and financial losses. Properly prescribed and conducted technical diagnostics is the basis of rational technical-economical, reliable and safe operation of the BWEs.

Acknowledgement

This work is a contribution to the Ministry of Education, Science and Technological Development of Serbia funded project TR 35006.

References

1 Araujo LS, de Almeida LH, Batista EM, Landesmann A. Failure of a bucket-wheel stacker reclaimer: metallographic and structural analyses. Journal of Failure Analysis and Prevention 2012; 12: 402-407, http://dx.doi.org/10.1007/s11668-012-9575-z.
2. Arsić M, Bošnjak S, Zrnić N, Sedmak A, Gnjatović N. Bucket wheel failure caused by residual stresses in welded joints. Engineering Failure Analysis 2011; 18(2): 700-712, http://dx.doi.org/10.1016/j.engfailanal.2010.11.009.
3. Bošnjak S, Zrnić N, Simonović A, Momčilović D. Failure analysis of the end eye connection of the bucket wheel excavator portal tie-rod support. Engineering Failure Analysis 2009; 16(3): 740-750, http://dx.doi.org/10.1016/j.engfailanal.2008.06.006.
4. Bošnjak S, Petković Z, Zrnić N, Simić G., Simonović A. Cracks, repair and reconstruction of bucket wheel excavator slewing platform. Engineering Failure Analysis 2009; 16(5): 1631-1642, http://dx.doi.org/10.1016/j.engfailanal.2008.11.009.
5. Bošnjak S, Arsić M, Zrnić N, Rakin M, Pantelić M. Bucket wheel excavator: integrity assessment of the bucket wheel boom tie - rod welded joint. Engineering Failure Analysis 2011; 18(1): 212-222, http://dx.doi.org/10.1016/j.engfailanal.2010.09.001.
6. Bošnjak S, Zrnić N. Dynamics, failures, redesigning and environmentally friendly technologies in surface mining systems. Archives of Civil and Mechanical Engineering 2012; 12(3): 348-359, http://dx.doi.org/10.1016/j.acme.2012.06.009.
7. Bošnjak S, Oguamanam D, Zrnić N. The influence of constructive parameters on response of bucket wheel excavator superstructure in the out-of-resonance region. Archives of Civil and Mechanical Engineering 2015 (article in press); http://dx.doi.org/10.1016/j.acme.2015.03.009.
8. Brkić A Đ, Maneski T, Ignjatović D, Jovančić P, Spasojević Brkić V K. Diagnostics of bucket wheel excavator discharge boom dynamic performance and its reconstruction. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2014; 16 (2): 188-197.
9. Bugaric U, Tanasijevic M, Polovina D, Ignjatovic D, Jovancic P. Lost production costs of the overburden excavation system caused by rubber belt failure. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2012; 14 (4): 333-341.
10. De Castro PMST, Fernandes AA. Methodologies for failure analysis: a critical survey. Materials & Design 2004; 25(2): 117-123, http://dx.doi.org/10.1016/j.matdes.2003.09.020.
11. DIN 17100. Steels for general structural purposes. Deutsches Institut für Normung; 1980.
12. DIN 22261-2. Bagger, Absetzer und Zusatsgeräte in Braunkohlentagebauen. Teil 2: Berechnungsgrundlagen. Deutsches Institut für Normung; 2006.
13. Dreyer E. Cost-effective prevention of equipment failure in the mining industry. International Journal of Pressure Vessels and Piping 1995; 61(2-3): 329-347, http://dx.doi.org/10.1016/0308-0161(94)00114-X.
14. EN 10002-1. Metallic materials - Tensile testing - Part 1: Method of test at ambient temperature. European Committee for Standardization; 1990.
15. EN 10045-1. Mechanical testing of metals - Charpy impact test - Part 1: Test method. European Committee for Standardization; 1990.
16. EN 1290. Non-destructive testing of welds - Magnetic particle testing of welds, European Committee for Standardization; 2004.
17. EN ISO 6506-1. Metallic materials - Brinell hardness test - Part 1: Test method. European Committee for Standardization; 2005.
18. EN ISO 2560. Welding consumables - Covered electrodes for manual metal arc welding of non-alloy and fine grain steels - Classification. European Committee for Standardization; 2009.
19. EN ISO 643. Steels - Micrographic determination of the apparent grain size, European Committee for Standardization; 2012.
20. EN 1024. Micrographic examination of the non-metallic inclusion content of steels using standard pictures, European Committee for Standardization; 2012.
21. Gagg CR. Failure of components and products by 'engineered-in' defects: Case studies. Engineering Failure Analysis 2005; 12(6): 1000-1026, http://dx.doi.org/10.1016/j.engfailanal.2004.12.008.
22. Golubović Z, Lekić Z, Jović S. Influence of bucket wheel vertical vibration on bucket-wheel excavator (BWE) digging force. Technical Gazette 2012; 19(4): 807-812.
23. Gottvald J. The calculation and measurement of the natural frequencies of the bucket wheel excavator SchRs 1320/4x30. Transport 2010; 25(3): 269-277. http://dx.doi.org/10.3846/transport.2010.33
24. Gottvald J. Analysis of vibrations of bucket wheel excavator SchRs 1320 during mining process. FME Transactions 2012; 40(4):165-170.
25. Hobbacher AF. Recommendations for fatigue design of welded joints and components. IIW Document XIII-2151r3-07/XV-1254r3-07. International Institute of Welding; 2008.
26. Hobbacher AF. The new IIW recommendations for fatigue assessment of welded joints and components - A comprehensive code recently updated. International Journal of Fatigue 2009; 31(1): 50-58, http://dx.doi.org/10.1016/j.ijfatigue.2008.04.002.
27. ISO 4516. Metallic and other inorganic coatings - Vickers and Knoop microhardness tests. International Organization for Standardization; 1980.
28. ISO 3057. Non-destructive testing, Metallographic replica techniques of surface examination, International Organization for Standardization, 2011.
29. Jovančić P, Tanasijević M, Ignjatović D. Relation between numerical model and vibration: Behavior diagnosis for bucket wheel drive assembly at the bucket wheel excavator. Journal of Vibroengineering 2010; 12(4):500-513.
30. Kowalczyk M, Czmochowski J, Rusiński E. Construction of diagnostic models of the states of developing fault for working parts of the multi-bucket excavator. Eksploatacja i Niezawodnosc- Maintenance and Reliability 2009; 42(2): 17-24
31. Mellor BG, Rainey RCT, Kirk NE. The static strength of end and T fillet weld connections. Materials & Design 1999; 20(4): 193-205, http://dx.doi.org/10.1016/S0261-3069(99)00027-8.
32. Milčić D, Miladinović S, Mijajlović M, Marković B. Determination of load spectrum of bucket wheel excavator SRs 1300 in coal strip mine Drmno. Transactions of FAMENA 2013; 37(1): 77-88.
33. Ognjanović M, Ristić M, Vasin S. BWE traction units failures caused by structural elasticity and gear resonances. Technical Gazette 2013; 20(4): 599-604.
34. Rusiński E, Czmochowski J, Iluk A, Kowalczyk M. An analysis of the causes of a BWE counterweight boom support fracture. Engineering Failure Analysis 2010; 17(1):179-191, http://dx.doi.org/10.1016/j.engfailanal.2009.06.001.
35. Rusiński E, Dragan S, Moczko P, Pietrusiak D. Implementation of experimental method of determining modal characteristics of surface mining machinery in the modernization of the excavating unit. Archives of Civil and Mechanical Engineering 2012; 12(4): 471-476, http://dx.doi.org/10.1016/j.acme.2012.07.002.

36. Rusiński E, Czmochowski J, Pietrusiak D. Selected problems in designing and constructing surface mining machinery. FME Transactions 2012; 40(4): 153-164.
37. Rusinski E, Czmochowski J, Pietrusiak D. Problems of steel construction modal models identification. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2012; 14(1): 54-61.
38. Rusiński E, Czmochowski J, Moczko P, Pietrusiak D. Assessment of the correlation between the numerical and experimental dynamic characteristics of the bucket wheel excavator in terms of the operational conditions, FME Transactions 2013; 41(4): 298-304.
39. Rusiński E, Moczko P, Odyjas P, Pietrusiak D. Investigations of structural vibrations problems of high performance machines. FME Transactions 2013; 41(4): 305-310.
40. Shukla AK, Das P, Dutta S, Ray S, Roy H. Failure analysis of a head gear pulley used in coal mines. Engineering Failure Analysis 2013; 31: 48-58, http://dx.doi.org/10.1016/j.engfailanal.2013.01.046.

Srđan BOŠNJAK
University of Belgrade
Faculty of Mechanical Engineering
Kraljice Marije 16, 11120 Belgrade 35, Serbia

Miodrag ARSIĆ
Institute for Testing of Materials IMS
Bulevar Vojvode Mišića 43, 11000 Belgrade, Serbia

Sreten SAVIĆEVIĆ
University of Montenegro
Faculty of Mechanical Engineering
Džordža Vašingtona bb, 81000, Podgorica, Montenegro

Goran MILOJEVIĆ
University of Belgrade
Faculty of Mechanical Engineering
Kraljice Marije 16, 11120 Belgrade 35, Serbia

Dušan ARSIĆ
University of Kragujevac
Faculty of Engineering
Sestre Janjić 6, 34000 Kragujevac, Serbia

E-mails: sbosnjak@mas.bg.ac.rs, miodrag.arsic@institutims.rs, sreto@ac.me, gmilojevic@mas.bg.ac.rs, dusan.arsic@fink.rs

Article citation info:
KOPECKI T, MAZUREK P, LIS T. The effect of the type of elements used to stiffen thin-walled skins of load-bearing aircraft structures on their operating properties. Experimental tests and numerical analysis. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 164–170, http://dx.doi.org/10.17531/ein.2016.2.2.

Tomasz KOPECKI
Przemysław MAZUREK
Tomasz LIS

THE EFFECT OF THE TYPE OF ELEMENTS USED TO STIFFEN THIN-WALLED SKINS OF LOAD-BEARING AIRCRAFT STRUCTURES ON THEIR OPERATING PROPERTIES. EXPERIMENTAL TESTS AND NUMERICAL ANALYSIS

WPŁYW RODZAJÓW USZTYWNIEŃ POKRYĆ CIENKOŚCIENNYCH STRUKTUR NOŚNYCH STATKÓW POWIETRZNYCH NA ICH WŁAŚCIWOŚCI EKSPLOATACYJNE. BADANIA EKSPERYMENTALNE I ANALIZA NUMERYCZNA*

The paper presents results of a study on thin-walled structures modelling representative fragments of aircraft fuselages subjected to bending and torsion. The type of the considered load and deformation corresponds to the state of such structures under in-flight conditions. The subject of the study were structures made of composite materials. Adopted assumptions include admissibility of post-buckling deformation in the operating load regime. Results of experimental studies are presented together with nonlinear numerical analyses carried out with the use of the finite elements method applied to a number of variant structures provided with various types of skin stiffening elements. Operating properties of the examined structures have been compared on the grounds of adopted criteria.

Keywords: *thin-walled structures, loss of stability, aircraft load-bearing structures, finite elements method, nonlinear numerical analyses, operating stability.*

Praca prezentuje wyniki badań ustrojów cienkościennych, stanowiących modele reprezentatywnych fragmentów struktur lotniczych, poddawanych zginaniu oraz skręcaniu. Rodzaj obciążenia oraz deformacji odpowiada stanowi struktury w warunkach eksploatacji. Przedmiotem rozważań były ustroje wykonane z kompozytów. Przyjęto założenie o dopuszczalności deformacji zakrytycznych dla obciążeń eksploatacyjnych. Przedstawiono wyniki badań eksperymentalnych i nieliniowych analiz numerycznych w ujęciu metody elementów skończonych szeregu wariantów ustrojów, zawierających różne rodzaje usztywnień pokryć. Dokonano porównania właściwości eksploatacyjnych badanych ustrojów, w oparciu o przyjęte kryteria.

Słowa kluczowe: *ustroje cienkościenne, utrata stateczności, lotnicze struktury nośne, metoda elementów skończonych, nieliniowe analizy numeryczne, trwałość eksploatacyjna.*

1. Introduction

Scientific research projects concerning the issue of the loss of stability of sub-structures constituting elements of larger load-bearing structures used in the engineering practice are usually focused on problems relating to determination of critical load values. Post-buckling states of structures are the subject of analyses much more seldom. It follows from the fact that in most of the areas of engineering, the instant at which a structure loses its stability is considered to be the moment of its destruction [1, 10, 13].

In the aircraft construction technology, in view of very specific nature of structures being the subject of considerations, equally specific standards have been established affecting both design processes and principles of operation. One of the rules, applicable to metal structures still most popular in the aircraft industry, provides admissibility of post-buckling deformations in selected types of structures under the in-flight load conditions [2, 11, 12].

It should be underlined that, similarly as in other disciplines of engineering, an iron rule is commonly accepted according to which a bar structure subjected to buckling is considered destroyed. Reliability of a structure depends therefore on adequately rigorous selection of geometrical parameters for all structure components which are represented by bar frameworks in the model design process. These include e.g. stringers, elements of frames and lattices, spar flanges, etc.

A completely different rule applies to sheet-shaped structures constituting components of semi-monocoque fuselages, however also in this case a number of limitations exist. In general, in view of the necessity to minimise the mass of the object, the loss of stability of the skin in operating conditions is assumed to be admissible provided the phenomenon has an elastic and local nature, i.e. occurs locally within the area of skin segments limited by components of the skeleton. Exceptions include skins of e.g. wing torsion boxes and other structural elements responsible for maintaining appropriate torsional rigidity, as well as these skin fragments large deformations of which are undesired in view of the necessity to maintain specific aerodynamic properties of the object [8, 14, 18, 19].

Although structures made of lightweight metals are still the basic components in most of currently operated load-bearing aircraft fuselages, recent years saw a distinct trend towards increased application of composites of various types. Those most commonly used in the

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

aircraft industry include layered composites based on glass, carbon, and aramid fibre fabrics and polymer resins.

In view of insufficient knowledge concerning the whole of changes occurring in mechanical properties of composites during prolonged operation, load-bearing components made with the use of this very technology were for many years designed and fabricated as monocoque structures with interlayers preventing stressed skins from the loss of stability.

Nowadays, striving to meet still more and more rigorous operating and economic criteria, aircraft structure designers have adopted the doctrine admitting a local loss of stability for some fragments of composite skins, similar to his accepted earlier for metal fuselage components [6, 7, 16].

This allowed to make use of semi-monocoque structures characterised with more favourable mechanical properties than those demonstrated by monocoque design solutions as far as the mass criterion is considered.

Before admitting structures of that type to operation it is necessary to carry out comprehensive studies on the effect of prolonged cyclic loads applied to composite structures on their condition and reliability.

Admitting the loss of stability in composite skins induces occurrence of design problems similar to those encountered in structures with metal sheet skins. One of such issues is related to the necessity to limit the phenomenon as effectively as possible, with the structure mass increase reduced to a minimum. It seems to be appropriate to achieve the target by employing integral stiffening elements. An indisputable merit of composite materials is the possibility of giving them arbitrary shapes. As a result, geometries of integral elements stiffening composite skins are usually different than those used for metal skins, and feasibility of realising individual design concepts is subject to different technological limitations.

Procedures applicable to design work on structures of that type processes employ universally numerical tools, including various software packages based on the finite elements method. To determine displacement distribution patterns in cases of existence of post-buckling deformations, it is necessary to realise nonlinear numerical analyses with geometrical nonlinearities taken into account. In view of the limited accuracy of numerical methods, results obtained by means of these tools are very frequently biased with significant errors [5, 9]. The only way to eliminate them consists in elaborating the numerical models by means of relevant experimental verification. In most cases, such verification can be based on simplified test carried out on models in the course of which measurements of displacements in a set of predetermined reference points are taken. Once conformity of deformation patterns and magnitudes calculated numerically and observed in the experiment is achieved, it is possible to obtain reliable reduced stress distributions within the framework of the selected mechanical strength hypothesis. The base on which such reliability is assumed is the rule of uniqueness of solutions according to which one and only one distribution of the reduced stress corresponds to each deformation state [3].

2. Purpose and scope of the study

The purpose of the study presented in this paper was to carry out a comparative analysis of several design solutions concerning a fragment of aeroplane wing with composite skin subjected to post-critical deformations under permissible operating load conditions. The subject of the research were structures with identical dimensions (Fig. 1) differing in the skeleton structure designs. In each of the cases, the front portion of the skin corresponding to the torsion box was given a larger thickness thus protecting it against the loss of stability. In the course of analyses, special attention was attached to areas between the spar and the trailing edge.

In all variants, the same skin making technology and thickness were used. As a result of the experiment, deformation distribution patterns have been obtained and representative equilibrium paths determined. The obtained results were used as a base for elaboration and verification of numerical models subjected to nonlinear analyses with the use of software based on the finite elements method.

The overall outcome of the study can be considered confirmation of appropriateness of the numerical models which can be used as a base for determining usefulness of another modifications of the analysed design solutions.

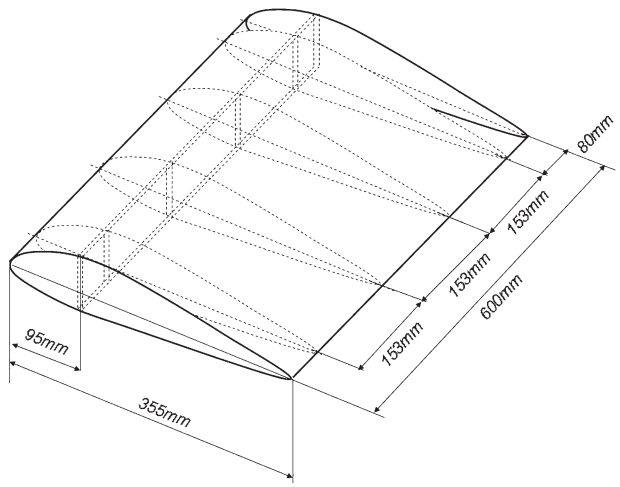


Fig. 1. Schematic views and dimensions of the examined structures

3. The experiment

The skeleton structure of the model used in the experiment was made of plywood and wooden slats with known mechanical parameters. The skin was made of an epoxy resin glass-fibre reinforced composite (GFRP).

As the composite reinforcement, Interglass glass fibre fabrics were used with weight ratios of 50 g/m^2 and 163 g/m^2. The matrix was a saturating mix based on epoxy resin MGS L285/H286 with known mechanical properties which corresponded to those measured for the composite material, i.e. E_{11} = 22000 MPa, E_{22} = 22000 MPa, ν_{12} = 0.11, G_{12} = 4600. In the torsion box area, skin of the model was made a structure comprising four layers of the symmetric fabric with reinforcement ratio of 50/50. The main directions of the composite ortotrophy were oriented at the angle of 45 degrees with respect to the direction of spar flanges.

The remaining portion of the structure contained three layers of fabric (Fig. 2). Such differentiation in the number of layers was aimed at

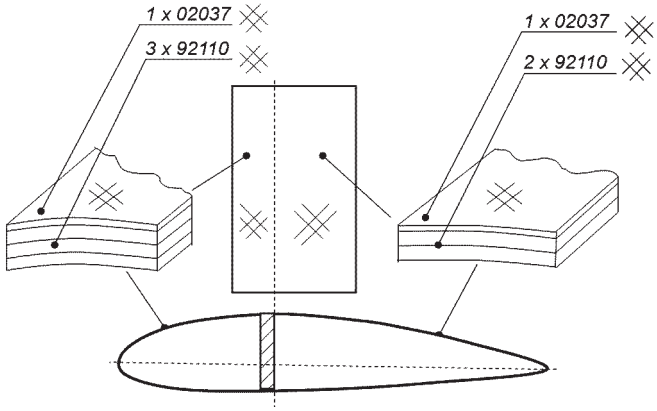
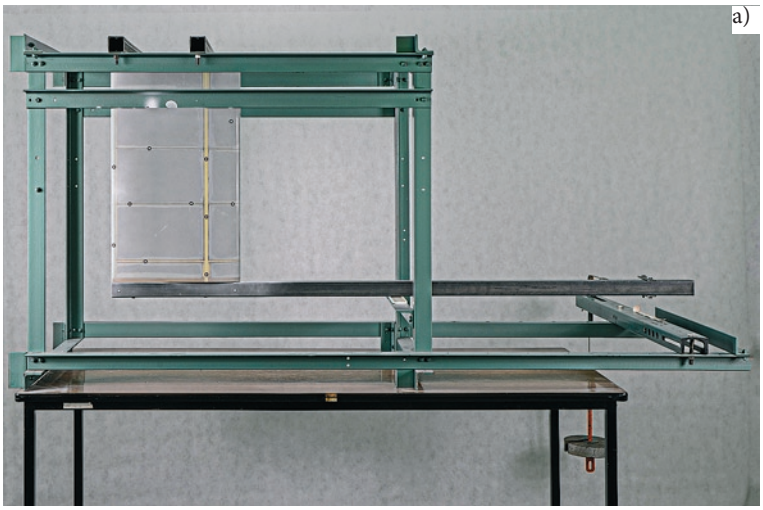


Fig. 2. Schematic representation of the laminated structure



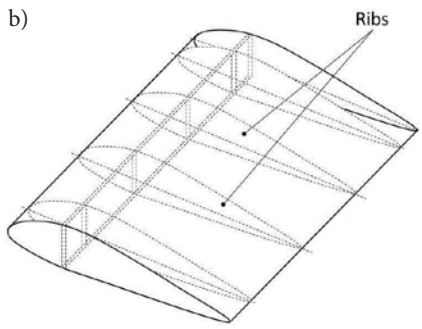


Fig. 3. (a) The experimental set-up; (b) schematic diagram of structure mounting and load application

protecting the torsion box surface against the loss of stability and creating conditions favourable to occurrence of post-critical deformations in the skin area between the spar and the trailing edge.

The experiments were carried out with the use of a specifically designed set-up (Fig. 3(a)). The examined models were subjected to torsion and bending which corresponded to the actual nature of loads applied to an aeroplane wing in the in-flight conditions (Fig. 3(b)). The load was applied gravitationally.

Three variants of the model were examined (Fig. 4).

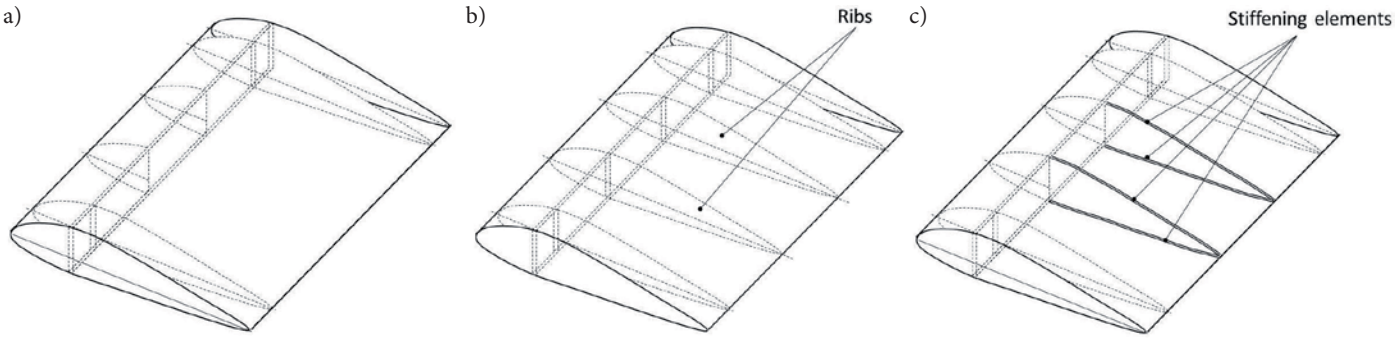


Fig. 4. Schematic sketches of examined model versions: (a) without stiffening elements; (b) with ribs; (c) with stiffening elements

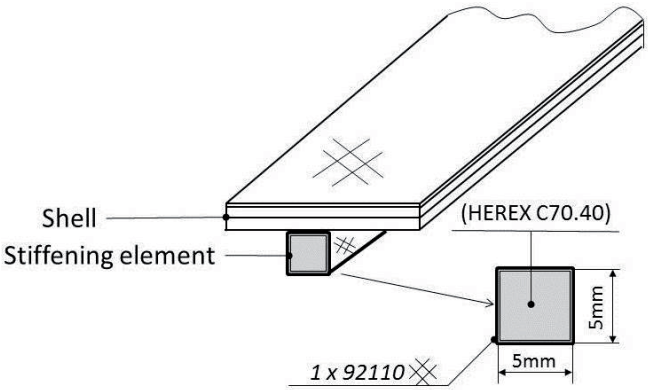


Fig. 5. A schematic view of skin stiffening elements

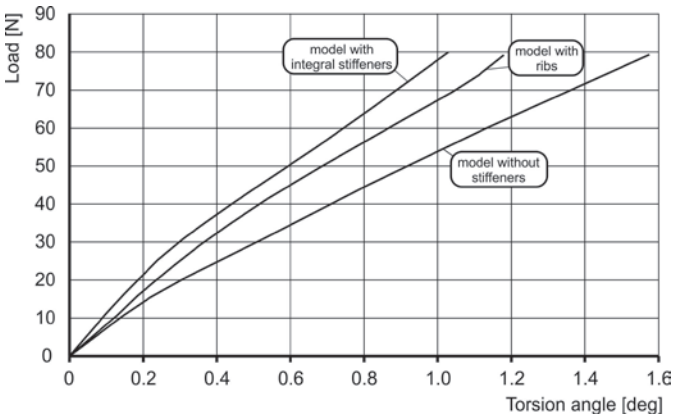


Fig. 7. Comparison of representative equilibrium paths

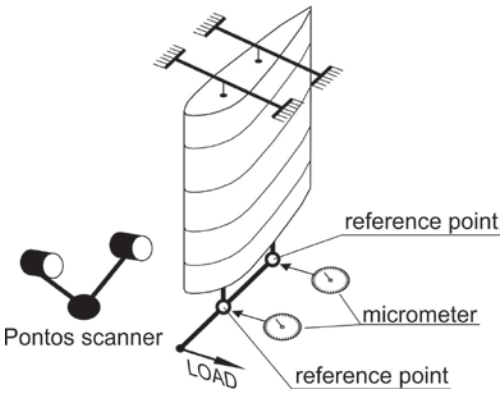


Fig. 6. Reference point displacement measuring technique

In the first variant, the tested skin fragment had no stiffening elements. Complete skeleton comprising middle ribs was applied in the second variant. In the third version of the model was provided with stiffening elements in the form of perimeters of a figures filled in with polymer foam (Fig. 5). Structure of that type would not be categorised as the integral stiffening element in the commonly adopted meaning of the term, however it serves the same function as an element in the form of perimeter of figure created by proper shaping of outer layer of the skin. Additionally, such solution is much easier to fabricate. For this reason, in the following this type of stiffening element will be referred to as the integral one.

In the course of the experiments, skin displacement measurements were taken in selected reference points for successive stationary deformation states of the structure. To this end, a system of micrometer gauges and PONTOS optical scanner by GOM Optical Measuring

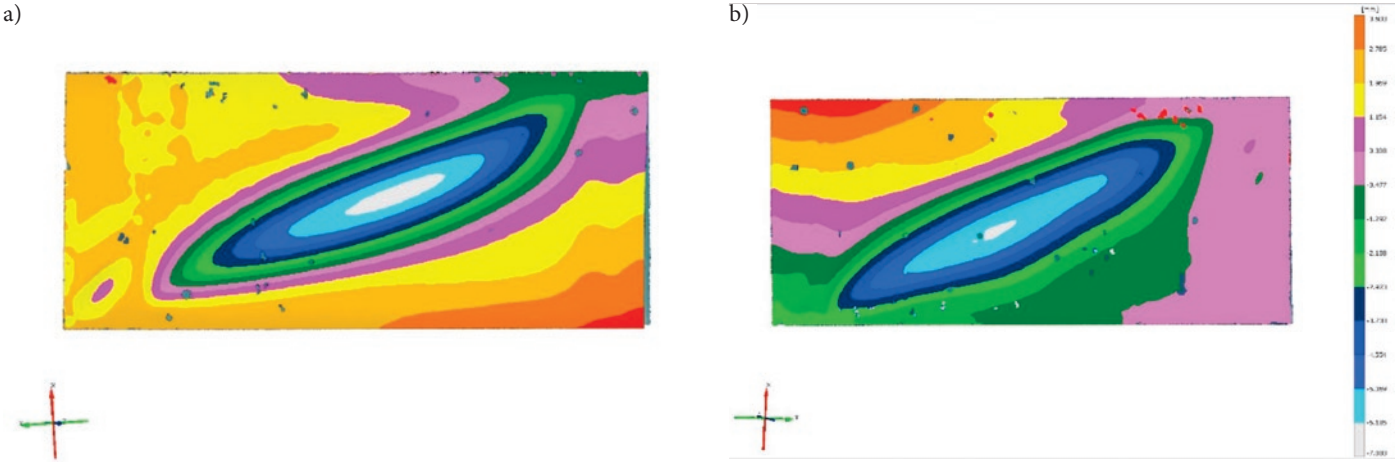


Fig. 8. Distribution of resultant displacement (in mm) — the model without stiffening elements: (a) upper skin; (b) lower skin (ribs mounted on the right)

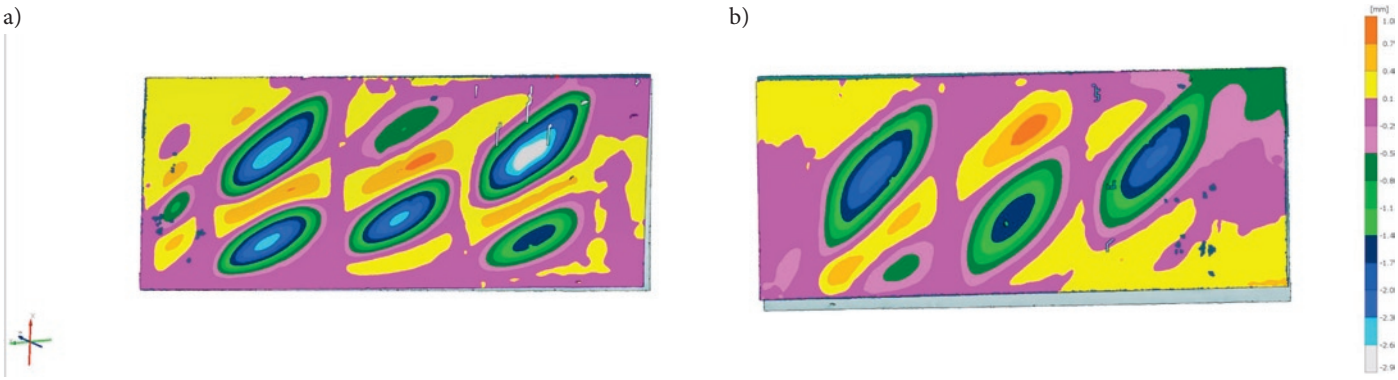


Fig. 9. Distribution of resultant displacement (in mm) — the model with ribs: (a) upper skin; (b) lower skin (ribs mounted on the right)

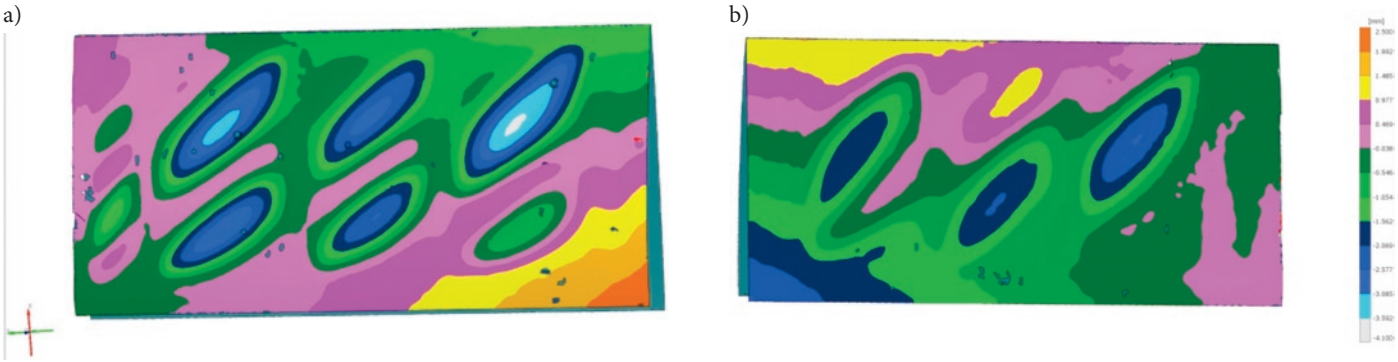


Fig. 10. Distribution of resultant displacement (in mm) — the model with stiffening elements: (a) upper skin; (b) lower skin (ribs mounted on the right)

was used (Fig. 6). As a result, representative equilibrium paths have been obtained, representing the relationship between the total angle of torsion of the structure and the load value (Fig. 7).

Further, the examined area of the deformed model was scanned with the use of ATOS scanner for the target load value. As a result, images of the deformation field were obtained (Figs. 8–10) on the grounds of which results of numerical calculations were verified.

4. Numerical analyses

A necessary complement to the experimental phase allowing to obtain information about stress distribution patterns in the examined skins consists in development of appropriate and effective calculation models implemented by means of the finite elements method.

Numerical modelling of the examined structures was realised with the use of commercial software MSC PATRAN/MARC effectiveness of which proved to be effective in case of analyses concerning post-buckling deformations of skins made of isotropic materials [11, 12]. As far as composite materials are concerned, the decisive model development phase consists in application of an algorithm the purpose of which is to determine properties of the laminate based on sets of constants characterising individual layers. The algorithm constitutes and integral subroutine of the preprocessor and any intervention of the user in its structure is impossible.

A distinctive feature of composite structures, making development of their numerical representation exceptionally difficult, is their non-homogeneity resulting not only from conditions in which individual layers are laminated but also from assembling operations, i.e. presence of local excesses of resin and diversified thickness of bonded joints. Such factors can induce local skin stiffness variations and have an effect on post-critical deformation patterns. Even small errors in selection of geometric parameters for the numerical model, introducing definite deviations from actual boundary conditions characterising a skin segment, generate significant errors in the course of nonlinear analyses.

The fundamental relationship in any nonlinear problem, determining quantitatively the relationship between condition of a structure and the load, is the so-called equilibrium path of the system constituting, in general, a hypersurface in the hyperspace of states [4, 9]. The relationship can be written as the following matrix equation of residual forces:

$$\mathbf{r}(\mathbf{u}, \mathbf{\Lambda}) = \mathbf{0}, \tag{1}$$

where u is the state vector containing displacement components of nodes of the structure corresponding to its current geometrical configuration, Λ is a matrix composed of control parameters corresponding to the actual load level, and r is the residual vector of uncompensated force components related to the current system deformation state. The set of control parameters can be represented by a single parameter λ being a function of the load. In such case, equation (1) takes the form:

$$\mathbf{r}(\mathbf{u}, \lambda) = \mathbf{0}, \tag{2}$$

called the single-parameter equation of residual forces.

The prediction-correction methods of determining the consecutive points of the equilibrium path used in contemporary software routines provide also for a correction phase based on the requirement that the system satisfies an additional equation called the increment control equation or the constraints equation [9, 17]:

$$c(\Delta\mathbf{u}_n, \Delta\lambda_n) = 0, \tag{3}$$

where the increments:

$$\Delta\mathbf{u}_n = \mathbf{u}_{n+1} - \mathbf{u}_n, \quad \text{and} \quad \Delta\lambda_n = \lambda_{n+1} - \lambda_n$$

correspond to transition from state n to state $n + 1$.

In view of lack of possibility to represent equilibrium paths for systems with more than two degrees of freedom in a form of easily interpretable plots, for the purpose of comparison, the so-called representative equilibrium paths are used in practice that represent a functional relationship between a selected parameter characterising deformation of the system and a single control parameter related to the applied load. Reliability of results obtained from FEM-based nonlinear numerical analyses is usually accepted when a satisfactory coherence is found between two representative equilibrium paths, namely the actual one determined in the course of an experiment and this obtained numerically. It is also necessary to obtain convergence between the forms of deformations following from the calculations [4, 9, 17] with results of relevant experiment. On the grounds of the above-mentioned rule of uniqueness of solutions, the obtained reduced stress distributions in the deformed skin can be therefore also considered reliable [3].

As the nonlinear numerical analysis is an iterative process aimed at finding successive equilibrium states, its correctness is to a large degree determined by correct choice of the prognostic method, the correction strategy, and a number of control parameters. In the case described here, the Newton-Raphson method was used in combination with the Crisfield hyperspherical correction strategy [4, 15].

By contrast with linear analyses where the goal is to obtain as high number of finite elements as possible, selection of excessively dense mesh of elements for nonlinear analyses leads to faulty calculation results in many cases with the calculation time significantly extended. After a series of numerical tests aimed at selection of a proper topology for the model, it has been decided to use a model comprising 5000 four-node skin elements. The necessity to employ such elements resulted from the fact that other types available in the MSC MARC software library to which properties of laminated composites could be assigned do not offer the possibility to reproduce geometrically complex objects, in view of the type and number of the degrees of freedom.

Models of the materials were developed taking into account mechanical properties of composites fabricated with the use of components used in the experimental phase and characterised by means of material constants quoted in preceding chapter.

As a result of the performed nonlinear numerical analyses, representative equilibrium paths have been determined and compared to corresponding characteristic obtained from the experiment (Fig. 11).

Very small discrepancies between corresponding representative curves were found in all the cases subjected to analysis which confirmed that the numerical methods and parameters controlling nonlinear procedures were selected correctly.

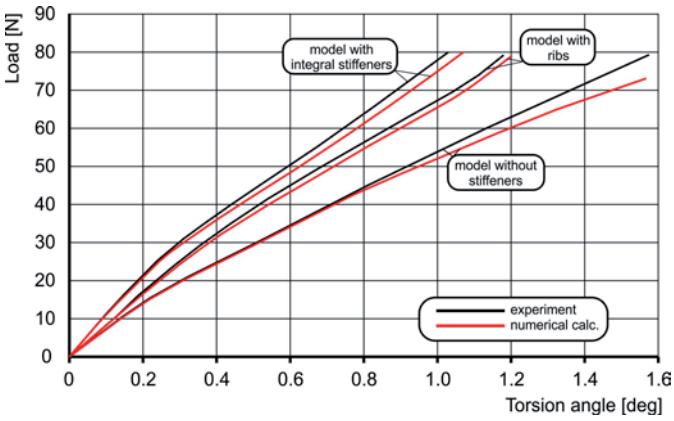


Fig. 11. Comparison of representative equilibrium paths

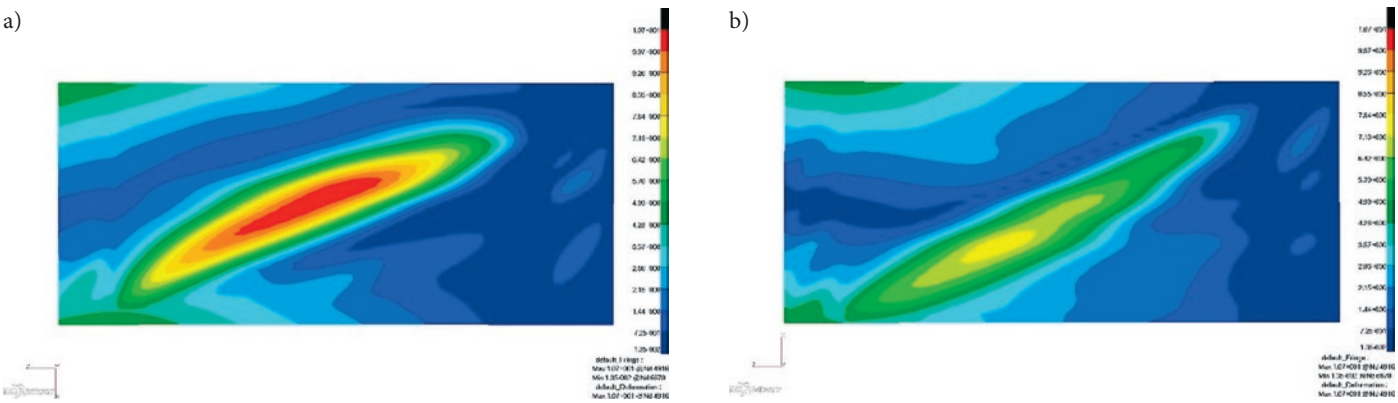


Fig. 12. Resultant displacement distribution patterns for the analysed skin fragments (in mm) — model without stiffening elements: (a) upper skin; (b) lower skin

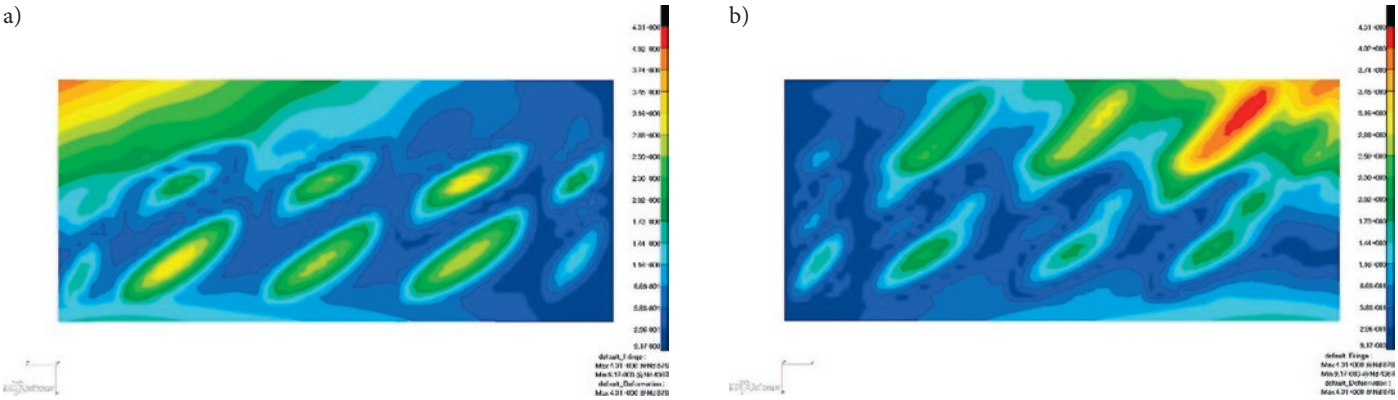


Fig. 13. Resultant displacement distribution patterns for the analysed skin fragments (in mm) — model with ribs: (a) upper skin; (b) lower skin

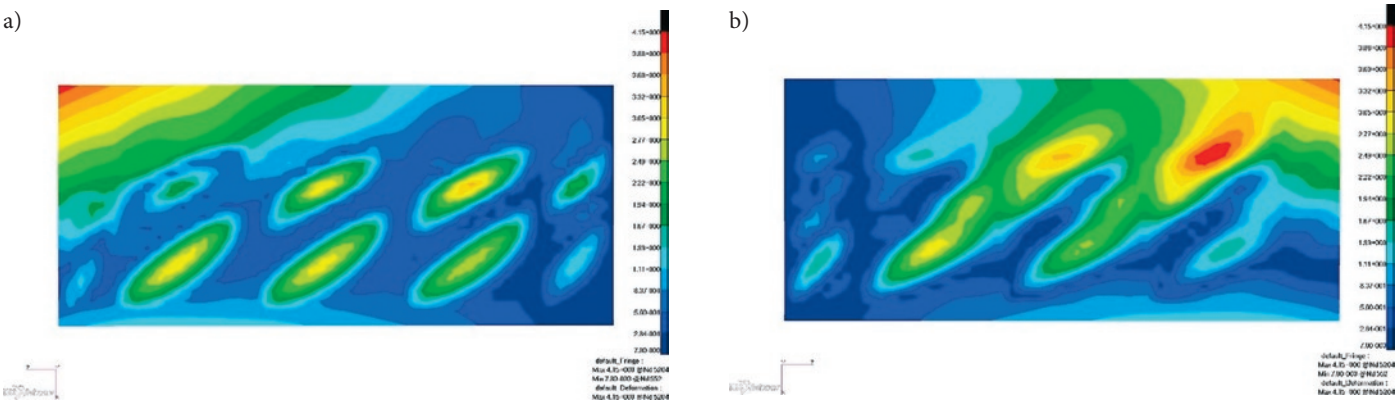


Fig. 14. Resultant displacement distribution patterns for the analysed skin fragments (in mm) — model with integral stiffening elements: (a) upper skin; (b) lower skin

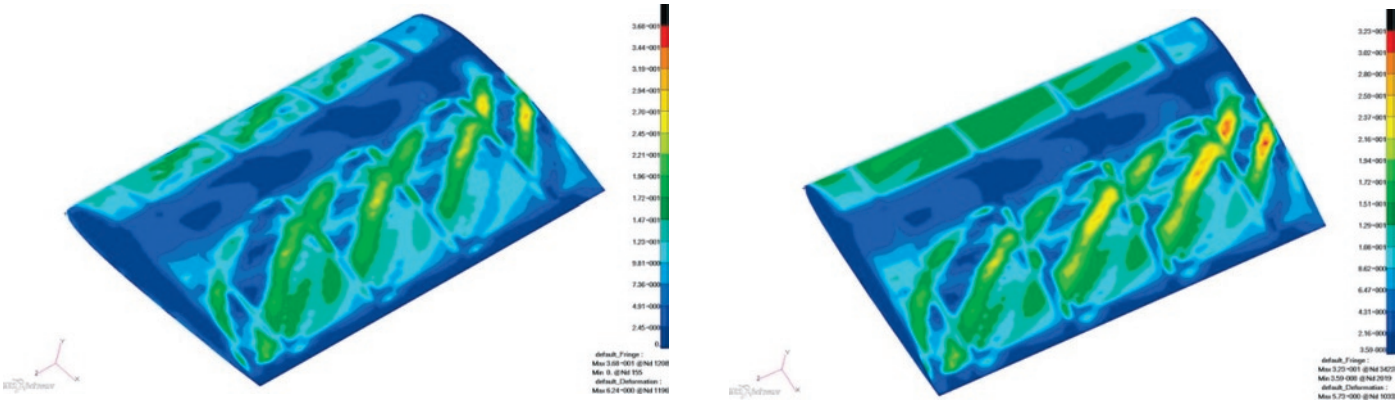


Fig. 15. Reduced stress distributions according to σmax hypothesis (in MPa): (a) model with ribs; (b) model with integral stiffening elements

As a result of the analyses, distributions patterns of resultant displacements of the finite element mesh nodes constituting skin models haven been also obtained (Figs. 12–14).

Taking into account the satisfactory similarity between results of analyses on one hand and the experiment on the other observed in the scope of both representative equilibrium paths and resultant displacements, it can be stated that properties of composites attributed to finite elements by PATRAN software, determined by the program based on data for individual layers of the composite, may be considered correct and corresponding to actual characteristics. However, it should be emphasised that, in case of occurrence of any defects in the real structure that may arise in the process of lamination, it is necessary to introduce appropriate correction in the numerical model accounting for the effect of such flaws on local stiffness of the skin.

The courses of the representative equilibrium paths allow to claim that the wing skin stiffening elements constitute a necessary component of the structure. In the model lacking these features, loss of stability occurred at very low load value as a result of occurrence of a field of tensions within the area of the observed segment.

On the other hand, comparison of solutions providing for ribs and integral stiffening elements have proven superiority of the latter which was also accompanied by a decrease of mass.

The increased rigidity of the whole structure can be explained by the fact that when ribs are employed, skin displacement in directions tangent to its surface at points where the skin is joined with the ribs becomes limited. This results in occurrence of different stress distributions in segments of the skin, more unfavourable from the point of view of overall effectiveness of the solution (Fig. 15).

5. Summary and conclusions

The combined experimental and numerical analysis of a number of design solutions for a typical fragment of thin-walled aircraft structure was focused on two fundamental goals. The first of the objectives

included determination and comparison of operating properties of the examined structures with a maximum possible accuracy. The study has revealed that integral elements stiffening the skin applied in one of the considered model version, although based on a very simple technological solution, turned out to be more effective than the traditional variant providing for skin division by means of ribs. This allows to mark out a direction for further research aimed at determining an integral stiffening variant with properties most favourable in operating conditions specific for given structure. A criterion for selection of a target solution could be the largest critical load value or the smallest value of a deformation adopted as the representative one, with the structure mass changes taken into account at the same time.

The presented results of experimental studies and nonlinear numerical analyses, as well as conclusions derived from them, should be therefore assessed in the context of a broader framed research project aimed at determining properties of a number of stiffening elements for aircraft composite skins subjected to post-critical deformations under operating load conditions.

Another objective of the research project, constituting an indispensable complement to the experimental phase allowing to obtain knowledge of stress distribution patterns in the examined skins, consisted in development of appropriate and sufficiently effective FEM calculation models. Verified numerical models can constitute a very effective tool useful in search for another design solutions developed by means of using various combinations of stiffening elements. It should be however stressed that the last step must consist in performing an experiment with the use of a model corresponding to the selected variant. This follows from the absolute imperative to verify numerical models. Nevertheless, such methodology allows to eliminate the experimental component from intermediate design development stages identified as those leading to solutions that do not meet the selected criteria.

In the light of the above conclusions it can be found that the presented research methodology may prove finally to be an effective tool allowing to design skin stiffening solutions most favourable from the point of view of mass- and rigidity-related criteria. Meeting these criteria means that the desired operating properties of the examined structures have been achieved.

References

1. Arborcz J. Post-buckling behavior of structures. Numerical techniques for more complicated structures. Lecture Notes In Physics 1985; 288: 83-142.
2. Aborcz J., Hol J.M.A. Recent development in shell stability analysis, Report LR-633 Faculty of Aerospace Engineering. Delft: University of Technology, 1990.
3. Brzoska Z. Statyka i stateczność konstrukcji. Warszawa: PWN, 1965.
4. Crisfield M. A Non-linear finite element analysis of solid and structures. J. Wiley & Sons, 1997.
5. Doyle J.F. Nonlinear analysis of thin-walled structures. Springer-Verlag, 2001, http://dx.doi.org/10.1007/978-1-4757-3546-8.
6. Dębski H., Kubiak T., Teter A. Experimental investigation of channel-section composite profiles' behavior with various sequences of plies subjected to static compression. Thin-Walled Structures, 2013; 71: 147-154, http://dx.doi.org/10.1016/j.tws.2013.07.008.
7. Dębski H. Experimental investigation of post-buckling behavior of composite column with top-hat cross-section. Eksploatacja i Niezawodnosc - Maintenance and Reliability, 2013; 16 (2): 1056-109.
8. Dobrzański P., Czarnocki P., Lorenz Z. Shell structures - theory and application. CRC Press, 2013, ISSN/ISBN 978-1-138-000-82-7, Vol. 3.
9. Felippa C. A., Crivelli L. A., Haugen B. A survey of the core-congruential formulation for nonlinear finite element. Archive of Computer Methods in Engineering, 1994, http://dx.doi.org/10.1007/BF02736179.
10. Jachimowicz J., Szymczyk E., Słowiński G., Derewońko A., Wronicz W. Modele globalne i lokalne MES w analizie struktur lotniczych na przykładzie fragmentu skrzydła samolotu M-28 Skytruck. Wyd. WAT, 2010, Biuletyn WAT Nr 1/2010.
11. Kopecki T., Mazurek P. Problems of numerical bifurcation reproducing in post-critical deformation states of aircraft structures. Journal of Theoretical and Applied Mechanics, 2013; 51(4): 969-977.
12. Kopecki T., Mazurek P. Numerical representation of post-critical deformations in the processes of determining stress distributions in closed multi-segment thin-walled aircraft load-bearing structures. Eksploatacja i Niezawodnosc - Maintenance and Reliability, 2014; 16(1):164-169.
13. Lynch C. A. Finite element study of the post buckling behavior of a typical aircraft fuselage panel. PhD Thesis, Queen's University Belfast, 2000.
14. Mielczarek G. Optimization of the structure of a multilayer cylindrical shell under stability loss conditions. Engineering Transaction, 1995, 43(4): 495-504.
15. Rakowski G., Kacprzyk Z. The finite elements method in mechanics of structures. Oficyna Wydawnicza Politechniki Warszawskiej, 1993.
16. Rudawska A., Dębski H. Experimental and numerical analysis of adhesively bonded aluminium alloy sheets joints, Eksploatacja i Niezawodnosc - Maintenance and Reliability, 2011; 49(1): 4-10.
17. Riks E. An incremental approach to the solution of snapping and buckling problems. International Journal of Solid and Structures, 1979; 15: 529-551, http://dx.doi.org/10.1016/0020-7683(79)90081-7.
18. Szymczyk E., Niezgoda T. Numeryczna analiza wyboczenia lokalnego użebrowanej powłoki cylindrycznej. Biuletyn Wojskowej Akademii Technicznej: budowa i eksploatacja maszyn, 2003; 1(605), 137-154.
19. Wiśniowski W. Sztywność i rozsztywnienie konstrukcji lotniczych. Prac Instytutu Lotnictwa No. 214, 2011, ISSN 0509-6669.

Tomasz KOPECKI
Przemysław MAZUREK
Tomasz LIS
Department of Mechanical Engineering and Aviation
Rzeszów University of Technology
al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
E-mails: tkopecki@prz.edu.pl, pmazurek@prz.edu.pl, list@prz.edu.pl

Article citation info:
NI X, ZHAO J, SONG W, GUO C, LI H. Nonlinear degradation modeling and maintenance policy for a two-stage degradation system based on cumulative damage model. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 171–180, http://dx.doi.org/10.17531/ein.2016.2.3.

Xianglong NI
Jianmin ZHAO
Wenyuan SONG
Chiming GUO
Haiping LI

NONLINEAR DEGRADATION MODELING AND MAINTENANCE POLICY FOR A TWO-STAGE DEGRADATION SYSTEM BASED ON CUMULATIVE DAMAGE MODEL

SYSTEM CHARAKTERYZUJĄCY SIĘ DWUETAPOWYM PROCESEM DEGRADACJI: NIELINIOWE MODELOWANIE DEGRADACJI ORAZ WYZNACZANIE STRATEGII EKSPLOATACJI SYSTEMU NA PODSTAWIE MODELU SUMOWANIA USZKODZEŃ

This paper attempts to take into account a two-stage degradation system which degradation rate is non-stationary and change over time. The system degradation is thought to be caused by shocks, and system degradation model is established based on cumulative damage model. The nonlinear degradation process is expressed by different shock damage and shock counting. And shock damage and shock counting are assumed to be Gamma distribution and non-homogeneous Poisson process, respectively. On the basis of these, system reliability model and nonlinear degradation model are given. In order to optimal maintenance policy for considered system, adaptive maintenance policy and time-dependent maintenance policy are studied, and mean maintenance cost rate is established to evaluate the maintenance policies. Numerical examples are given to analyze the influences of degradation model parameters and find optimal maintenance policy for considered system.

Keywords: *two-stage, nonlinear, degradation modeling, cumulative damage model, maintenance policy.*

W przedstawionym artykule badano system, w którym proces degradacji zachodzi dwuetapowo, a szybkość degradacji jest zmienna w czasie. Przyjęto, że do degradacji systemu dochodzi w wyniku wstrząsów. Model degradacji systemu oparto na modelu sumowania uszkodzeń. Nieliniowy proces degradacji określono jako taki, w którym uszkodzenie powodowane wstrząsem oraz częstotliwość wstrząsów są wartościami zmiennymi. Przyjęto, że uszkodzenie powodowane wstrząsem ma rozkład gamma a częstotliwość wstrząsów jest niejednorodnym procesem Poissona. Na tej podstawie utworzono model niezawodności systemu oraz model degradacji nieliniowej. W celu opracowania optymalnej strategii eksploatacji dla rozpatrywanego systemu, rozważono dwa typy strategii utrzymania ruchu: strategię adaptacyjną oraz strategię czasowo-zależną. Strategie te oceniano określając średni poziom kosztów eksploatacji. Przykłady numeryczne posłużyły do analizy wpływu parametrów modelu degradacji oraz pozwoliły określić optymalną strategię utrzymania dla rozpatrywanego systemu.

Słowa kluczowe: *dwu-etapowy, nieliniowy, modelowanie degradacji, model sumowania uszkodzeń, strategia eksploatacji.*

Nomenclature

M_k	The kth (k=1,2) stage of system degradation process
x_i^k	System damage value due to the ith shock in degradation stage M_k
N_k	The number of shock counting in degradation stage M_k
t_c	The change-point that degradation stage from M_1 to M_2
$Y(t)$	System whole degradation level at time t
$Y^k(t)$	Cumulative degradation quantity just for the kth stage
$\lambda_k(t)$	The Poisson parameter that system at time t in degradation stage M_k
$g_c(t)$	The probability density function of change-point t_c
Y_f	System failure threshold
T_f	The time point of system failure
Y_k	System alarm threshold in degradation stage M_k
T_i	The ith inspection time
ΔT_k	The interval of inspection in degradation stage M_k
ΔT^i	The ith interval of inspection
$E(C)$	Mean maintenance cost rate
C_I	Unit cost of an inspection action
C_P	Unit cost of a preventive maintenance action
C_C	Unit cost of a corrective maintenance action
$E[C(T)]$	Total maintenance cost in a renewal cycle T
$E[T]$	Mean time length of renewal cycle T
$E[N_I(T)]$	Average number of inspection counting in a renewal cycle T
P_P	Probability of performing preventive maintenance in a renewal cycle T
P_C	Probability of performing corrective maintenance in a renewal cycle T

1. Introduction

Degradation analysis is a research hotspot for prognostic and health management (PHM), which can be used for estimating failure-time distribution [16], predicting remaining useful life (RUL) distri-

bution [3, 23] and exploring preventive maintenance policy [12, 25]. Especially, degradation process modeling is an important approach for evaluating the reliability of high reliable products [1, 2].

Stationary degradation process has been studied intensively to optimize maintenance problems. However, the degradation process of some systems present two-stage feature due to the influences of internal mechanism and external environment etc [7], where the degradation rate is suddenly increased. For example, the vibration-based degradation signals of bearings [8] and vibration signals special frequency band energy of gearboxes [14] exhibit two-stage characteristic in degradation test. There are two typical models with independent random increment, continuous time model [7] and cumulative damage model [21], that can be used to present system degradation process. Some researchers have studied on the degradation process modeling for two-stage degradation system. But in most articles [5, 6, 17, 19] the degradation processes are assumed to be continuous Gamma process, and degradation rates for different degradation stages are presented by different Gamma parameters. Wiener process is also used for two-stage degradation modeling [9]. In existing studies, the degradation process is mostly supposed to be continuous and modelling by continuous time model. But for some systems, their degradations are caused by shocks and their increases of degradation levels are step, such as reciprocating machine. Furthermore, some system degradation quantities, which are collected by interval monitoring, can be considered as causing by shocks, even if the system degradation process is continuous.

Condition-based maintenance (CBM) is an importance approach for reducing maintenance cost to gradual degradation system [11]. This maintenance decision-making method is also effective for two-stage degradation system. But the degradation rate suddenly increased will bring significant impact on maintenance policy. On the basis that change-point of degradation rate can be monitored perfected, Saassouh [19] put forward an activation zone to plan the maintenance action for a two-stage system. Fouladirad [5, 7] proposed an adaptive maintenance policy based on online change detection procedures, where alarm thresholds were diverse in different degradation stage as the degradation rate change. Ponchet [17] assumed that change-point of degradation rate cannot be monitored, and he developed two condition-based maintenance optimization models with and without considering the change-point in system degradation process, respectively, the numerical results showed that it can bring considerable benefits if degradation rate changing was considered in maintenance policy. In these existing studies, the mean degradation rates in the first stage and the second stage were both considered as fixed, and the increased process of degradation level presented linear. In 2011, Fouladirad [6] took into account a system with time-dependent degradation rate after change-point, but the degradation process was assumed to be continuous Gamma process. Meanwhile, he studied a condition-based maintenance policy with time-dependent alarm thresholds in the second degradation stage. But the interval of inspection was considered as fixed no matter how the degradation rate changed, it was difficult to achieve the best maintenance policy.

This paper considers degradation modeling and maintenance policy for a two-stage degradation system, which degradation process is nonlinear and degradation rate is changed over time in both stages. The main contributions of this study are: (a) Considering some system degradation are caused by shocks, cumulative damage model is used for two-stage degradation process modeling, and the degradation rate are presented by different shock damage and shock counting. (b) As the degradation rate is changed over time, a time-dependent maintenance policy is proposed, which the interval of inspection is time-dependent.

The remainder of this paper is organized as follows. In section 2, a two-stage degradation system is presented and system degradation modelling method is studied. Two kinds of maintenance policy and maintenance policy evaluation method are given in section 3. In section 4, numerical examples are used to analyze the influences of different parameters for two-stage degradation model. Conclusions are made in section 5.

2. Nonlinear degradation modeling for a two-stage degradation system

2.1. Two-stage degradation system

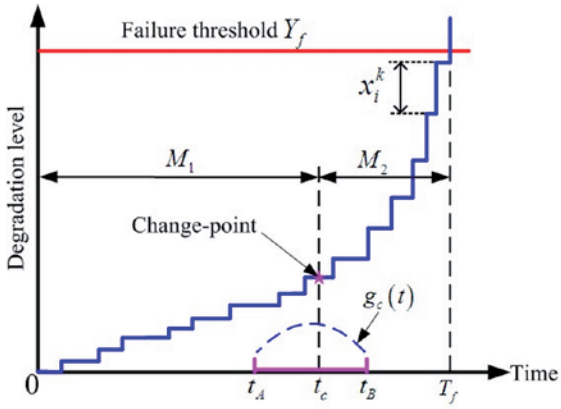


Fig. 1. Two-stage degradation process

The system with two-stage degradation process considered in this paper is described as follows (as shown in Fig. 1):

- In system degradation process, mean degradation rate suddenly increase at a random time point. And the time point is denoted by change-point t_c. Before t_c, system is in nominal degradation stage M_1 and mean degradation rate is small. After t_c, system is in accelerated degradation stage M_2 and mean degradation rate is large. The degradation rate of whole degradation process is non-stationary and become larger in terms of working time, so the degradation process is nonlinear.
- System degradation level at time t can be summarized by a scalar aging variable $Y(t)$. There is no doubt that $Y(t)$ is an increasing stochastic scalar. System initial state is assumed to be intact in this paper, namely $Y(0)=0$. System will be considered as failed if degradation level $Y(t)$ exceeds failure threshold Y_f. And system stops functioning either for economic reasons or for safety reasons when $Y(t)$ is greater than Y_f.
- System degradation process can be thought as step, and degradation level is the sum of large numbers of tiny damage values. Therefore, the whole degradation level at working time t can be expressed by cumulative damage model [22] as

$$Y(t)=\sum_{i=1}^{N_1}x_i^1\cdot \mathrm{I}_{\{t\le t_c\}}+\left(\sum_{i=1}^{N_1}x_i^1+\sum_{j=1}^{N_2}x_j^2\right)\cdot \mathrm{I}_{\{t>t_c\}} \tag{1}$$

Where N_k (k=1,2) is the number of shock counting in degradation stage M_k, x_i^k is the shock damage value for ith shock in degradation stage M_k. $I_{\{E\}}=1$ if E is true and otherwise $I_{\{E\}}=0$. When $t>t_c$, degradation level is the sum of the damage in the first stage M_1 and the second stage M_2. In this case, the working time length of system degradation in stage M_1 is t_c and in stage M_2 is t-t_c.

As the characteristic of cumulative damage model [17,22], system degradation rate is determined by damage value per shock and shock counting per unit time. In this paper, in order to show shock counting

changes over time, the shock counting N_k is assumed to be non-homogeneous Poisson process (NHPP) and with Poisson parameter $\lambda_k(t)$ at time t in stage M_k [4,24]. That is, the probability of shock counting N_k equals to m during (0, t) in stage M_k can be written as:

$$P\left(N_k = m\right) = \frac{\left(\int_0^t \lambda_k\left(u\right)du\right)^m}{m!} \cdot e^{-\int_0^t \lambda_k(u)du} \tag{2}$$

Because all the shocks are independent in whole degradation process, the probability of shock counting in the first stage N_1=m and in the second stage N_2=n is:

$$P\left(N_1 = m, N_2 = n\right) = \frac{\left(\int_0^{t_c} \lambda_1\left(u\right)du\right)^m}{m!} \cdot \frac{\left(\int_{t_c}^{t} \lambda_2\left(\tau - t_c\right)d\tau\right)^n}{n!} \cdot e^{-\int_0^{t_c} \lambda_1(u)du - \int_{t_c}^{t} \lambda_2(\tau - t_c)d\tau} \tag{3}$$

As Gamma process is suitable for describing monotonic degradation [13, 15], shock damage is assumed to be Gamma distribution in this study (It is important to note that the Gamma distribution is used for shock damage by a shock in this study, but in literatures [5-7, 17, 19] the Gamma distribution is used for degradation level of whole continuous degradation process). In this paper, if the ith shock in stage M_k occurs at t_i, the shock damage is $x_i^k \sim Ga\left(\alpha_k\left(t_i\right), \beta_k\right)$ ($\alpha_k\left(t_i\right)$ is shape-parameter, β_k is scale-parameter). $Y^k(t)$ is the cumulative degradation quantity just for the kth stage (namely whole degradation level $Y(t)$=$Y^1(t)$+$Y^2(t)$ when t>t_c). Meanwhile, as every shock is independent in degradation process, it can be known from Gamma theorem that $Y^k(t)$ also follows Gamma distribution. N_k is the shock counting in stage M_k. When N_k=1,2,3,⋯, $Y^k(t)$ can be written as:

$$Y^k\left(t\right) = \sum_{i=1}^{N_k} x_i^k \sim Ga\left(\sum_{i=1}^{N_k} \alpha_k\left(t_i\right), \beta_k\right) \tag{4}$$

The corresponding probability density function (PDF) is:

$$f_{N_k}^k\left(y\right) = \frac{1}{\Gamma\left(\sum_{i=1}^{N_k} \alpha_k\left(t_i\right)\right) \cdot \beta_k^{\sum_{i=1}^{N_k} \alpha_k\left(t_i\right)}} \cdot y^{\sum_{i=1}^{N_k} \alpha_k\left(t_i\right) - 1} \cdot e^{-y/\beta_k} \cdot \mathrm{I}_{\{y \ge 0\}} \tag{5}$$

Where Γ is the Euler's Gamma function, $\Gamma\left(\alpha\right) = \int_0^\infty u^{\alpha-1} e^{-u} du$.

When N_k =0, the considered system is undamaged in degradation stage M_k and $Y^k(t)$=0.

2.2. Reliability modeling

There are many reasons can cause the transition of degradation rate. Therefore, in engineering practice, change-point t_c should be not a fixed parameter but a variable in degradation process. And it is difficult to monitor the specific time point of change-point t_c in engineering. But observing some system degradation data and experimental data, it can be found that the change-point mostly falls in a certain range. Moreover, the change-point information can be obtained by statistical from degradation data. In this paper, it is assumed that change-point t_c falls in time interval [t_A, t_B] with PDF $g_c(t)$ and $\int_{t_A}^{t_B} g_c\left(t\right)dt = 1$, as shown in Fig. 1.

The system is reliable when degradation level $Y(t)$ does not exceed the failure threshold Y_f. As key parameters, shock damage x_i^k, Poisson parameter $\lambda_k\left(t\right)$ and change-point t_c all should be considered in reliability modeling of two-stage degradation system. System reliability modeling is divided into three periods, before change-point (0≤ t ≤t_A), after change-point (t >t_B) and change-point interval (t_A<t≤t_B).

When 0≤ t ≤t_A, the system reliability is:

$$\begin{aligned} R_1\left(t\right) &= P\left(Y\left(t\right) < Y_f\right) = P\left(N_1 = 0\right) + \sum_{m=1}^{\infty} P\left(\sum_{i=1}^{N_1} x_i^1 < Y_f \middle| N_1 = m\right) \cdot P\left(N_1 = m\right) \\ &= e^{-\lambda_1 t} + \sum_{m=1}^{\infty}\left(\int_0^{Y_f} f_{m\alpha_1,\beta_1}\left(w\right)dw \cdot \frac{\left(\lambda_1 t\right)^m}{m!}\right) \cdot e^{-\lambda_1 t} \end{aligned} \tag{6}$$

When t >t_B, the system reliability is:

$$\begin{aligned} R_2(t) &= P\left(Y\left(t\right) < Y_f\right) \\ &= P\left(N_1 = 0, N_2 = 0\right) + \sum_{m=1}^{\infty} P\left(\sum_{i=1}^{N_1} x_i^1 < Y_f \middle| N_1 = m\right) \cdot P\left(N_1 = m\right) \cdot P\left(N_2 = 0\right) + \\ &\quad P\left(N_1 = 0\right) \cdot \sum_{n=1}^{\infty} P\left(\sum_{j=1}^{N_2} x_j^2 < Y_f \middle| N_2 = n\right) \cdot P\left(N_2 = n\right) + \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\left(\sum_{i=1}^{N_1} x_i^1 + \sum_{j=1}^{N_2} x_j^2 < Y_f \middle| N_1 = m, N_2 = n\right) \\ &= \int_{t_A}^{t_B}\left(e^{-\lambda_1 u - \lambda_2(t-u)} \cdot g_c\left(u\right)\right)du + \int_{t_A}^{t_B}\left(\sum_{m=1}^{\infty}\left(\int_0^{Y_f} f_{m\alpha_1,\beta_1}\left(w\right)dw \cdot \frac{\left(\lambda_1 u\right)^m}{m!}\right) \cdot e^{-\lambda_1 u - \lambda_2(t-u)} \cdot g_c\left(u\right)\right)du + \\ &\quad \int_{t_A}^{t_B}\left(\sum_{n=1}^{\infty}\left(\int_0^{Y_f} f_{n\alpha_2,\beta_2}\left(v\right)dv \cdot \frac{\left(\lambda_2\left(t-u\right)\right)^n}{n!}\right) \cdot e^{-\lambda_1 u - \lambda_2(t-u)} \cdot g_c\left(u\right)\right)du + \\ &\quad \int_{t_A}^{t_B}\left(\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\left(\int_0^{Y_f}\int_0^{Y_f - w}\left(f_{m\alpha_1,\beta_1}\left(w\right) \cdot f_{n\alpha_2,\beta_2}\left(v\right)\right)dvdw \cdot \frac{\left(\lambda_1 u\right)^m}{m!} \cdot \frac{\left(\lambda_2\left(t-u\right)\right)^n}{n!}\right) \cdot e^{-\lambda_1 u - \lambda_2(t-u)} \cdot g_c\left(u\right)\right)du \end{aligned} \tag{7}$$

The same to formulas (6) and (7), when $t_A < t \le t_B$, the system reliability is:

$$\begin{aligned}
R_3(t) &= P\left(Y(t) < Y_f \,|\, t_A < t \le t_B\right) \\
&= P\left(Y(t) < Y_f \,|\, t_A < t \le t_c\right) \cdot P\left(t_A < t \le t_c\right) + P\left(Y(t) < Y_f \,|\, t_c < t \le t_B\right) \cdot P\left(t_c < t \le t_B\right) \\
&= \left(e^{-\lambda_1 t} + \sum_{m=1}^{\infty} \left(\int_0^{Y_f} f_{m\alpha_1,\beta_1}(w)\,dw \cdot \frac{(\lambda_1 t)^m}{m!} \right) \cdot e^{-\lambda_1 t} \right) \cdot \int_t^{t_B} g_c(u)\,du + \\
&\quad \int_{t_A}^{t} \left(e^{-\lambda_1 u - \lambda_2 (t-u)} \cdot g_c(u) \right) du + \int_{t_A}^{t} \left(\sum_{m=1}^{\infty} \left(\int_0^{Y_f} f_{m\alpha_1,\beta_1}(w)\,dw \cdot \frac{(\lambda_1 u)^m}{m!} \right) \cdot e^{-\lambda_1 u - \lambda_2 (t-u)} \cdot g_c(u) \right) du + \\
&\quad \int_{t_A}^{t} \left(\sum_{n=1}^{\infty} \left(\int_0^{Y_f} f_{n\alpha_2,\beta_2}(v)\,dv \cdot \frac{\left(\lambda_2 (t-u)\right)^n}{n!} \right) \cdot e^{-\lambda_1 u - \lambda_2 (t-u)} \cdot g_c(u) \right) du + \\
&\quad \int_{t_A}^{t} \left(\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \left(\int_0^{Y_f} \int_0^{Y_f - w} \left(f_{m\alpha_1,\beta_1}(w) \cdot f_{n\alpha_2,\beta_2}(v) \right) dv\,dw \cdot \frac{(\lambda_1 u)^m}{m!} \cdot \frac{\left(\lambda_2 (t-u)\right)^n}{n!} \right) \cdot e^{-\lambda_1 u - \lambda_2 (t-u)} \cdot g_c(u) \right) du
\end{aligned} \tag{8}$$

2.3. Nonlinear degradation modeling

In general, the degradation rate is gradually increasing with the increase of system degradation level. In degradation modeling based on cumulative damage model, the degradation rate is determined by damage value of per shock and shock counting per unit time. In other words, any change of shock damage or shock counting, the system degradation rate is affected. Therefore, both shock damage and shock counting should be considered in the transition of degradation rate.

2.3.1. Shock damage change

In order to simplify the calculation, it is assumed that the relationship between the damage values of two consecutive shocks is proportional. The shape-parameter of (i+1)th shock is q_k times as large as ith shock in stage M_k, namely $\alpha_k(t_{i+1}) = q_k \cdot \alpha_k(t_i)$ ($q_k > 0$, i=1,2,3,⋯).

The shape-parameter of the first shock in stage M_k is $\alpha_k(t_1) = \alpha_k$, and the (i+1)th shock is $\alpha_k(t_{i+1}) = (q_k)^i \cdot \alpha_k$. Therefore, when the shock counting in stage M_k is N_k, the equation (4) $(q_k \ne 1)$ becomes:

$$Y^k(t) = \sum_{i=1}^{N_k} x_i^k \sim Ga\left(\frac{\left(1 - (q_k)^{N_k}\right)\alpha_k}{1 - q_k}, \beta_k \right) \tag{9}$$

When $q_k > 1$, the shock damage shows increasing trend; when $q_k < 1$, the shock damage shows decreasing trend; when q_k=1, system degradation process is linear, and:

$$Y^k(t) = \sum_{i=1}^{N_k} x_i^k \sim Ga\left(N_k \alpha_k, \beta_k\right) \tag{10}$$

2.3.2. Shock counting change

Shock counting per unit time is associated with system working time. Hence, the probability of shock counting is best related to working time. In this paper, it is assumed that Poisson parameter $\lambda_k(t)$ is variable function and shows as follows:

$$\lambda_k(t) = \lambda_k \eta_k t^{\eta_k - 1} \quad (\lambda_k, \eta_k > 0) \tag{11}$$

Therefore, the probability equations (2) and (3) of shock counting become:

$$P(N_k = m) = \frac{\left(\lambda_k t^{\eta_k}\right)^m}{m!} \cdot e^{-\lambda_k t^{\eta_k}} \tag{12}$$

$$P(N_1 = m, N_2 = n) = \frac{\left(\lambda_1 t^{\eta_1}\right)^m}{m!} \cdot \frac{\left(\lambda_2 (t - t_c)^{\eta_2}\right)^n}{n!} \cdot e^{-\lambda_1 t^{\eta_1} - \lambda_2 (t - t_c)^{\eta_2}} \tag{13}$$

It is similar to Weibull process, when $0 < \eta_k < 1$, the time interval of two consecutive shocks shows increasing trend; when $\eta_k > 1$, the time interval of two consecutive shocks shows decreasing trend; when $\eta_k = 1$, the mean time interval of two consecutive shocks are fixed, and the shock process is a homogeneous Poisson process.

3. Maintenance policy

Research of maintenance decision-making is one of focuses for two-stage degradation modeling. As CBM is an effective maintenance policy method for various systems, CBM policy is chose to monitor considered system for the purpose of reducing maintenance cost. In the framework of this study, there are three possible maintenance actions are considered, inspection, preventive maintenance and corrective maintenance, respectively.

3.1. Adaptive maintenance policy

According to the characteristic that degradation rate is diverse in different degradation stage for two-stage degradation system, Saas-

souh et al. [5, 7, 19] put forward adaptive maintenance policy. And this maintenance policy has been proved to be useful.

In adaptive maintenance policy, the alarm threshold (Y_A) and interval of inspection (ΔT) are defined as follows:

$$Y_A = Y_1 \cdot \mathrm{I}_{\{t \le t_c\}} + Y_2 \cdot \mathrm{I}_{\{t > t_c\}} \tag{14}$$

$$\Delta T = \Delta T_1 \cdot \mathrm{I}_{\{t \le t_c\}} + \Delta T_2 \cdot \mathrm{I}_{\{t > t_c\}} \tag{15}$$

Set Y_k as the alarm threshold and ΔT_k as the interval of inspection for degradation stage M_k. After the ith inspection (the inspection time is T_i) in degradation stage M_k, the possible maintenance actions which can put into practice are defined as follows:

- If $Y(T_i)<Y_k$, do nothing and system is left as it is until next inspection time $T_{i+1}=T_i+\Delta T_k$.
- If $Y_k \le Y(T_i)<Y_f$, system is serious deteriorated and needs to be preventively repaired.
- If $Y(T_i)\ge Y_f$, system is fault and needs to be correctively repaired.

As the degradation rate in the second stage M_2 is greater than the first stage M_1, the parameters of adaptive maintenance policy have the following relationship: $Y_1 > Y_2$, $\Delta T_1 > \Delta T_2$. The rule of adaptive maintenance policy is illustrated in Fig. 2.

3.2. Time-dependent maintenance policy

As the degradation rate is faster and faster in nonlinear degradation process, the interval of inspection should be shorter and shorter in terms of working time. The maintenance decision-making method that the interval of inspection changes over time is called time-dependent maintenance policy in this paper. In order to facilitate engineering application, the (i+1)th interval of inspection is r times than ith, namely $\Delta T^{i+1}=r\cdot\Delta T^i$ and $r<1$.

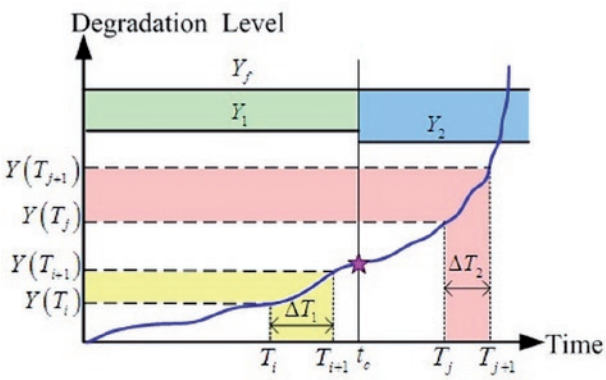


Fig. 2. Adaptive maintenance policy

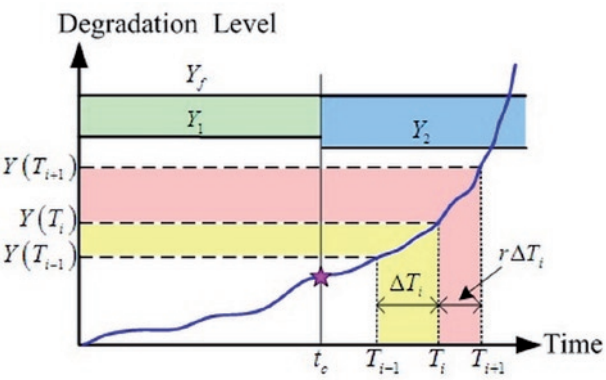


Fig. 3 Time-dependent maintenance policy

The rule of time-dependent maintenance policy (alarm threshold, possible maintenance action) is similar to adaptive maintenance policy, the only difference is that the interval of inspection $\Delta T^{i+1}=r\cdot\Delta T^i$. The rule of time-dependent maintenance policy is illustrated in Fig. 3.

3.3. Maintenance policy evaluation

3.3.1. Evaluation method

Maintenance cost occurs when a maintenance action is performed. The mean maintenance cost rate over an infinite time span is used to evaluate maintenance policy in this study. System is perfectly monitored through periodic monitor, and system state restores to be as good as new after a preventive/corrective maintenance action with negligible time. Mean maintenance cost rate can be obtained by renewal reward theory [20] as follows:

$$E(C) = \lim_{t\to\infty} \frac{E[C(t)]}{t} = \frac{E[C(T)]}{E[T]} \tag{16}$$

Where $C(t)$ is the total maintenance cost during time [0, t], T is the mean time length of a renewal cycle.

The total maintenance cost in a renewal cycle T can be written as:

$$E[C(T)] = C_I E[N_I(T)] + C_P P_P + C_C P_C \tag{17}$$

The mean time length of a renewal cycle T can be expressed as

$$E[T] = P_P T_P + P_C T_f \tag{18}$$

3.3.2. The probability of corrective maintenance

According to the rule of maintenance policy, system is considered as failure if any one of the following events (A_{C1}, A_{C2}, A_{C3}) occurs. In other words, system needs to be correctively repaired and it will cause corrective maintenance cost C_C. Take the event A_{C1} as a example, system degradation process is in stage M_1 ($T_z<T_{z+1}\le t_c$), if the degradation level $Y(T_z)<Y_1$ for zth inspection and $Y(T_{z+1})>Y_f$ for (z+1)th inspection, corrective maintenance action will be performed.

$$A_{C1} = \{Y(T_z) < Y_1 \cap Y(T_{z+1}) \ge Y_f \cap T_z < T_{z+1} \le t_c\}$$

$$A_{C2} = \{Y(T_z) < Y_2 \cap Y(T_{z+1}) \ge Y_f \cap t_c < T_z < T_{z+1}\}$$

$$A_{C3} = \{Y(T_z) < Y_1 \cap Y(T_{z+1}) \ge Y_f \cap T_z < t_c \le T_{z+1}\}$$

The probability for a corrective maintenance in a renewal cycle is the sums of probabilities for A_{C1}, A_{C2}, A_{C3}. It is written as:

$$P_C = P(A_{C1}) + P(A_{C2}) + P(A_{C3}) \tag{19}$$

3.3.3. The probability of preventive maintenance

It is considered that system needs to be preventively repaired if any one of the following events (A_{P1}, A_{P2}, A_{P3}) occurs, and it will cause preventive maintenance cost C_P.

$$A_{P1} = \left\{ Y(T_z) < Y_1 \bigcap Y_1 \le Y(T_{z+1}) < Y_f \bigcap T_z < T_{z+1} \le t_c \right\}$$

$$A_{P2} = \left\{ Y(T_z) < Y_2 \bigcap Y_2 \le Y(T_{z+1}) < Y_f \bigcap t_c < T_z < T_{z+1} \right\}$$

$$A_{P3} = \left\{ Y(T_z) < Y_1 \bigcap Y_2 \le Y(T_{z+1}) < Y_f \bigcap T_z < t_c \le T_{z+1} \right\}$$

The probability for a preventive maintenance in a renewal cycle is expressed as:

$$P_P = P(A_{P1}) + P(A_{P2}) + P(A_{P3}) \tag{20}$$

3.3.4. Continuous monitoring events

The system is left until next inspection time if any one of the following events (A_{I1}, A_{I2}) occurs, and it will cause monitoring cost C_I.

$$A_{I1} = \left\{ Y(T_z) < Y_1 \bigcap T_z \le t_c \right\}$$

$$A_{I2} = \left\{ Y(T_z) < Y_2 \bigcap T_z > t_c \right\}$$

The probability for system left until next inspection in a renewal cycle can be written as:

$$P_I = P(A_{I1}) + P(A_{I2}) \tag{21}$$

The mean number of times of inspection actions in a renewal cycle T is:

$$E\left[N_I(T) \right] = \sum_{z=1}^{\infty} z P_I \tag{22}$$

3.3.5. Mean time length of a renewal cycle

As formula (18) shown, the mean time length of a renewal cycle is determined by lifetime length T_f when system ends with corrective maintenance and mean working time length T_P when system ends with preventive maintenance. If the degradation level $Y(t)$ exceeds failure threshold Y_f, the system is considered as failed and will not work any time. That is to say, the lifetime length T_f is the time interval for $Y(t)$ from initial value 0 to Y_f. However, the mean working time length T_P does not mean that system cannot work. It is just shown that if a preventive maintenance action performed is better for system in inspection time T_z. Therefore, the system working time length when system ends with preventive maintenance is T_z.

4. Numerical example

This section aims to present some characteristics of two-stage degradation system: (a) In order to find the optimal maintenance policy for two-stage degradation system, mean cost rates of different maintenance policy are compared. (b) For the purpose of improving the understanding in two-stage degradation system, the influences of different parameter in degradation modeling are analyzed. The following numerical evaluations of the maintenance cost rate for two-stage degradation system are obtained from Monte Carlo simulations.

4.1. Choice of parameters values

In this paper, the considered two-stage degradation system has the following features: The degradation process is linear and mean degradation rate is stationary in the first stage M_1, the model parameters are $\alpha_1(t_{i+1}) = \alpha_1 = 1, \beta_1 = 1, \lambda_1 = 1, \eta_1 = 1, q_1 = 1$. The degradation process is nonlinear and mean degradation rate is change over time in the second stage M_2, the model parameters are $\alpha_2(t_1) = \alpha_2 = 1,\ \alpha_2(t_{i+1}) = q_2 \cdot \alpha_2(t_i) = q_2^{\,i} \cdot \alpha_2,\ \beta_2 = 1,\ \lambda_2 = 1$, and in order to present different nonlinear degradation process η_2, q_2 ($\eta_2, q_2 > 1$) will been evaluated as the need of studying.

The failure threshold Y_f is chosen in considering with the intrinsic properties of a two-stage degradation system. It is considered that Y_f=200 in this study. Meanwhile, in order to ensure the optimal result of mean cost rate $E(C)$ for maintenance policy is creditable, the unit costs are evaluated as other literatures [5, 7, 17, 19], so C_I=5, C_P=50, C_C=100.

Because the distribution of change-point t_c is affected by many factors, it is difficult to determine the PDF of t_c. In this study, the t_c PDF $g_c(t)$ is assumed to follow uniform distribution for the convenience of calculation. In order to analyze the influence of t_c, different uniform distribution of t_c are considered:

- Whole change-point distribution: $t_c \sim U(1,120)$.
- Early change-point distribution: $t_c \sim U(1,60)$.
- Middle change-point distribution: $t_c \sim U(30,90)$.
- Late change-point distribution: $t_c \sim U(60,120)$.

The upper bound value of the uniform distribution is evaluated as 120, it is considered that system fault occurs mostly in the second degradation stage M_2 on this occasion. Early and late change-point distributions present the first and second half of whole change-point distribution, respectively.

4.2. Influence of maintenance policy

The degradation level monitoring method for different maintenance policy is different, which includes alarm threshold and interval of inspection. Meanwhile, the mean maintenance cost rate is impacted by monitoring method. The method for obtaining optimal parameters and minimum mean cost rate of maintenance policy has been mentioned in some literatures [10,17]. The optimal parameters of main-

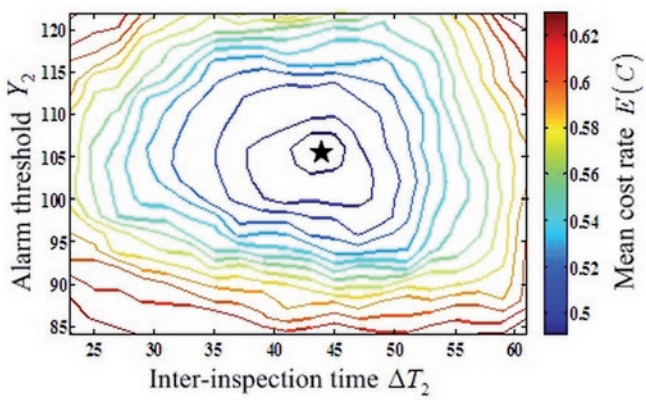


Fig. 4. Mean cost rate E(C) when Y_1=126, ΔT_1=71, t_c~U(30,90)

Table 1. Influence of maintenance policy and tc when shock damage change ($\eta_2 = 1, q_2 = 1.01$)

Change-point	Maintenance policy	Optimal parameters	Mean cost rate	Impact
$t_c \sim U(1,120)$	Adaptive	$Y_1 = 126, \Delta T_1 = 76$ $Y_2 = 109, \Delta T_2 = 38$	$E_1(C) = 0.4827$	
	Time-dependent	$Y_1 = 122, \Delta T_1 = 74$ $Y_2 = 102, r = 0.66$	$E_2(C) = 0.4608$	0.0219 (4.54%)
$t_c \sim U(1,60)$	Adaptive	$Y_1 = 128, \Delta T_1 = 73$ $Y_2 = 104, \Delta T_2 = 34$	$E_1(C) = 0.5478$	
	Time-dependent	$Y_1 = 122, \Delta T_1 = 71$ $Y_2 = 102, r = 0.60$	$E_2(C) = 0.5341$	0.0137 (2.50%)
$t_c \sim U(30,90)$	Adaptive	$Y_1 = 126, \Delta T_1 = 71$ $Y_2 = 103, \Delta T_2 = 44$	$E_1(C) = 0.4803$	
	Time-dependent	$Y_1 = 125, \Delta T_1 = 72$ $Y_2 = 105, r = 0.69$	$E_2(C) = 0.4622$	0.0181 (3.77%)
$t_c \sim U(60,120)$	Adaptive	$Y_1 = 126, \Delta T_1 = 74$ $Y_2 = 107, \Delta T_2 = 51$	$E_1(C) = 0.4421$	
	Time-dependent	$Y_1 = 128, \Delta T_1 = 74$ $Y_2 = 108, r = 0.78$	$E_2(C) = 0.4204$	0.0217 (4.91%)

Table 2. Influence of maintenance policy and tc when shock counting change ($q_2 = 1, \eta_2 = 1.1$)

Change-point	Maintenance policy	Optimal parameters	Mean cost rate	Impact
$t_c \sim U(1,120)$	Adaptive	$Y_1 = 128, \Delta T_1 = 73$ $Y_2 = 114, \Delta T_2 = 53$	$E_1(C) = 0.4763$	
	Time-dependent	$Y_1 = 131, \Delta T_1 = 73$ $Y_2 = 111, r = 0.69$	$E_2(C) = 0.4522$	0.0241 (5.06%)
$t_c \sim U(1,60)$	Adaptive	$Y_1 = 131, \Delta T_1 = 75$ $Y_2 = 113, \Delta T_2 = 45$	$E_1(C) = 0.5182$	
	Time-dependent	$Y_1 = 134, \Delta T_1 = 72$ $Y_2 = 110, r = 0.63$	$E_2(C) = 0.5047$	0.0135 (2.61%)
$t_c \sim U(30,90)$	Adaptive	$Y_1 = 126, \Delta T_1 = 72$ $Y_2 = 116, \Delta T_2 = 55$	$E_1(C) = 0.4751$	
	Time-dependent	$Y_1 = 130, \Delta T_1 = 69$ $Y_2 = 111, r = 0.81$	$E_2(C) = 0.4544$	0.0207 (4.36%)
$t_c \sim U(60,120)$	Adaptive	$Y_1 = 129, \Delta T_1 = 75$ $Y_2 = 114, \Delta T_2 = 60$	$E_1(C) = 0.4269$	
	Time-dependent	$Y_1 = 129, \Delta T_1 = 75$ $Y_2 = 108, r = 0.84$	$E_2(C) = 0.4024$	0.0245 (5.74%)

tenance policy can be achieved after simulations. That is to say, the mean maintenance cost rate $E(C)$ can be found under simulations with different alarm threshold and interval of inspection. Take adaptive maintenance policy as an example, as maintenance policy evaluation method studies in section 3.3, the minimum mean cost rate is $E(C)$=0.4803 when Y_1=126, Y_2=103, ΔT_1=71, ΔT_2=44 and $t_c \sim U(30,90)$, as shown in contour map Fig. 4 ($E(C)$ are equal in the same contour). All the optimal parameters of maintenance policy under different cases can be achieved by a similar way.

As adaptive maintenance policy has been proved to be effective by Ponchet [19], taking the mean cost rate of adaptive maintenance policy as a basis of comparison. For instance (as Tab. 1 shown), when $t_c \sim U(30,90)$, $\eta_2 = 1, q_2 = 1.01$, the minimum mean cost rate of adaptive maintenance policy is $E_1(C)$=0.4803, and the minimum mean cost rate of time-dependent maintenance policy is $E_2(C)$=0.4622. $E_2(C)$ have a decrease of 0.0181 compares to $E_1(C)$, so that the optimal rate is 0.0181/0.4803=3.77%.

Nonlinear degradation process can be expressed by shock damage change and shock counting change. As shown in Tab. 1 ($\eta_2 = 1, q_2 = 1.01$) and Tab. 2 ($t_c \sim U(1,120)$), it is obvious that the mean cost rate of time-dependent maintenance policy is smaller than adaptive maintenance policy. In other words, time-dependent maintenance policy is better than adaptive maintenance policy for given system. Because degradation rate in stage M_2 for given system is nonlinear and becomes faster and faster over working time, and the interval of inspection in stage M_2 for time-dependent maintenance policy ΔT_2 is shorter and shorter as inspection time goes on. But the interval of inspection for adaptive maintenance policy ΔT_2 is fixed and never changed in stage M_2.

4.3. Influence of change-point distribution

The time distribution of change-point t_c can influence the choice of system maintenance policy. Hence, the influence of change-point distribution is studied under both maintenance policies. The optimal results under different t_c distribution for shock damage change and shock counting change are shown in Tab. 1 and Tab. 2, respectively. Taking Tab. 1 as an example, the analyzed results can be achieved as follows:

(a) Although the t_c distribution and maintenance policy are different, the variety of alarm thresholds (Y_1) and intervals of inspection (ΔT_1) for the first degradation stage M_1 is very small. The change of alarm thresholds (Y_2) for the second degradation stage M_2 is also very small, but the variety of intervals of inspection (ΔT_2) is great. It is because that mean system lifetime is change corresponding to different t_c distribution, the earlier change-point t_c occurs, the shorter mean system lifetime.

(b) When t_c fall in time interval (1,60), (30,90) and (60,120), the mean cost rates for adaptive maintenance policy are 0.5478, 0.4803 and 0.4421, respectively. It means that the mean maintenance cost rate is drop off with the increase of average time of t_c distribution. This feature also conforms to time-dependent maintenance policy.

Table 3. Influence of shock damage change ($\eta_2 = 1,\ t_c \sim U(30,90)$)

Model parameter	Maintenance policy	Optimal parameters	Mean cost rate	Impact
$q_2 = 1.000$	Adaptive	$Y_1 = 124, \Delta T_1 = 75$ $Y_2 = 124, \Delta T_2 = 75$	$E_1(C) = 0.3750$	
	Time-dependent	$Y_1 = 126, \Delta T_1 = 85$ $Y_2 = 126, r = 0.80$	$E_2(C) = 0.3711$	0.0039 (1.04%)
$q_2 = 1.005$	Adaptive	$Y_1 = 126, \Delta T_1 = 74$ $Y_2 = 115, \Delta T_2 = 62$	$E_1(C) = 0.4270$	
	Time-dependent	$Y_1 = 122, \Delta T_1 = 74$ $Y_2 = 110, r = 0.84$	$E_2(C) = 0.4141$	0.0129 (3.02%)
$q_2 = 1.010$	Adaptive	$Y_1 = 126, \Delta T_1 = 71$ $Y_2 = 103, \Delta T_2 = 44$	$E_1(C) = 0.4803$	
	Time-dependent	$Y_1 = 125, \Delta T_1 = 72$ $Y_2 = 105, r = 0.69$	$E_2(C) = 0.4622$	0.0181 (3.77%)
$q_2 = 1.015$	Adaptive	$Y_1 = 124, \Delta T_1 = 68$ $Y_2 = 104, \Delta T_2 = 39$	$E_1(C) = 0.5207$	
	Time-dependent	$Y_1 = 122, \Delta T_1 = 71$ $Y_2 = 103, r = 0.63$	$E_2(C) = 0.4970$	0.0237 (4.55%)
$q_2 = 1.020$	Adaptive	$Y_1 = 123, \Delta T_1 = 68$ $Y_2 = 106, \Delta T_2 = 35$	$E_1(C) = 0.5569$	
	Time-dependent	$Y_1 = 128, \Delta T_1 = 68$ $Y_2 = 99, r = 0.60$	$E_2(C) = 0.5257$	0.0312 (5.60%)

(c) As shown in section 4.2, time-dependent maintenance policy is better than adaptive maintenance policy for given system. The optimal rates are 2.50%, 3.77% and 4.91% when t_c fall in time interval (1,60), (30,90) and (60,120), respectively. It is obvious that there is more interest in using a time-dependent maintenance policy instead of an adaptive maintenance policy when the change-point t_c occurs more later.

(d) When t_c fall in time interval (1,120) and (30,90), the optimal rates are 4.54% and 3.77%, respectively. The average time of both (1,120) and (30,90) equal to 60. It can be known that when the change-point t_c is defined on a larger time interval the more interest can be achieved in using a time-dependent maintenance policy instead of an adaptive maintenance policy.

4.4. Influence of nonlinear degradation process

As shown in Tab. 3 and Tab. 4, different model parameters η_2 and q_2 present different nonlinear degradation process. The larger of η_2 and q_2, the faster of degradation rate increase in the second stage M_2. When $\eta_1 = \eta_2 = 1$ and $q_1 = q_2 = 1$, the system mean degradation rate is stationary and never change, the degradation process is linear and single-stage.

The optimal results under different model parameters q_2 for shock damage change are shown in Tab. 3, and the optimal results under different model parameters η_2 for shock counting change are shown in Tab. 4. The optimal results are similar between Tab. 3 and Tab. 4. Taking Tab. 3 as an example, the analyzed results can be achieved as follows:

(a) Because degradation process is linear and degradation rate is stationary in the first stage M_1 for considered system, the variety of alarm thresholds (Y_1) and intervals of inspection (ΔT_1) for stage M_1 is very small although q_2 are different, especially for adaptive maintenance policy.

(b) As the growth of model parameter q_2, the alarm thresholds (Y_2) and intervals of inspection (ΔT_2) for stage M_2 become smaller and smaller. For example, q_2 equal to 1.000, 1.005, 1.010, 1.015 and 1.020, the intervals of inspection of adaptive maintenance policy are 75, 62, 44, 39 and 35, respectively.

(c) When q_2 equal to 1.000, 1.005, 1.010, 1.015 and 1.020, the mean cost rates for adaptive maintenance policy are 0.3750, 0.4270, 0.4803, 0.5207 and 0.5569, respectively. It means that the mean maintenance cost rate is going up with the increase of model parameter q_2. This feature also conforms to time-dependent maintenance policy.

(d) As seen previously, the optimal rates are 1.04%, 3.02%, 3.77%, 4.55% and 5.60% when q_2 equal to 1.000, 1.005, 1.010, 1.015 and 1.020, respectively. It is obviously that there is more interest in using a time-dependent maintenance policy instead of an

Table 4. Influence of shock counting change ($q_2 = 1,\ t_c \sim U(30,90)$)

Model parameter	Maintenance policy	Optimal parameters	Mean cost rate	Impact
$\eta_2 = 1$	Adaptive	$Y_1 = 126, \Delta T_1 = 76$ $Y_2 = 126, \Delta T_2 = 76$	$E_1(C) = 0.3776$	
	Time-dependent	$Y_1 = 124, \Delta T_1 = 76$ $Y_2 = 118, r = 0.87$	$E_2(C) = 0.3745$	0.0031 (0.82%)
$\eta_2 = 1.05$	Adaptive	$Y_1 = 126, \Delta T_1 = 77$ $Y_2 = 118, \Delta T_2 = 65$	$E_1(C) = 0.4118$	
	Time-dependent	$Y_1 = 126, \Delta T_1 = 78$ $Y_2 = 116, r = 0.73$	$E_2(C) = 0.4017$	0.0101 (2.45%)
$\eta_2 = 1.1$	Adaptive	$Y_1 = 126, \Delta T_1 = 72$ $Y_2 = 116, \Delta T_2 = 55$	$E_1(C) = 0.4751$	
	Time-dependent	$Y_1 = 130, \Delta T_1 = 69$ $Y_2 = 111, r = 0.81$	$E_2(C) = 0.4544$	0.0207 (4.36%)
$\eta_2 = 1.15$	Adaptive	$Y_1 = 126, \Delta T_1 = 71$ $Y_2 = 118, \Delta T_2 = 47$	$E_1(C) = 0.5241$	
	Time-dependent	$Y_1 = 126, \Delta T_1 = 66$ $Y_2 = 116, r = 0.71$	$E_2(C) = 0.4926$	0.0315 (6.01%)
$\eta_2 = 1.2$	Adaptive	$Y_1 = 124, \Delta T_1 = 75$ $Y_2 = 111, \Delta T_2 = 39$	$E_1(C) = 0.5615$	
	Time-dependent	$Y_1 = 128, \Delta T_1 = 63$ $Y_2 = 112, r = 0.59$	$E_2(C) = 0.5191$	0.0424 (7.55%)

adaptive maintenance policy when model parameter q_2 greater. Meanwhile, when q_2=1 the system mean degradation rate is fixed in whole degradation process, the alarm thresholds and inter-inspection times for both stage M_1 and M_2 are equal. In this case, the adaptive maintenance policy is the same to traditional condition-based maintenance policy (namely global maintenance policy [17]), and it is no necessary to use a time-dependent maintenance policy instead of an adaptive maintenance policy.

5. Conclusions

This paper takes into account degradation modeling and maintenance policy for a two-stage degradation system, which degradation process is nonlinear and degradation rate is change over time in both stages. The system degradation process is considered as step, and it is modeled based on cumulative damage model. The nonlinear degradation process is modeled by shock damage change and shock counting change. In order to explore optimal maintenance policy for considered system, two maintenance policies have been investigated and assessed through their mean maintenance cost rates.

Moreover, influence analysis of different model parameter and maintenance policy is studied in numerical examples, and results prove that: (a) It is necessary to consider monitoring method for considered system, optimal maintenance policy can help to reduce mean cost rate. (b) It is obvious that the mean maintenance cost rate and maintenance policy are impacted by change-point distribution, shock damage and shock counting.

Acknowledgements

This research report was partially supported by science fund for youths under grant number 71401173.

References

1. Bae SJ, Kuo W, Kvam PH. Degradation models and implied lifetime distributions. Reliability Engineering and System Safety 2007; 92:601-608, http://dx.doi.org/10.1016/j.ress.2006.02.002.
2. Bae SJ, Yuan T, Ning S, Kuo W. A Bayesian approach to modeling two-phase degradation using change-point regression. Reliability Engineering and System Safety 2015; 134:66-74, http://dx.doi.org/10.1016/j.ress.2014.10.009.
3. Chen N, Tsui KL. Condition monitoring and remaining useful life prediction using degradation signals: revisited. IIE Transactions 2013; 45(9):939-952, http://dx.doi.org/10.1080/0740817X.2012.706376.
4. Farid A, David P. Estimating arrival rate of nonhomogeneous Poisson processes with semidefinite programming. Ann Oper Res 2013; 208:291-308, http://dx.doi.org/10.1007/s10479-011-1020-2.

5. Fouladirad M, Grall A. A maintenance decision rule with embedded Bayesian online change detection for gradually deteriorating systems. Journal of Risk and Reliability 2008; 22(3):359-369.
6. Fouladirad M, Grall A. Condition-based maintenance for a system subject to a non-homogeneous wear process with a wear rate transition. Reliability Engineering and System Safety 2011; 96:611-618, http://dx.doi.org/10.1016/j.ress.2010.12.008.
7. Fouladirad M, Grall A, Dieulle L. On the use of on-line detection for maintenance of gradually deteriorating systems. Reliability Engineering and System Safety 2008; 93(12):1814-1820, http://dx.doi.org/10.1016/j.ress.2008.03.020.
8. Gebraeel NZ, Lawley MA, Li R, Ryan JK. Residual-life distributions from component degradation signals: a Bayesian approach. IIE Transactions 2005; 37(6):543-557, http://dx.doi.org/10.1080/07408170590929018.
9. Guo CM. Condition-based Maintenance Optimization with independent increments processes. National University of Defense Technology, Changsha, China, 2013,07.
10. Huynh KT, Barros A, Be´renguer C, Castro IT. A periodic inspection and replacement policy for systems subject to competing failure modes due to degradation and traumatic events. Reliability Engineering and System Safety 2011; 96:497-508, http://dx.doi.org/10.1016/j.ress.2010.12.018.
11. Jiang RY, Prabhakar Murthy DN. Maintenance decision models for management. Science Press, 2008,06.
12. Keedy E, Feng Q. A physics-of-failure based reliability and maintenance modeling framework for stent deployment and operation. Reliability Engineering and System Safety 2012; 103:94-101, http://dx.doi.org/10.1016/j.ress.2012.03.005.
13. Meier-Hirmer C, Sourget F, Roussignol M. Maintenance optimization for a system with a gamma deterioration process and intervention delay: application to track maintenance. Journal of Risk and Reliability 2009; 223(3):189-198, http://dx.doi.org/10.1243/1748006XJRR234.
14. Ni XL, Zhao JM, Wang GY, Teng HZ. Maintenance policy for two-stage deteriorating mode system based on cumulative damage model. Journal of Vibroengineering 2015; 17(3):1266-1285.
15. Noortwijk van JM. A survey of the application of gamma processes in maintenance. Reliability Engineering and System Safety 2009; 94:2-21, http://dx.doi.org/10.1016/j.ress.2007.03.019.
16. Peng W, Li YF, Yang YJ, Huang HZ, Zuo MJ. Inverse Gaussian process models for degradation analysis: a Bayesian perspective. Reliability Engineering and System Safety 2014; 130:175-189, http://dx.doi.org/10.1016/j.ress.2014.06.005.
17. Ponchet A, Fouladirad M, Grall A. Assessment of a maintenance model of a multi-deteriorating mode system. Reliability Engineering and System Safety 2010; 95(11):1244-1254, http://dx.doi.org/10.1016/j.ress.2010.06.021.
18. Qian CH, Nakamura S, Nakagawa T. Cumulative damage model with two kinds of shocks and its application to the backup policy. Journal of the Operations Research 1999; 42(4):501-511.
19. Saassouh B, Dieulle L, Grall A. Online maintenance policy for a deterioration system with random change of mode. Reliability Engineering and System Safety 2007; 92(12):1677-1685, http://dx.doi.org/10.1016/j.ress.2006.10.017.
20. Sheldon MR. Stochastic processes for insurance and finance, wiley series in probability and statistics. Johon Wiley & Sons, New York, 1996:1-639.
21. Song SL, Coit DW, Feng QM, Peng H. Reliability analysis for multi-component systems subject to multiple dependent competing failure process. IEEE Transactions on Reliability 2014; 63:331-345, http://dx.doi.org/ 10.1109/TR.2014.2299693.
22. Toshio N. Shock and damage models in reliability theory. Springer, 2006,06.
23. Wang X, Balakrishnan N, Guo B. Residual life estimation based on a generalized Wiener degradation process. Reliability Engineering and System Safety 2014; 124:13-23, http://dx.doi.org/10.1016/j.ress.2013.11.011.
24. Ye ZS, Tang LC, Xu HY. A distribution-based systems reliability model under extreme shocks and natural degradation. IEEE Transactions on Reliability 2011; 60(1):246-256, http://dx.doi.org/10.1109/TR.2010.2103710.
25. Zhao Z, Wang FL, Jia MX, Wang S. Predictive maintenance policy based on process data. Chemometrics and Intelligent Laboratory Systems 2010; 103:137-143, http://dx.doi.org/10.1016/j.chemolab.2010.06.009.

Xianglong NI
Jianmin ZHAO
Wenyuan SONG
Chiming GUO
Haiping LI
Mechanical Engineering College
Shijiazhuang, 050003, Hebei province, China

E-mails: ni_phm@163.com, jm_zhao@hotmail.com, phm_cbm@163.com, guochiming@nudt.edu.cn, hp_li@hotmail.com

Article citation info:
PRZYBYŁEK P, SIODŁA K. Application of capacitive sensor for measuring water content in electro-insulating liquids. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 181–185, http://dx.doi.org/10.17531/ein.2016.2.4.

Piotr PRZYBYŁEK
Krzysztof SIODŁA

APPLICATION OF CAPACITIVE SENSOR FOR MEASURING WATER CONTENT IN ELECTRO-INSULATING LIQUIDS

ZASTOSOWANIE CZUJNIKA POJEMNOŚCIOWEGO DO POMIARU ZAWARTOŚCI WODY W CIECZACH ELEKTROIZOLACYJNYCH*

The article discusses the problem of water content measurement in electro-insulating liquids using capacitive sensors. The article describes coefficients affecting reliability of the water content measurement. The authors discussed issues connected with water saturation limit in electro-insulating liquids. The authors also proposed a method which allows determining coefficients by means of which it is possible to calculate the water saturation limit in electro-insulating liquid as a function of temperature. Determining the coefficients allows proper calculating of the water content in ppm by weight by means of relative water saturation of the investigated liquid, what was measured with a capacitive probe. Propositions included in the article improve reliability of the method to determine water content in electro-insulating liquids and thus contribute to breakdown-free operation of electric power equipment insulated with these liquids.

Keywords: *capacitive sensor, water content measurement, electro-insulating liquid, transformer.*

W artykule omówiono problematykę pomiaru zawartości wody w cieczach elektroizolacyjnych przy wykorzystaniu czujników pojemnościowych. Opisano czynniki wpływające na wiarygodność pomiaru zawartości wody. Autorzy pracy omówili zagadnienia związane z granicznym nasyceniem cieczy elektroizolacyjnych wodą. W artykule zaproponowana została metoda umożliwiająca wyznaczenie współczynników, za pomocą których możliwe jest obliczenie granicznego nasycenia cieczy elektroizolacyjnej wodą w funkcji temperatury. Wyznaczenie współczynników umożliwia poprawne obliczenie zawartości wody w ppm wagowo za pomocą zmierzonego sondą pojemnościową względnego nasycenia badanej cieczy wodą. Propozycje zawarte w artykule poprawiają niezawodność metody wyznaczania zawartości wody w cieczach elektroizolacyjnych, a przez to przyczyniają się do bezawaryjnej eksploatacji urządzeń elektroenergetycznych izolowanych tymi cieczami.

Słowa kluczowe: *czujnik pojemnościowy, pomiar zawartości wody, ciecz elektroizolacyjna, transformator.*

1. Introduction

The presence of water in the electro-insulating system of power devices is a serious operational problem. This problem refers first of all to devices insulated with cellulose materials impregnated with electro-insulating liquids [4, 5]. Examples of such devices are power transformers and instrument transformers. This problem also refers to transformer bushing insulators.

With passing years of operation of a given device, its insulation moisture rises. This problem has been discussed in different scientific publications many times [2, 4, 15], mainly in terms of oil-paper insulation moisture of power transformers. Authors of these articles notice that water is not only a decomposition product of cellulose insulation but it also contributes to this decomposition. As a result of water presence in the insulating system and exposure to high temperature, the process of cellulose depolimerization takes place, which results in a decrease of its mechanical strength [1, 12, 17]. Other disadvantageous consequences of water presence in the insulating system are: electric strength drop of the insulation, lowering the inception voltage of partial discharges [19], and the threat of the bubble effect appearance [8, 11], which increases with more moistened cellulose insulation. High insulation moisture forces restrictions of the device load. One of positive aspects of water presence in solid insulation is the increase of its thermal conductivity, which to a certain extent improves cooling of the transformer windings [7]. However, negative results of water presence are predominant and there is a trend to reduce the insulation moisture of power devices as much as possible.

Due to long-lasting operation of power devices, the problem of moisture is serious and concerns not only the Polish electric power system but also the systems of most countries. The problem of insulation moisture has been known to the international community for a few decades and it is still a current issue, which is confirmed by the newest publication of the worldwide range: they concern, among others, moisture measurement methods of transformer insulation [6, 13, 20], also problems connected with forms of water presence [3, 13, 21] and its migration in the cellulose−electro-insulating liquid system [2, 15]. Nowadays, the activities of the international CIGRE Working Group WG D1.52 *Moisture measurement in insulating fluids and transformer insulation* are in progress. Their purpose is, among others, to improve reliability of water content measurement methods in electro-insulating liquids.

2. Capacitive probe for measuring water content in electro-insulating liquid

The most frequently used method of measuring water content in electro-insulating liquids is titration method which makes use of the Karl Fischer reaction (KFT – Karl Fischer Titration). It is a standardized method [10] which is characterized by high accuracy of water content measurement. The water content measured by means of the

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

KFT method is expressed in ppm by weight. In order to do the measurement of water content it is necessary to take a liquid sample from the investigated device.

In recent years strong emphasis has been put on monitoring quantities which make evaluation of the device state possible, particularly power transformers of strategic importance in power system. The devices available on the market (Fig. 1), equipped with capacitive sensors allow moisture monitoring of the electro-insulating liquid, which certainly is their great advantage.

Fig. 1. Devices applied for monitoring moisture and temperature of the electro-insulating liquid, the three devices on the right side of the image can also measure some selected gases dissolved in the liquid

Figure 2 presents an example of monitoring mineral oil moisture in 160 MVA autotransformer.

It should be stressed, however, that incorrect operation of the capacitive probes can lead to wrong diagnoses concerning the state of the electric power device.

The probes mentioned before, apart from capacitive sensors, are also equipped with temperature sensors. Relative saturation (*RS*) of the investigated liquid is determined by means of a capacitive sensor, which is expressed by means of the following equation:

$$RS = \frac{WCL}{S} \cdot 100, \tag{1}$$

where *WCL* means the water content in the liquid expressed in ppm by weight, whereas *S* means water saturation limit in the liquid, which is also expressed in ppm by weight.

Water saturation limit in the liquid should be understood as the maximum amount of water which can be dissolved in it at a given temperature. Exceeding water saturation limit in the liquid results in appearing of dispersed water. Water saturation limit depends not only on temperature but also on many other factors such as:

- kind of liquid: synthetic esters (S=1758 ppm for 20°C), natural esters (S=858 ppm for 20°C), and silicon oils (S=169 ppm for 20°C) have much higher water solubility than mineral oils (S=47 ppm for 20°C),

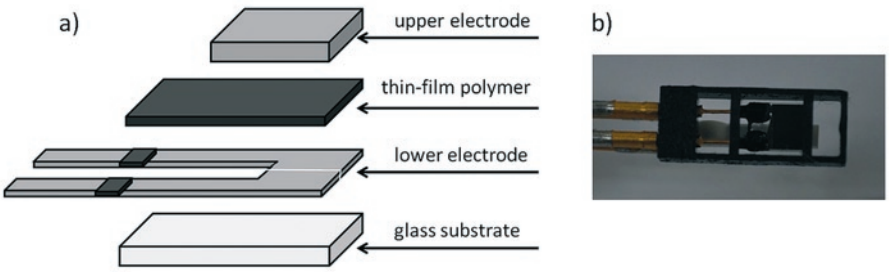


Fig. 3. Capacitive sensor: a) construction, b) photograph of the sensor [13]

Fig. 2. Monitoring system of water and gas content in oil in 160 MVA autotransformer

- liquid composition: e.g. water solubility in mineral oil rises with an increase of the aromatic fraction content,
- ageing degree: aged liquids with the higher neutralization value, have polar ageing products, which affect water solubility increase.

Relative saturation of the liquid is measured indirectly by means of a capacitive sensor which is proportional to the sensor's capacity that is the main element of the capacitive probe. The construction of the sensor is presented in Figure 3.

The capacitive sensor consists of two electrodes between which a thin-film of hygroscopic polymer is placed. The electrodes are made in such a way so as to allow contact of the polymer with the investigated insulating liquid. Water particles present in the insulating liquid penetrate to the hygroscopic polymer, in the amount depending on relative humidity of the environment in which the sensor is immersed. Water content increase in the polymer causes increase of its electric permittivity, which in turn is connected with increase of the sensor's electric capacitance [13].

The principle of the sensor's work described above allows correct measurement of relative saturation of electro-insulating liquids. The problem appears in a situation when it is required to know the water content expressed in ppm by weight, which is enforced (not necessarily fully reasonably) by standards and operation manuals of electric power equipment [9, 16]. There is a possibility to calculate water concentration *WCL* in the liquid in ppm by weight on the basis of Equation (1). In order to do it, it is necessary to calculate water saturation limit in the liquid using the Arrhenius equation:

$$\log S(T) = A - \frac{B}{T}, \tag{2}$$

where A and B are coefficients characteristic for a given insulating liquid, whereas T means temperature expressed in K.

Calculating water saturation limit in the liquid is possible when coefficients A and B are known. Unfortunately, these coefficients are dependent on numerous, mentioned above factors which affect water solubility in the electro-insulating liquid. The capacitive probes fixed in the devices have a built-in algorithm which allows calculating water content in the electro-insulating liquid in ppm by weight, however, for the calculations, they most often make use of coefficients A and B, which were determined previously for a given new mineral oil. ***The water content measurement in ppm by weight using the capacitive probe, in liquids of different water solubility than new mineral oil, can be significantly undervalued. This is why it is necessary to determine coefficients A and B for liquid taken from investigated, by means of capacitive probe, electric power device.***

3. Determining coefficients A and B

3.1. Method description

Within CIGRE Working Group WG D1.52 (*CIGRE − Conseil International des Grands Réseaux Électriques*) there are activities whose goal is to propose methods allowing reliable determination of coefficients A and B, which describe water saturation limit in electro-insulating liquids. Below is described one of the methods proposed to submit in the CIGRE brochure. The proposed method was created, among others, on the basis of experiences presented in [3, 6, 13, 18].

This method consists in conditioning of the electro-insulating liquid in a tightly closed vessel at three different temperature values. After achievement of the state of moisture equilibrium, for each temperature level the following oil parameters are measured: water content, relative saturation of the liquid and temperature. Coefficients A and B are determined on the basis of the data obtained in above described way.

3.2. Measurement system

The measurement system (Fig. 4) consists of a glass vessel with volume of about one litre, filled with electro-insulating liquid. The vessel should be tightly closed by means of a lid made of polytetrafluoroethylene (PTFE). Instead of glass and PTFE it is possible to use the other materials which are characterized with low hygroscopicity and do not react chemically with the investigated liquid.

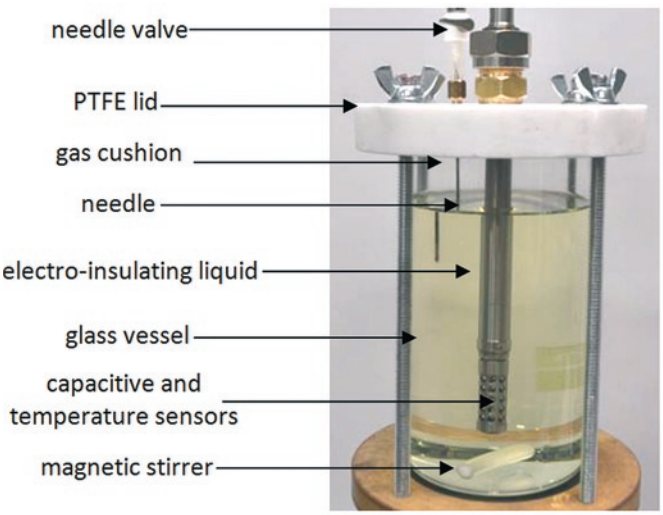


Fig. 4. Measurement system for determining coefficients A and B [14]

There are two slots in the lid. The first of them allows fixing the probe with the capacitive and temperature sensors. The other slot is for fixing the needle with the valve. A sample of the liquid is taken with the needle to measure water content by means of the Karl Fischer Titration method. In time of conditioning process the investigated liquid is mixed with the magnetic stirrer. The capacitive and temperature sensors are placed directly above the magnetic stirrer.

3.3. Measurement procedure

Oil conditioning is carried out for three temperature levels selected in the range from 20 to 60°C, where the difference between the successive temperature levels should be at least 10°C. In turn, relative saturation of the liquid in the mentioned above temperature range should be within the range from 15% (for a high value of liquid temperature) to 75% (for a low value of liquid temperature). Because of this it is necessary to do preliminary liquid moistening to the level of about 70% of relative saturation at the temperature of 20°C. The conditioning of the vessel with oil should take place in the climatic chamber whose task is to enforce a proper temperature of the liquid. Moreover, it is recommended that after stabilizing a preset oil temperature value, the value of air relative humidity in the chamber should be set at the same level as relative saturation of the investigated liquid. This is a way to prevent water migration if leaks occur in the measurement system. After stabilizing the temperature, the needle valve should be opened for a very short time in order to rebalance the air pressure in the vessel with atmospheric pressure.

The time of liquid conditioning for each temperature level should be long enough, so that the system could reach the equilibrium state of moisture and temperature.

For each temperature level n, after reaching the equilibrium state, the value of relative water saturation RS_n and temperature T_n of the investigated liquid should be recorded. Moreover, a liquid sample should be taken and water content WCL_n should be measured using the Karl Fischer Titration method.

Using the obtained data sets (RS_n, T_n, WCL_n) and Equation (3) it is possible to determine coefficients A and B. Equation 3 was obtained from Equations (1) and (2).

$$\log\left(\frac{WCL_n}{RS_n}\cdot 100\right) = A - \frac{B}{T_n}. \tag{3}$$

3.4. Research results – example of method application

Figure 5 presents an application of the described above procedure for determining coefficients A and B for new mineral oil. The conditioning time was equal to 24 h for each temperature level.

On the basis of the research results presented in Figure 5 and making use of Equation (3), coefficients A=7.288 and B=1646.897 were determined for new mineral oil. This experiment was repeated twice, and the obtained results are listed in Table 1 and in Figure 6.

On the basis of the experiment results presented in Table 1 and in Figure 6, it can be find high reproducibility of the results obtained by means of the proposed method.

For all measurements in Table 1, the authors determined expanded uncertainty $U(S)$ for the confidence level k=2. The following equation was used for the calculations:

$$U(S_n) = k\cdot u_c(S_n) = k\cdot\sqrt{\left[\frac{\partial\left(\frac{WCL_n}{RS_n}\cdot 100\right)}{\partial WCL_n}\cdot u(WCL)\right]^2 + \left[\frac{\partial\left(\frac{WCL_n}{RS_n}\cdot 100\right)}{\partial RS_n}\cdot u(RS)\right]^2}. \tag{4}$$

Table 1. Comparison of coefficients A and B, and water saturation limit S in mineral oil, calculated on their basis

Temperature [°C]	0	10	20	30	40	50	60	70	80	90	100
Measurement 1 – A=7.288,B=1646.897 S±U(S) [ppm]	18±2	30±3	47±4	72±7	107±10	155±15	221±21	308±30	421±42	566±57	749±76
Measurement 2 – A=7.246,B=1635.942 S±U(S) [ppm]	18±2	29±3	46±4	71±7	105±10	153±15	216±21	301±30	411±42	551±57	727±76
Measurement 3 – A=7.342,B=1662.676 S±U(S) [ppm]	18±2	30±3	47±4	72±7	108±10	157±15	224±22	314±31	430±43	580±58	769±78

U(S) – expanded uncertainty

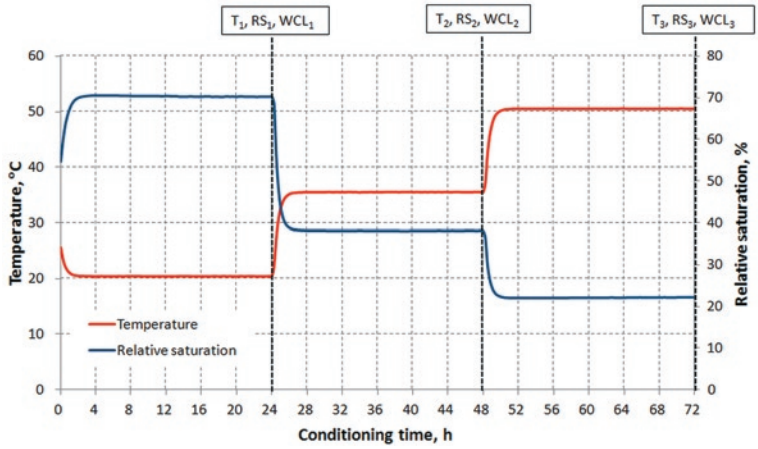


Fig. 5. Procedure for determining coefficients A and B; results obtained for new mineral oil: T_1=20.47°C, RS_1=70.21%, WCL_1=33.6 ppm; T_2=35.54°C, RS_2=38.17%, WCL_2=34.2 ppm; T_3=50.59°C, RS_3=22.25%, WCL_3=35.4 ppm

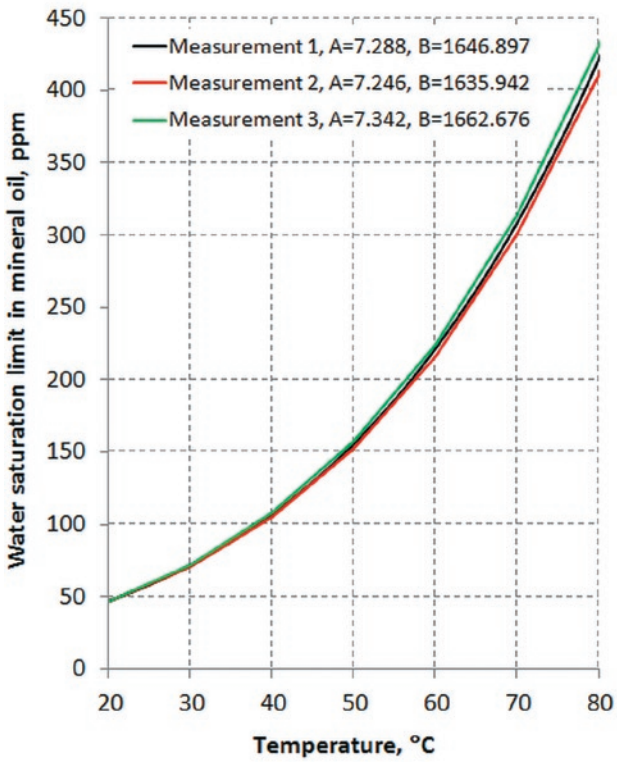


Fig. 6. Comparison of measurement results of water saturation limit in new mineral oil as a function of temperature

In order to determine measurement uncertainty of water saturation limit in mineral oil, standard uncertainty of the water content measurement using the KFT method equal to u(WCL)=1,5 ppm as well as standard uncertainty of the measurement of relative saturation of oil equal to u(RS)=0,5% was assumed. The uncertainty range for measurement 1 of water saturation limit in mineral oil is presented in Figure 7.

It should be noted that reliability of the coefficients A and B allowing description of water saturation limit in electro-insulating liquid as a function of temperature is highly affected by accuracy of the water content measurement using the Karl Fischer Titration method. Even a slight measurement mistake can cause a significant error of the determined water solubility, particularly in the range of high temperature values.

4. Conclusions

The authors propose a method for determining coefficients A and B which allow description of water solubility in the electro-insulating liquid as a function of temperature. The knowledge of water solubility as a function of temperature is needed when it is necessary to recalculate relative saturation of the investigated liquid, measured by means of the capacitive sensor, into water content expressed in ppm by weight.

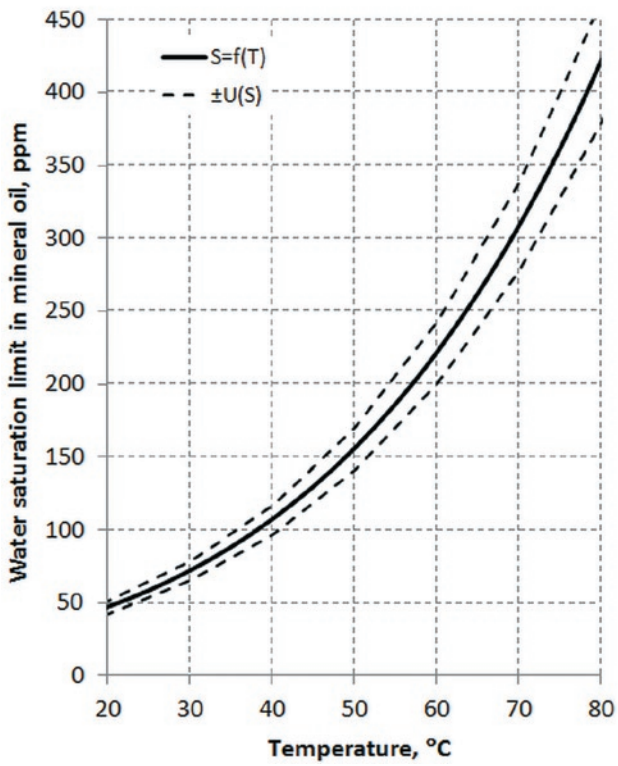


Fig. 7. Uncertainty range of water saturation limit in mineral oil: S=f(T) – result of water saturation limit in oil for measurement 1 (Table 1), U(S) – expanded uncertainty

The article presents a procedure for determining coefficients using new mineral oil as an example. However, it should be emphasized that determined coefficient values are correct only for the mineral oil investigated within the experiment. Their application for other electro-insulating liquids, particularly aged mineral oils, silicon oils and esters will make results of water content in ppm by weight significantly undervalued.

The research done so far allow to claim that the presented method for determining coefficients, which are used for description of water solubility as a function of temperature, can be applied for different electro-insulating liquids, such as: mineral oils, silicon oils, synthetic and natural esters. The authors found high reproducibility of results obtained using the above described method.

Acknowledgement

The research was financed from resources of the Ministry of Science and Higher Education for statutory activities No 04/41/DS-PB/4181, name of the task: Water solubility in electro-insulating liquids – measurement procedure, determining coefficients which allow calculating water solubility in mineral oil as a function of temperature.

References

1. Cigré Brochure 323, Ageing of cellulose in mineral-oil insulated transformers, 2007.
2. Cigré Brochure 349, Moisture equilibrium and moisture migration within transformer insulation systems, 2008.
3. Fofana I., Arakelian V.G., Water in oil-filled high-voltage equipment, Part I: States, Solubility, and Equilibrium in Insulating Materials, IEEE Electrical Insulation Magazine 2007; 23(4): 15-27, http://dx.doi.org/10.1109/MEI.2007.386480.
4. Gielniak J., Graczkowski A., Morańda H., Przybylek P., Walczak K., Nadolny Z., Moscicka-Grzesiak H., Gubanski S.M., Feser K., Moisture in cellulose insulation of power transformers: statistics, IEEE Transactions on Dielectrics and Electrical Insulation 2013; 20(3): 982-987, http://dx.doi.org/10.1109/TDEI.2013.6518968.
5. Gielniak J., Przybylek P., Moisture distribution in RBP bushing core (Rozkład zawilgocenia w rdzeniu izolatora przepustowego o izolacji typu RBP), in proc. of the X Conference Power and Special Transformers (Transformatory Energetyczne i Specjalne), 2014; 51-60 (in Polish).
6. Gradnik T., Koncan-Gradnik M., Petric N., Muc N., Experimental evaluation of water content determination in transformer oil by moisture sensor, in proc. of the IEEE International Conference on Dielectric Liquids (ICDL) 2011; 1-4, http://dx.doi.org/10.1109/icdl.2011.6015440.
7. Lopatkiewicz R., Nadolny Z., Przybylek P., The influence of water content on thermal conductivity of paper used as transformer windings insulation, in proc. of the 10th IEEE International Conference on the Properties and Applications of Dielectric Materials (ICPADM) 2012; 1-4, http://dx.doi.org/10.1109/icpadm.2012.6318991.
8. Oommen T.V., Lindgren S.R., Bubble evolution from transformer overload, in proc. of the Transmission and Distribution Conference and Exposition 2001; 137-142.
9. IEC 60422, Mineral insulating oils in electrical equipment. Supervision and maintenance guidance, 2006.
10. IEC 60814, Insulating liquids - Oil-impregnated paper and pressboard - Determination of water by automatic coulometric Karl Fischer titration, 2002.
11. Przybylek P., Investigations of the temperature of bubble effect initiation in oil-paper insulation (Badania temperatury inicjacji efektu bąbelkowania w izolacji papier-olej), Przeglad Elektrotechniczny (Electrical Review) 2010; 86(11B): 166-169 (in Polish).
12. Przybylek P., Moscicka-Grzesiak H., The influence of water content and ageing degree of paper insulation on its mechanical strength, in proc. of the 10th IEEE International Conference on the Solid Dielectrics (ICSD) 2010; 1-3, http://dx.doi.org/10.1109/icsd.2010.5568018.
13. Przybylek P., Water solubility in insulating liquids in aspect of moisture investigation using capacitive sensor (Rozpuszczalność wody w cieczach izolacyjnych w aspekcie badania ich zawilgocenia przy użyciu sondy pojemnościowej), Przeglad Elektrotechniczny (Electrical Review) 2012; 88(11A): 347-350 (in Polish).
14. Przybylek P., Siodla S., Determination of water solubility coefficients - text proposal for the Brochure, materiał niepublikowany opracowany na potrzeby grupy roboczej CIGRE D1.52, Helsinki 2015.
15. Przybylek P., The influence of temperature and cellulose polymerization degree on water distribution in oil-paper insulation, IEEE Transactions on Dielectrics and Electrical Insulation 2013; 20 (2): 552-556, http://dx.doi.org/10.1109/TDEI.2013.6508758.
16. Frame Instruction of Transformers Exploitation (Ramowa Instrukcja Eksploatacji Transformatorów), ZPBE Energopomiar, Gliwice, 2012 (in Polish).
17. Shroff D.H., Stannett A.W., A review of paper aging in power transformers, in the IEE Proceedings C. Generation, Transmission and Distribution 1985; 132(6): 312-319, http://dx.doi.org/10.1049/ip-c.1985.0052.
18. Vaisala HUMICAP Moisture and Temperature Transmitter Series HMT330 - User's Guide, 2008.
19. Ziomek W., Kuffel E., Sikorski W., Staniek P., Siodła K., Location and recognition of partial discharge sources in a power transformer using advanced acoustic emission method, Przeglad Elektrotechniczny (Electrical Review) 2008; 10/2008(84): 20-23.
20. Żukowski P., Kołtunowicz T.N., Kierczyński K., Subocz J. Szrot M., Gutten M., Sebok M., Jurcik J., An analysis of AC conductivity in moist oil-impregnated insulation pressboard, IEEE Transactions on Dielectrics and Electrical Insulation 2015; 22(4): 2156-2164, http://dx.doi.org/10.1109/TDEI.2015.004606.
21. Żukowski P., Kołtunowicz T.N., Kierczyński K., Subocz J. Szrot M., Formation of water nanodrops in cellulose impregnated with insulating oil, Cellulose 2015, 22(1): 861-866, http://dx.doi.org/10.1007/s10570-015-0543-0.

Piotr PRZYBYŁEK
Krzysztof SIODŁA
Institute of Electric Power Engineering
Poznan University of Technology
ul. Piotrowo 3a, 60-965 Poznań, Poland
E-mails: piotr.przybylek@put.poznan.pl, krzysztof.siodla@put.poznan.pl

Article citation info:
ESTRADA Q, SZWEDOWICZ D, MAJEWSKI T, MARTINEZ E, RODRIGUEZ-MENDEZ A. Effect of quadrilateral discontinuity size on the energy absorption of structural steel profiles. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 186–193, http://dx.doi.org/10.17531/ein.2016.2.5.

Quirino ESTRADA
Dariusz SZWEDOWICZ
Tadeusz MAJEWSKI
Eladio MARTINEZ
Alejandro Rodriguez-MENDEZ

EFFECT OF QUADRILATERAL DISCONTINUITY SIZE ON THE ENERGY ABSORPTION OF STRUCTURAL STEEL PROFILES

WPŁYW WIELKOŚCI CZWOROBOCZNYCH NIECIĄGŁOŚCI NA POCHŁANIANIE ENERGII W STALOWYCH PROFILACH STRUKTURALNYCH*

In this paper the effect of discontinuity size on energy absorption performance of steel square profiles is reported. The analysis consists of finite element simulations and experimental results of the compression strength of steel profiles with discontinuities. The discontinuities were placed at the mid span of the profiles in two walls opposite to each other. Square, rectangular and diamond initiators were evaluated at different scales. The numerical results determined the size intervals that present a good energy absorption performance in each case. Energy absorption capabilities were increased up to 12.54% with respect to a structure without discontinuities. Additionally, the peak load value (P_{max}) was decreased 25.97% with the implementation of a diamond initiator. For structures with discontinuities with major axis close to the profile width, a buckling effect was observed. Finally, it was observed that the size of the initiators contributes to reduce the peak load (P_{max}) value.

Keywords: *thin-walled structure, geometrical discontinuities, finite element method, energy performance.*

W pracy przedstawiono analizę wpływu rozmiaru nieciągłości na pochłanianie energii przez stalowe profile o przekroju kwadratowym. Analiza przedstawia wyniki symulacji elementami skończonymi próby ściskania profili stalowych z nieciągłościami oraz porównanie z danymi eksperymentalnymi. Nieciągłości zostały usytuowane w środku profilu w dwóch przeciwległych ścianach. W pracy zostały przebadane nieciągłości o formach kwadratowych, prostokątnych i rombowych dla różnych wymiarów. Stwierdzono wzrost o 12,54% możliwości pochłaniania energii w porównaniu dla struktur bez nieciągłości. Dodatkowo, w przypadku nieciągłości rombowych stwierdzono spadek wartości siły maksymalnej (Pmax) o 25,97%. Zaobserwowano występowanie efektu wyboczenia dla nieciągłości rombowej gdy wymiar jej osi zbliża się do szerokości profilu. Zaobserwowano, ze rozmiar nieciągłości wpływa na redukcje wartości maksymalnego obciążenia oraz w tym samym czasie na obniżenie pochłanianej energii.

Słowa kluczowe: *struktury cienkościenne, nieciągłości geometryczne, metoda elementów skończonych, pochłanianie energii.*

1. Introduction

Worldwide, an estimated of 1.2 million people die in road crashes every year and 50 million are injured. This evidence can be increased by 65% for the next 20 years [16]. With the introduction of the crashworthiness concept in previous decades and the addition of the efforts to ensure the life of the passengers, the use of thin walled structures as energy passive absorbers is taking relevance. Many characteristics can be attributed to them; however, the most important is the high performance which absorbs energy by plastic deformation. With the objective to optimize the energetic behavior of structural members, numerical and experimental studies have been realized. In these studies, factors such as cross section geometry [21, 11], length of the profile [10] and manufacture material [17] have been evaluated. Other studies have focused their efforts in the implementation of geometrical discontinuities to reduce the peak load value. Many shapes of imperfections have been analyze such as circular [2, 4, 5], slotted [13], elliptical [9] and dimples [8]. In all cases the effectiveness of discontinuities have been corroborated, nevertheless according to Szwedowicz et al [18] the energy abortion capacity also can be modified by location of imperfections along the structure. In [14] the effect of location of circular discontinuities on square profile was studied, concluding that the best performance is obtained by placing discontinuities at middle height. Likewise, a numerical comparison between circular, elliptical and slotted discontinuities was carried out by [6]. In the analysis the initiators were located at the midpoint of the square aluminum tube in two opposite walls. A reduction of 11.7% in peak load value was obtained with a pair of holes. In this paper the effect of discontinuity size on the crashworthiness characteristics of square steel profiles is analyzed. For this purpose, numerical simulations were realized using Abaqus finite element software. Square, rectangular and elliptical geometries discontinuities were evaluated at difference scales.

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

2. Measurement of the energy absorption performance

As mentioned there are two important parameters to determine the response of structural members under crushing force; the energy absorbed (E_a) and the maximum value of crushing load called peak load (P_{max}). The energy absorption by elastic and plastic deformation can be calculated by integration of the area below the load-displacement curve, using the following equation [7]:

$$E_a = \frac{1}{2}\sum_{i}^{n-1}\left(F(\delta)_{i+1} + F(\delta)_i\right)\cdot(\delta_{i+1} - \delta_i) \quad (1)$$

where: F is the crushing force and δ is the displacement in the axial direction.

After reaching the P_{max} value, the initial stiffness of the structure is broken. Then the force required to continue with plastic deformation decreases with an oscillating value. This force is denominated mean crushing force and it is equal to the ratio of the absorbed energy (E_a) and displacement (δ) [7]:

$$P_m = \frac{E_a}{\delta} \quad (2)$$

A dimensionless parameter that associates the energy absorbed (E_a), the peak load (P_{max}) value and the displacement (δ) is known as the energy efficiency (E_e) defined by [19], is expressed in %:

$$E_e = \frac{E_a}{P_{\max\cdot\delta}}\cdot 100\% \quad (3)$$

Finally the specific energy (S_e) is used when the structures evaluated have different mass. It is defined as the ratio of energy absorbed (E_a) to the mass (m) of the structure [12]:

$$S_e = \frac{E_a}{m} \quad (4)$$

3. Numerical model and experimental validation

A first numerical model about the crushing process of a steel square profile without discontinuities was developed in Abaqus/Explicit. The geometry evaluated corresponds to a structure with square cross-section (50 x 50 mm), 240 mm in length, thickness of 1.32 mm and rounded corners (radius of 2 mm). According to [3, 18] and in addition with the quasi-static nature of the compression test, the profile was modeled with elastic-plastic material properties and the hardening effects were negated. The structural member was modeled as a deformable body with shell elements (S4R) with a thickness of 1.32 mm. The crushing process was carried out by two rigid plates using R3D4 elements. A fixed displacement restriction was applied to a lower plate while an upper rigid plate could move only in the y- direction to allow the compression of the structure. The boundary conditions applied to a profile were a tie restriction between the upper tip and the lower rigid plate beside a contact interaction between the upper tip and the top rigid plate. Finally a general contact condition was established to ensure the internal and external contact during the forming of wrinkles. A frictional coefficient of 0.15 was used for all contact conditions. A smooth post-buckling response was introduced in the discrete model by the *IMPERFECTION command. In this sense were introduced first two eigenvalues with an imperfection sensivity equal to 0.1 of thickness value. The eigenvalues were obtained from buckle analysis in Abaqus. The elastic-plastic properties of the material were established using an isotropic plasticity model. The mechanical properties of the material used for the development of the numerical model, were [15]: Young modulus = 200 GPa, Poisson's ratio = 0.26, yield stress = 250 MPa, density = 7850 kg/m^3. The hardening effects due to the load velocity were not considered in the discrete models.

The numerical model of the structure was validated experimentally by a quasi-static crushing test. The structural member was subjected to compression load by universal test machine (Shimadzu UH-300 kNI) with a velocity of 6 mm/min. During the evaluation test, the profile was located between two compression plates, after a maximum displacement of 160 mm was programmed. The figures 1, 2, 3 and 4 show the mechanical behavior of the profile to compare numerical and experimental results. In all cases a good agreement it was obtained. In this form the usefulness of the discrete model of the test specimen was experimentally validated and it was used to modeling discontinuities.

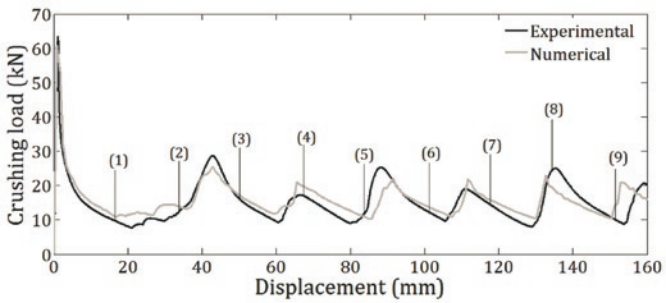


Fig. 1. Comparison of numerical and experimental load/displacement curves

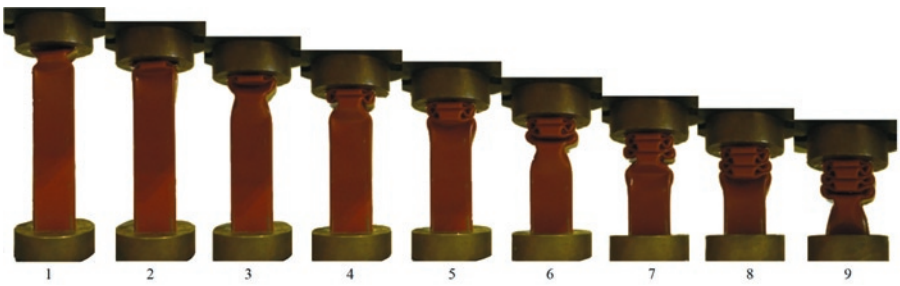


Fig. 2. Progressive collapse of specimen ST-01 (without holes)

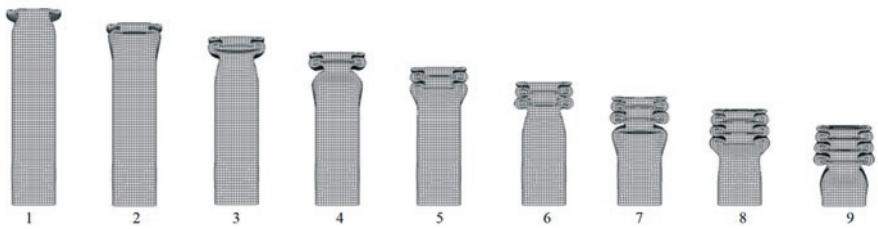


Fig. 3. Numerical simulation of the progressive collapse process of specimen ST-01 (without holes)

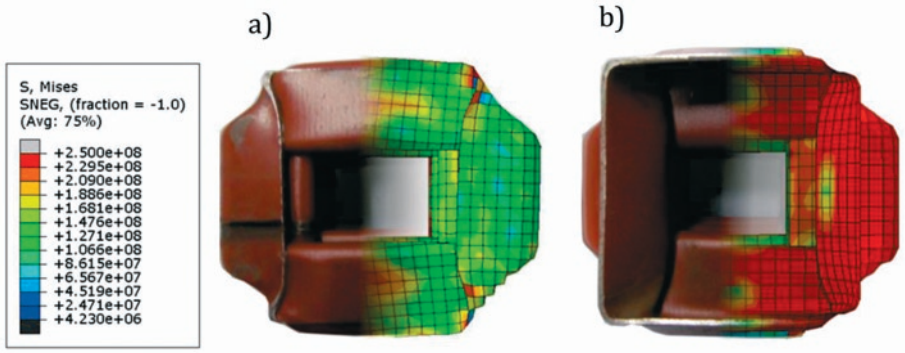


Fig. 4. Experimental and numerical final deformation state for the straight square profile (ST-01), where: a) top view and b) bottom view, S in MPa

4. Development of numerical simulation with discontinuities

The effect of discontinuity size on the response of square structures is evaluated by means of several numerical models. Since the best position to bore imperfections is at the mid span of the specimen walls in symmetric arrangements [14, 18], square, rectangular and diamond discontinuities were placed at mid span on the structure in two opposite walls (120 mm). The initial sizes of the imperfections were determined from the Von Karman principle based on effective width. In this way the implementation of the discontinuity modifies the mechanical response of the structure.

The structures were divided in three groups depending on the shape of the discontinuity. In each case, the discontinuity size was increased by a scale factor; this factor defines the resizing rate of the discontinuities in respect to the smallest size of the initiator for every group. The geometry of the structures is given in Table 1 for groups I, II and III, respectively.

Table 1. Geometry employed in group I, II, III

Group I. Geometrical details				
Profile code	Shape of discontinuity	Size of discontinuity (mm)	Scale Factor	Mass (gr)
		Side		
I-A	Square	5.32	-	477.94
I-B		7.98	1.50	477.21
I-C		10.63	2.00	476.18
I-D		13.29	2.50	474.86
I-E		15.95	3.00	473.25
I-F		18.61	3.50	471.35

Group II. Geometrical details					
Profile code	Shape of discontinuity	Size of discontinuity (mm)		Scale Factor	Mass (gr)
		Major axis (horizontal)	Minor axis (vertical)		
II-A	Rectangular	6.65	4.25	-	477.94
II-B		9.97	6.38	1.50	477.21
II-C		13.29	8.51	2.00	476.18
II-D		16.62	10.63	2.50	474.86
II-E		19.94	12.76	3.00	473.25
II-F		23.26	14.89	3.50	471.35

Group III. Geometrical details					
Profile code	Shape of discontinuity	Size of discontinuity (mm)		Scale Factor	Mass (gr)
		Major axis (horizontal)	Minor axis (vertical)		
III-A	Diamond	9.40	6.02	-	477.94
III-B		14.10	9.02	1.50	477.21
III-C		18.80	12.03	2.00	476.18
III-D		23.50	15.04	2.50	474.86
III-E		28.20	18.05	3.00	473.25
III-F		32.90	21.06	3.50	471.35

5. Results and discussion

5.1. Group I. Square discontinuities

The crushing behaviour of structures in group I was obtained by force-displacement curves, which are presented in figures 5 and 6. The crushing load and plastic folds of structures I-A, I-B and I-C present similar behaviours. Even if the mechanical response of the structures were similar, small differences are found at P_{max} due to the size of discontinuities. After reaching the peak value (P_{max}), the crushing load continues to deform the material with a mean crushing force of 17.62 kN. Besides at displacement of 110 mm, a second increase of crushing load is produced.

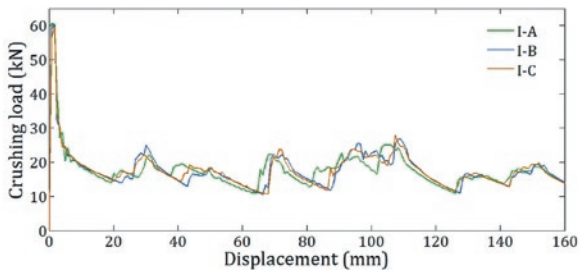


Fig. 5. Load–displacement curves for structure with square discontinuities (first part)

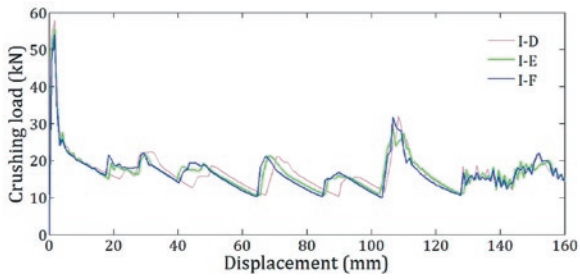


Fig. 6. Load–displacement curves for structure with square discontinuities (second part)

In respect to profiles I-D, I-E and I-F, the effect of size on P_{max} has a better appreciation. A decrease of this value was noticed by increasing the size of the initiators. The mean crushing force (P_m) for these profiles presents an approximated value of 16.87 kN. This means a decrease of P_m with respect to the first structures. Small differences occurred during the plastic fold mechanism, although the appearing of a second pulse of crushing load was presented in all cases, at a displacement of 105 mm (see figure 6).

The energy absorption capacities of the profiles depend on the quantity and mode of deformation. The absorbed energy occurs in

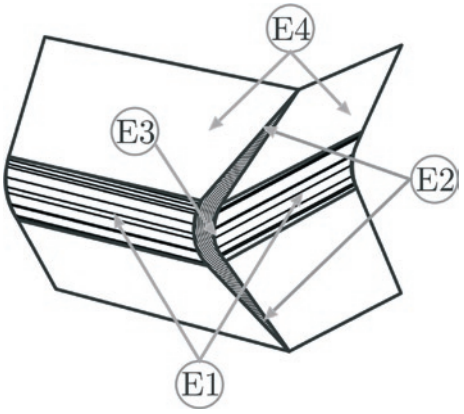


Fig. 7. Forming of plastic surfaces at basic folding mechanism element [20]

Table 2. Numerical results for group I.

Profiles with square discontinuities (group I)						
Specimen code	P_{max} (kN)	P_m (kN)	E_a(kJ)	E_e(%)	S_e(J/gr)	Deformation mode
ST-01	61.03	15.94	2.55	26.11	5.32	s
I-A	60.56	17.69	2.83	29.21	5.92	s
I-B	59.99	17.68	2.82	29.47	5.93	s
I-C	59.70	17.77	2.84	29.77	5.97	s
I-D	57.87	16.93	2.70	29.25	5.70	s
I-E	55.40	16.96	2.71	30.61	5.73	s
I-F	53.98	16.80	2.68	31.12	5.70	s

Fig. 8 Effect of size of discontinuities on formation hinge lines for group I

four different ways by the formation of cylindrical (E1), conical (E2), toroidal (E3) and trapezoidal surfaces (E4), (see figure 7). For all analysed structures, the formation of plastic wrinkles was determined by localization of the discontinuity. Additionally, it was observed, that as the size of discontinuities increases, the formation of hinge lines presents low resistance on collapsing, in this way the peak load value diminished (see figure 8). However, the size of discontinuities was increased; a buckling effect appeared at the vicinity of the discontinuity, causing just a partial deformation in the structure.

After the first wrinkle was created, the structures showed two different ways of fold formation. Some structures were deformed with direction to the upper tip of the profile (e.g. I-A) and others continued to the bottom tip profile (e.g. I-D). The final deformation state for all structures is shown in figure 9. The global collapse mode of the structures was symmetric (s), independent of the hole´s size. A summary of the performance of the structures belonging to the group I is shown in Table 2. In all cases, the P_{max} value decreases within a range of 0.77–11.55% compared with a profile without discontinuities (ST-01). The energy absorbed increased until reaching a maximum value of 2.84 kJ (I-C), later an increase in the scale factor produced a reduction of energy absorption in structures I-D, I-E and I-F. The largest energy efficiency (E_e) was obtained for the profile named I-F with a value of 31.12%, a value close to 100% represents the optimal efficiency of the structure. Due to the quasi-static nature of the compression test simulation and the symmetry of the discontinuities, the final deformation state of the specimens, was symmetric (s).

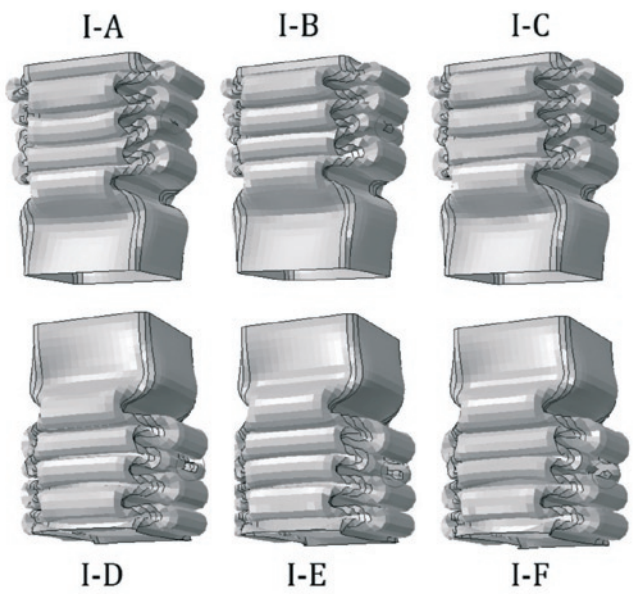


Fig. 9. Final deformation state for profiles with square discontinuities

5.2. Group II. Rectangular discontinuities

The crushing load vs displacement curves for these profiles are shown in figures 10 and 11. A perceptible difference was noticed between the profile II-A and II-B. Moreover, structures II-B and II-C described a similar crushing behaviour during the plastic deformation process. The peak load (P_{max}) at the beginning of the compression process was diminished by increasing the size of the discontinuity.

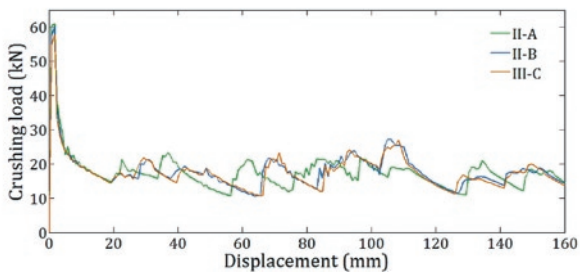


Fig. 10. Load–displacement curves for structures with rectangular discontinuities (first part)

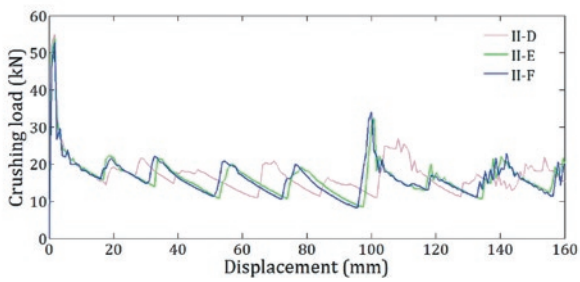


Fig. 11. Load–displacement curves for structure with square discontinuities (second part)

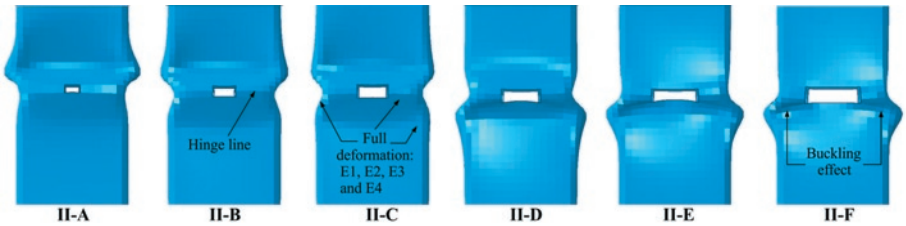


Fig. 12. Effect of size of rectangular discontinuities on formation hinge lines for group II

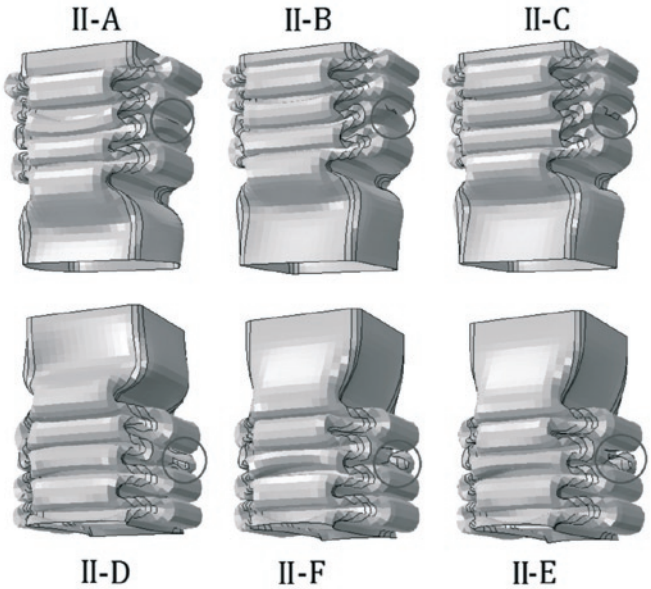


Fig. 13. Final deformation state for profiles with rectangular discontinuities.

Table 3. Numerical results for group II

Profiles with rectangular discontinuities (group II)						
Specimen code	P_{max} (kN)	P_m (kN)	E_a(kJ)	E_e(%)	S_e(J/gr)	Deformation mode
ST-01	61.03	15.94	2.55	26.11	5.32	s
II-A	61.08	17.60	2.81	28.81	5.89	s
II-B	60.00	17.88	2.86	29.79	5.99	s
II-C	57.81	17.74	2.83	30.68	5.96	s
II-D	54.88	17.09	2.73	31.14	5.76	s
II-E	53.78	16.44	2.63	30.57	5.56	s
II-F	52.81	16.40	2.62	31.05	5.57	s

The tendency of reduction of the peak load continues to appear for structures II-D, II-E and II-F. Very close mechanical behaviours are described by profiles II-E and II-F, particularly during the emergence of an increase in the crushing load at approximately 97 mm. This increase in crushing load is due to a change in the direction of the formation of the plastic folds. The mean crushing force (P_m) for II-E and II-F presented similar values during the plastic deformation process with an approximate value of 16. 42 kJ.

According to numerical simulations, effects such as the formation of hinge lines during the first plastic wrinkles and the energy absorption capabilities (E_a), are directly associated to the size of the discontinuities. The energy absorption occurs by plastic deformation of the structure (see figure 7). An increase on the discontinuity size generates a buckling effect at the near regions of the discontinuity. This effect provokes a reduction of the peak load and at the same time a reduction of the energy absorption capacity was observed due to a partial deformation of the structure (see figure 12).

After the collapse of the structure for the first time, the plastic deformation process was propagated in two directions, one at a time, depending on the hinge line formation. Consequently, all the material was deformed in one direction, the process repeats in the opposite direction until reaching the final state of deformation. The structures showed symmetric collapse modes independent of the size of the hole (see figure 13).

Table 3 presents the results obtained for the profiles in group II. Accordingly, as the scale factor is increased, the P_{max} values decrease to around 1.69-13.47% with respect to a structure without holes (ST-01). Also, the energy absorbed (E_a) increases until it reaches a maximum value of 2.86 kJ (12.15%) for the structure II-B. Later the profiles showed a gradual decay of performance on E_a by the resizing of discontinuities. The largest energy efficiency (E_e) was registered by profile II-D with a value of 31.14%. A symmetric collapse mode was observed in all the profiles evaluated.

5.3. Group III. Diamond discontinuities

The load–displacement curves for structures with diamond shape discontinuities are presented in figures 14 and 15. According to these results, the size of the discontinuities did not modify the behaviour of structure III-B compared to III-A. Regarding to profile III-C, a second major value of peak load was noticed at a 100 mm displacement. This increase of peak load is due to a change in the direction of the formation of wrinkles.

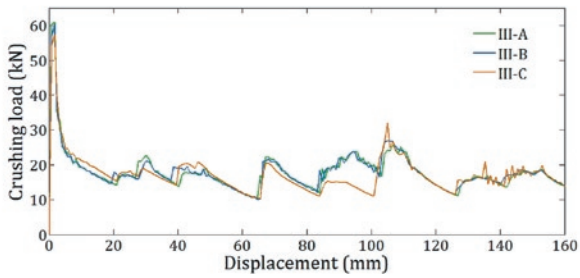


Fig. 14. Load–displacement curves for structure with diamond discontinuities (first part)

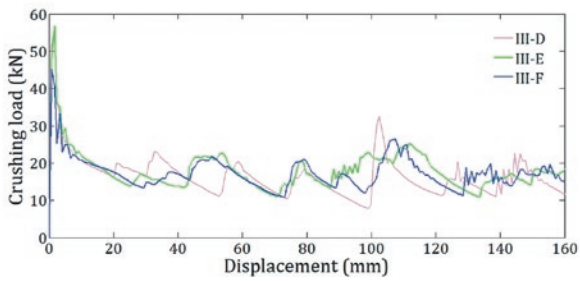


Fig. 15. Load–displacement curves for structure with diamond discontinuities (second part)

The effect of size of the diamond-shape hole on peak load (P_{max}) projected from the structure III-F, where a reduction of peak load (P_{max}) was obtained with respect to structures III-D and III-E, which show similar peak load values. The crushing response of the structures presents similar loops, however when the mean crushing force (P_m) was calculated, some differences were found. Thus, the mean

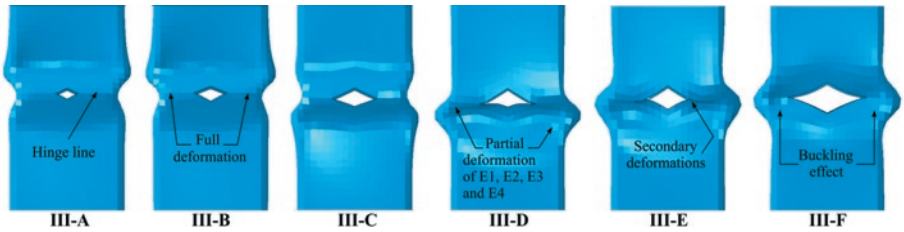


Fig. 16. Effect of size of diamond discontinuities on formation hinge lines for group III

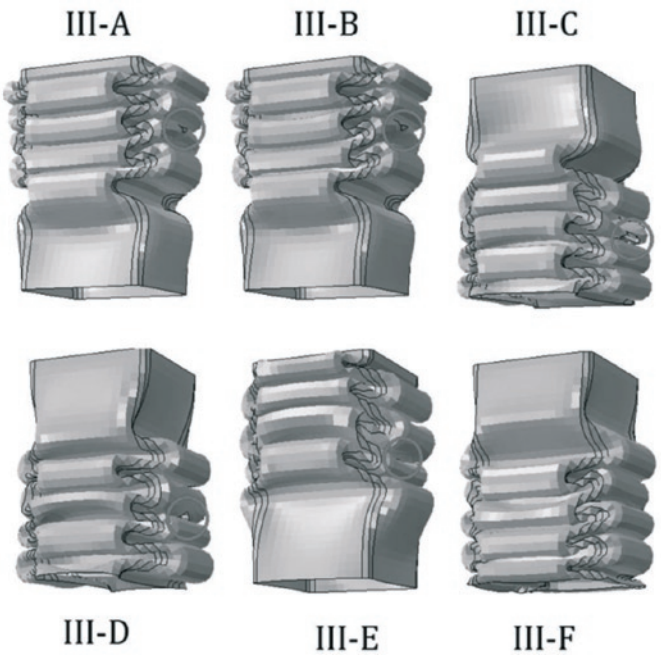


Fig. 17. Final deformation state for profiles with diamond discontinuities

crushing force values were 16.56, 17.94 and 17.34 kN for structures III-D, III-E and III-F, respectively.

The formation of hinge lines was determined by the location of the discontinuity. In this part, the crushing load required for its formation depended on the size of the discontinuities, decreasing as the scale factor increased. During the creation of the external cylindrical face (E1) the structure III-A and III-B showed a full deformation of material while it was partially interrupted in structure III-C, contributing to diminish the energy absorption performance (see figure 16). For structures with discontinuities of length of major axis, close to the width of the profile, the formation of the external and internal cylindrical faces was interchanged (III-D, III-E and III-F).

The final deformation state of the structures with diamond shape holes is shown in figure 17. Despite the fact that the profile described

Table 4. Numerical results for group III

Profiles with diamond discontinuities (group III)						
Specimen code	P_{max} (kN)	P_m (kN)	E_a(kJ)	E_e(%)	S_e(J/gr)	Deformation mode
ST-01	61.03	15.94	2.55	26.11	5.32	s
III-A	60.93	17.84	2.85	29.23	5.97	s
III-B	60.16	17.92	2.86	29.71	6.01	s
III-C	57.32	17.32	2.77	30.20	5.82	s
III-D	55.72	16.56	2.64	29.61	5.58	s
III-E	56.82	17.94	2.87	31.57	6.06	s
III-F	45.18	17.34	2.77	38.32	5.89	s

different directions of propagation after the first collapse; the complete structure reached symmetric modes of deformation.

A summary of the results obtained is given in Table 4. According to these results, the peak load (P_{max}) decreases as the size of the discontinuities increases. The minimum value of P_{max} reached was 45.18 kN for structure III-F, this quantity represents a decrease of 25.97% respect to a square profile without discontinuities (ST-01). A maximum value of energy absorbed of 2.87 kJ (12.54%) was obtained by III-E, this means a good relation of specific energy (S_e) of 6.06 J/gr. However, according to E_e, the structure III-F presented the best energetic performance with a value of 38.32 %. Finally a symmetric (s) mode of deformation was observed in all structures evaluated.

5.4. Discussion of results

In order to visualize the response of the structures due to variation of geometry and initiator's size, a comparison of peak load (P_{max}) values for all structures is presented in figure 18. If the outlier value for structure III-E is neglected, a tendency of reduction in the peak load due to an increase in size of the discontinuity is observed. For structures with denomination A and B (Groups I, II and III) the value of P_{max} was independent of the initiator geometry. The geometrical factor gains importance in the structures with denomination C.

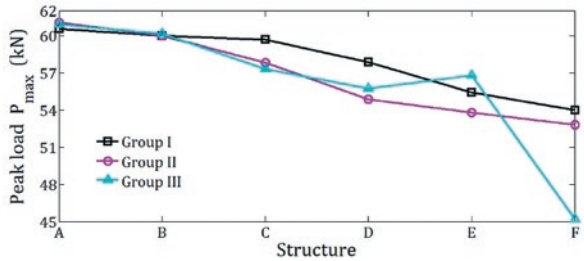


Fig. 18. Comparison of P_{max} values obtained for different groups of structures

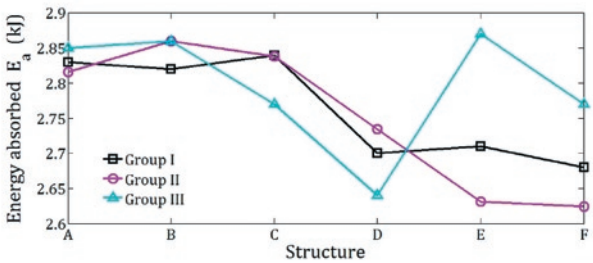


Fig. 19. Comparison of energy absorbed for different groups

Independently of geometry of holes a reduction of energy absorbed performance was noticed for groups I and II (profiles with square and rectangular initiators). In this way a maximum value of 2.84 kJ and 2.86 kJ was obtained for profiles I-B and II-B, respectively. The structures with diamond discontinuities (Group III) followed the same tendency of decrease of structure D. Subsequent profiles III-E and III-F show a secondary effect that caused an increase of energy absorption (see figure 19).

The increase of energy absorption characteristics, contrary to what is expected for specimens III-E and III-F, is directly associated with the mode of formation of the first wrinkle. Figures 20 is presented where the size of discontinuities causes torsional effects (2) along the major axis of the initiator, furthermore, a movement of inward curl

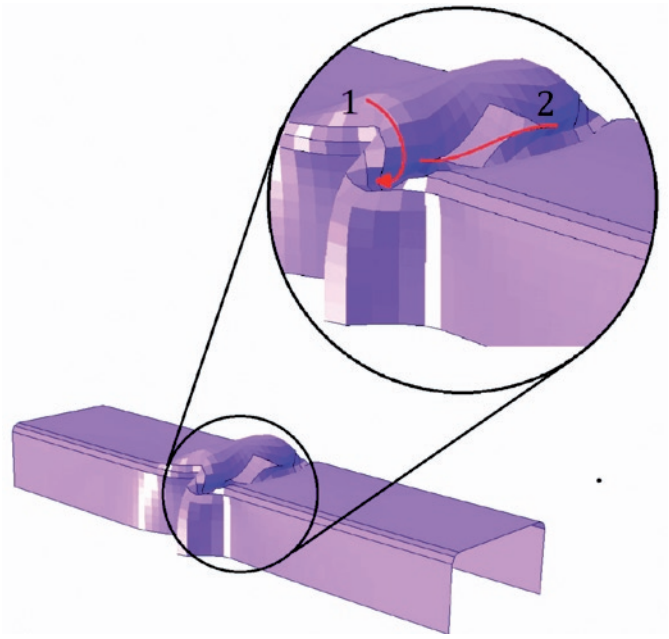


Fig. 20. Secondary effects by increase the size of the discontinuity in profile III-E

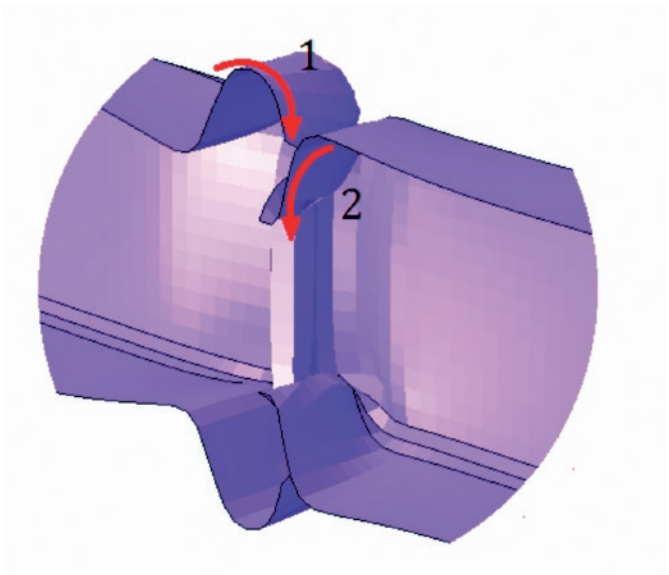


Fig. 21. Secondary effects by increase the size of the discontinuity in profile III-E

was developed (1), both effects contribute to increase the energy absorption performance.

Finally, with a cut view of profile III-E, a third mechanism of energy absorption is exposed, which indirectly improves the performance of the profile. After the inward curl process is completed (1), a contact between folds (external and internal) is presented. It produces additional work to deform the inner wrinkle counterclockwise (3), such as seen in figure 21.

Finally according to figure 19 for the specific case of this paper, there is a maximum energy value that can be absorbed by the structures suggesting a limit for the sizing of the discontinuities. In this way the limits related to the sizing of the square, rectangular and diamond discontinuities were numerically found according to the width of structure C, side (S) and the length of the major axis of the discontinuity (D).

For square discontinuities:

$$0.10C \leq S \leq 0.21C$$

For rectangular discontinuities:

$$0.13C \leq D \leq 0.27C$$

For diamond discontinuities:

$$0.19C \leq D \leq 0.28C$$

5. Conclusion

A numerical study was carried out to evaluate the effect of size of quadrilateral discontinuities on crashworthiness performance of square profiles. According to results obtained in the 'perfect' profile (ST-01), the implementation of both square, rectangular and diamond discontinuities, showed a reduction on peak load (P_{max}) value within a range of 0.77 – 25.97%. As the size of discontinuities grows, a reduction of peak value was registered independently of the initiator's shape. In all groups, the size of discontinuities determined the appearing of buckling effects at the beginning of the collapse of the profiles. It was observed that the reduction of the peak value (P_{max}) is influenced by the length of the major axis of the initiator in a major scale than in the minor axis. For discontinuities with a major axis length close to the width of the structure, the occurrence of hinge lines was faster and with less effort than discontinuities with denomination A and B. In all groups, the shape of the discontinuities is of great importance on peak load values from structures with denomination C holes (2.00 factor scale). With respect to energy absorption characteristics (E_a), the implementation of any type and size of discontinuities showed a better performance compared with a profile without discontinuities within a range of 2.74 – 12.54%. The values of E_a in all cases were increased until reaching an approximate maximum value of 2.86 kJ, after a decrease of E_a was noticed. A particular case was observed with structures III-E and III-F, which has a tendency to decrease E_a by increasing the size of initiator, these structures registered a second increase on energy absorption capabilities. Finally, it was found that the reduction of E_a capabilities resulted in partial deformations by buckling effect during the hinge line formation. This effect is major for structures with initiators with scale factors of 2.50, 3.00 and 3.50. For the specific case of profiles III-E and III-F it was observed that the relation between the minor axis and the major axis in addition to the diamond shape, have effects such as torsional and secondary deformations, improving the energy absorption characteristics (see fig. 20 and 21).

References

1. ABAQUS/Explicit User's Manual, Theory and examples manual and post manual. Version 5.8, Explicit, HKS, Inc., 1998.
2. Abramowicz W, Jones N. Dynamic axial crushing of circular tubes. International Journal of Impact Engineering 1984; 2(3): 263-281, http://dx.doi.org/10.1016/0734-743X(84)90010-1.
3. Aljawi A A N. Axial crushing of square steel tubes. The 6th Saudi Engineering Conference, KFUPM, Dhahran 2002; 3: 3-18.
4. Arnold B, Altenhof W. Experimental observations on the crush characteristics of AA6061 T4 and T6 structural square tubes with and without circular discontinuities. International Journal of Crashworthiness 2004; 9(1): 73-87, http://dx.doi.org/10.1533/ijcr.2004.0273.

5. Chenga Q, Altenhof W, Li L. Experimental investigations on the crush behaviour of AA6061-T6 aluminium square tubes with different types of through-hole discontinuities. Thin-Walled Structures 2006; 44(4): 441-454, http://dx.doi.org/10.1016/j.tws.2006.03.017.
6. Chiu Y S, Jenq S T. Crushing behaviour of metallic thin-wall tubes with triggering mechanisms due to quasi-static axial compression, Journal of the Chinese Institute of Engineers 2014; 37:4, 469:478.
7. Chung Kim Yuen S., Nurick G. N. The energy-absorbing characteristics of tubular structures with geometric and material modifications: An overview. ASME Applied Mechanics Reviews 2008; 61:020802, 1-15.
8. Ferdynus M. An energy absorber in the form of a thin-walled column with square cross-section and dimples. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2013; 15 (3): 253-258.
9. Huang M Y, Tai Y S. Dynamic crushing characteristics of high strength steel cylinders with elliptical geometric discontinuities. Theoretical and Applied Fracture Mechanics 2010; 54(1): 44-53, http://dx.doi.org/10.1016/j.tafmec.2010.06.014.
10. Jensen O, Langseth M, Hopperstad O S. Experimental investigations on the behaviour of short to long square aluminium tubes subjected to axial loading. International Journal of Impact Engineering 2004; 30(8-9): 973-1003, http://dx.doi.org/10.1016/j.ijimpeng.2004.05.002.
11. Kim D K, Lee S. Impact energy absorption of 6061 aluminium extruded tubes with different cross-sectional shapes. Materials and Design 1999; 20(1): 41-49, http://dx.doi.org/10.1016/S0261-3069(98)00042-9.
12. Krauss C A, Lananen D H. A parametric study of crush initiators for a thin-walled tube. International Journal of Vehicle Design 1994; 15(3/4/5): 385-401.
13. Lee, S, Hahn C, Rhee M, Oh J E. Effect of triggering on the energy absorption capacity of axially compressed aluminum tubes. Mechanics and Design 1999; 20(1): 31-40, http://dx.doi.org/10.1016/s0261-3069(98)00043-0.
14. Mamalis A G, Manolakos D E, Spentzas K N, Ioannidis M B, Koutroubakis S, Kostazos P K. The effect of the implementation of circular holes as crush initiators to the crushing characteristics of mild steel square tubes: experimental and numerical simulation. International Journal of Crashworthiness 2009; 14(5): 489-50, http://dx.doi.org/10.1080/13588260902826547.
15. Matweb, http://www.matweb.com, 2014.
16. Peden M., Scurfield R., Sleet D., Mohan D., Adnan A. Hyder A. A., Eva Jarawan E. Mathers C., World report on road traffic injury prevention. Geneva, World Health Organization, 2004.
17. Simhachalam B, Krishna Srinivas K, Lakshmana Rao C. Energy absorption characteristics of aluminium alloy AA7XXX and AA6061 tubes subjected to static and dynamic axial load. International Journal of Crashworthiness 2014; 19(2): 139-152, http://dx.doi.org/10.1080/13588265.2013.878974.
18. Szwedowicz D, Estrada Q, Cortes C, Bedolla J, Alvarez G, Castro F. Evaluation of energy absorption performance of steel square profiles with circular discontinuities. Latin American Journal of Solids and Structures 2014; 11: 1744-1760, http://dx.doi.org/10.1590/S1679-78252014001000003.
19. Thornton, P. H., Mahmood, H. F., and Magee, C. L. Energy absorption by structural collapse. Structural Crashworthiness 1983, 96-117.
20. Wierzbicki T, Abramowicz W. On the crushing mechanics of thin-walled structures J. Appl. Mech 1983; 50(4a), 727-734, http://dx.doi.org/10.1115/1.3167137.
21. Zhang X, Huh H. Crushing analysis of polygonal columns and angle elements. International Journal of Impact Engineering. 2010; 37(4): 441-451, http://dx.doi.org/10.1016/j.ijimpeng.2009.06.009.

Quirino Estrada
Institute of Technology and Engineering
Autonomous University of Juarez City
Av. Plutarco Elías Calles 1210 C.P 32310, Ciudad Juarez, Chihuahua, Mexico

Dariusz Szwedowicz
Mechanical Engineering Department
National Center for Research and Technological Development
Internado Palmira S/N C.P 62490, Cuernavaca, Mor., Mexico

Tadeusz Majewski
Mechanical and Industrial Engineering Department
Universidad de las Americas
Cholula S/N C.P. 72810, Puebla, Mexico

Eladio Martinez
Mechanical Engineering Department
National Center for Research and Technological Development
Internado Palmira S/N C.P 62490, Cuernavaca, Mor., Mexico

Alejandro Rodriguez-Mendez
Department of Mechanical Engineering
University of California
Berkeley, CA 94720, USA

E-mails: quirino.estrada@uacj.mx, d.sz@cenidet.edu.mx,
tadeusz.majewski@udlap.mx, mare@cenidet.edu.mx, aleromende@berkeley.edu

Article citation info:
GRONOSTAJSKI Z, HAWRYLUK M, KASZUBA M, ZIEMBA J. Application of a measuring arm with an integrated laser scanner in the analysis of the shape changes of forging instrumentation during production. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 194–200, http://dx.doi.org/10.17531/ein.2016.2.6.

Zbigniew GRONOSTAJSKI
Marek HAWRYLUK
Marcin KASZUBA
Jacek ZIEMBA

APPLICATION OF A MEASURING ARM WITH AN INTEGRATED LASER SCANNER IN THE ANALYSIS OF THE SHAPE CHANGES OF FORGING INSTRUMENTATION DURING PRODUCTION

ZASTOSOWANIE RAMIENIA POMIAROWEGO ZE ZINTEGROWANYM SKANEREM LASEROWYM DO ANALIZY ZMIAN KSZTAŁTU OPRZYRZĄDOWANIA KUŹNICZEGO W TRAKCIE PRODUKCJI*

In the article, the authors present an innovative approach consisting in an analysis of the wear (shape changes) of one of the forging tools directly during production, with the use of a laser scanner, without the necessity of their disassembly from the forging unit. The tests consisted in direct measurements of the shape changes of cyclically sampled forgings during the forging process (every 1000 item), and next, based on the proceeding wear, an indirect analysis was performed of the shape change of the impression of the selected tool, i.e. a filler. At the time of the short technological intervals in the process, direct measurements were performed of the tool itself, with the purpose of verifying the results of the forgings' measurement in relation to the actual changes in the tool. The performed analyses showed a good agreement of the geometrical properties of the surfaces (of the selected forgings representing the proceeding wear of the tool) and the geometrical defect of the working impression of the tool, based on the direct measurements during the production process. The obtained results allow for a fast analysis of the forging tool life with respect to the quality and the quantity (of material defect), which, in consequence, leads to significant economical savings. The proposed method makes it possible to make decisions on the time period of the tool operation based on the tools' actual wear, instead of – as has been the case in forging plants up till now – after the given maximal number of forgings has been made or a premature tool damage has been observed.

Keywords: *scanning of forging tools, tool life, failure mechanisms.*

W artykule autorzy przedstawili innowacyjne podejście polegające na analizie zużywania się (zmian kształtu) jednego z narzędzi kuźniczych bezpośrednio w trakcie produkcji przy wykorzystaniu skanera laserowego, bez konieczności ich demontażu z agregatu kuźniczego. Badania polegały na bezpośrednich pomiarach zmian kształtu cyklicznie pobieranych odkuwek podczas procesu kucia (co 1000szt), a następnie na podstawie postępującego zużywania dokonywana była w sposób pośrednia analiza zmian kształtu wykroju wybranego narzędzia - wypełniacza. Natomiast w momencie krótkich przerw technologicznych w procesie przeprowadzano bezpośrednie pomiary samego (analizowanego) narzędzia w celu weryfikacji wyników pomiaru odkuwek w stosunku do rzeczywistych zmian narzędzia. Przeprowadzone analizy wykazały dużą zgodność cech geometrycznych powierzchni (wybieranych odkuwek odzwierciedlających zużywanie się narzędzia), a ubytkiem geometrycznym wykroju roboczego narzędzia na podstawie bezpośrednich pomiarów podczas produkcji. Uzyskane rezultaty pozwoliły na dokonanie szybkiej analizy trwałości narzędzia kuźniczego pod względem jakościowym i ilościowym (ubytku materiału), co w konsekwencji prowadzi do znacznych oszczędności. Zaproponowana przez autorów metoda pozwala na podejmowanie decyzji o czasie eksploatowania narzędzi na podstawie ich rzeczywistego zużycia, a nie, jak to ma miejsce obecnie w kuźniach, po określonej maksymalnej ilości wykonanych odkuwek lub zaobserwowanego w tym okresie przedwczesnego uszkodzenia narzędzia.

Słowa kluczowe*: skanowanie narzędzi kuźniczych, trwałość, mechanizmy niszczące.*

1. Introduction

The high competition on the forged product supply market more and more often makes the quality of the offered forgings, beside the price of the product, the decisive parameter in the selection of the supplier. This is especially important when the recipient of the product is the automotive or aviation industry, where these requirements are at the highest level. The process of die forging is one of the most difficult production processes with respect to implementation. Despite the fact that this technology is relatively well-known, the proper production, especially of forgings of a complicated shape, which will meet the requirements regarding the precision and quantity made by the recipients, requires high experience from the technologists and operators. At each stage of the forging process, there is a potential risk of error, which lowers the quality of the produced forgings. One of the basic factors affecting the forging quality is the life of the ap-

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

plied tools, as their wear causes a change in the shape of the product, and any surface flaws of the tools (cracks, defects) are represented on the forged product. That is why a detailed quality control is so important during production. At present, such check-ups consist in a visual assessment of the quality of the produced element as well as a measurement of selected control points with the use of traditional measuring devices or simple curve gauges. These methods do not, however, provide the possibility to assess the quality and shape of the whole element and so, more and more often, other measuring methods are applied, such as the ones which use the coordinate measuring technique [3, 7, 9, 10].

The coordinate measuring technique provides a lot of possibilities for the development of modern metrological thought [12, 16, 27]. In the industrial coordinate technology, one can observe new trends. The most conspicuous change is the necessity of using 3D models during measurements which are in accordance with the standards of the GPS measurement chain (Geometrical Product Specifications) [10, 11, 21, 26]. Another important trend is increasing the number of measurement points necessary for determining the analyzed geometrical properties. One of those properties is the volumetric wear parameter, which can also be used in the analysis of the shape changes of forging tools, with the purpose to prognosticate their wear process. Among others, this parameter is used in medicine. For example, in works [1, 17, 18], the authors present the use of a coordinate measuring machine equipped with a contact measuring head for analyzing the volumetric wear parameters of the spherical surfaces of joint prostheses. In work [15], the authors expand the research concerning the analysis of the effect of the change of numerous scanning parameters, with the use of the contact scanning measuring head, whereas the authors of works [2, 18, 25] analyze the 'beneficial' effect of increasing the number of measurement points on the accuracy of the determined volumetric parameters.

At present, the industry is increasingly interested in mobile measuring devices. These undoubtedly include measuring arms equipped with linear laser scanners with dedicated specialized software. Measuring arms, through their mobility and universality, are an alternative for the coordinate measuring machines in applications which allow for a lower measuring accuracy [14, 20]. For example, the accuracy of a mobile measuring arm was discussed in work [4], whose authors performed tests consisting in an evaluation of the representation of the nominal shape with the use of the arm as well as the coordinate machine. The 3D scanning technique, also with the use of measuring arms, is mainly applied for the product end quality control [14, 22, 28]. These measurements are the most often based on the assessment of the shape errors of the determined contour and surface [5, 14]. The available literature more and more frequently discusses applications of this kind of methods for the measurement, control and evaluation of the state of swaging tools. An example of such application of the 3D scanning method [13, 23] is the use of an optical scanner for the determination of the shape errors of the given surface and next, on the basis of the obtained data, providing the geometrical specification for the process of rebuilding. Another application of the 3D scanning method with the use of scanners [5, 6, 8, 19, 24] is the use of the analysis of the shape errors of the given surface for the evaluation of the wear of the forging tools – nitrided or coated with hybrid layers. These analyses consist in a comparison of the images obtained from the scanning of the new forging tool before its operation or a reference CAD model and next of the same tool after the forging process, by way of determining the shape errors of the analyzed surface.

The aim of the work is to prove the possibilities of applying a measuring arm with an integrated laser scanner in the control and analysis of the wear process of the selected forging tool, i.e. an upper filler (used in the second operation of die forging), on the basis of the measurement of the shape changes of cyclically sampled forgings, directly during the production process.

2. Test subject

In the industrial production process, controlling the quality and state of the applied tools is quite difficult and troublesome, as it is often connected with the necessity of the tools' disassembly, which causes long production intervals. A continuous analysis of the quality and life of the tools directly on the production line is important due to the fact that the flaws occurring on the tool are 'automatically' transferred to the forging. Often, even a minor defect or wear of the tool in the first operation (preliminary forging) of the multi-operational forging process makes obtaining a forging which meets all of the requirements difficult in the last operation, i.e. the finishing (calibrating) forging [7]. That is why a useful solution can be the possibility to evaluate the shape changes of the tool impressions in the particular operations, performed based on a cyclic control of the tools (measurement through scanning), without the necessity of their disassembly from the forging press (Fig. 1).

Fig. 1. View of the press together with assembled forging tools for gearbox lid forging and the measurement by way of filler scanning with the use of the measuring arm, performed directly on the production line

The industrial process of hot die forging of a gearbox lid forging (Fig. 2), used in one of motor-car brands was chosen for the analysis. This element is assembled at the output of the transmission drive shaft. The process is realized in three operations on the P-1800T press. The consecutive operations are: upsetting, blocking and final forging. The forged material is C45 steel with the billet dimensions: diameter 55 cm, length 95 mm, weight 1,77 kg. The initial temperature

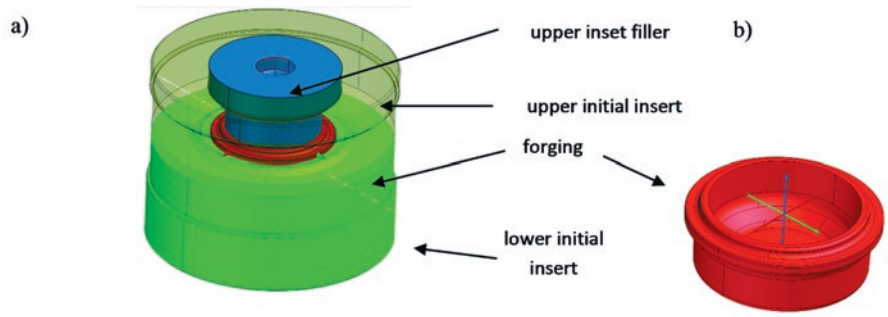


Fig. 2. Schematics of blocking for the produced element: a) tool set, b) forging

of the billet is 1150 °C. The forging tools were made of tool steel for hot operations – WCL (1.2343). After the thermal treatment, the tools assigned for the second and third operation additionally undergo thermo-chemical treatment (nitriding), with the purpose to increase their abrasive wear resistance. During their operation, these tools undergo very high cyclic thermal (from 80 to 600°C) and mechanical (0-800 MPa) loads. As a result of such extreme working conditions, their wear rate is high.

A detailed analysis was performed of one of the tools used in the second forging operation – the upper insert filler (Fig. 2a). The tools used in the second operation (blocking) undergo the highest loads, due to the fact that in this operation, the forging is formed to the highest extent. For the selected filler, tests were conducted in order to determine its tool life directly on the production line (without its disassembly from the forging press), on the basis of the measurements of the forgings' shape changes during the forging process (Fig. 2b) by way of scanning. At the time of the short technological intervals in the process, direct measurements were performed of the (analyzed) tool itself, with the purpose to verify the results of the forging measurement in relation to the actual changes in the tool (Fig. 3).

The average tool life is relatively low and equals about 6700 forgings, which is connected with the occurrence of many destructive mechanisms simultaneously. The average tool life for the remaining tools in this process equals about 9000 items.

The most frequent defects of this tool include: plastic deformations, mechanical microcracks, thermal microcracks, wear through abrasion, fracture and chipping off of tool parts. At the time of the whole production process, the shape changes proceed with varying intensity at the particular stages of operation, which additionally complicates the analysis of the tool life as well as the actions connected with this measurement [9].

Fig. 3. Analyzed upper insert filler: a) new – before operation, with the forging from the beginning of operation, b) worn – after 12 500 forgings, with the forging from the end of tool exploitation

3. Tool and measuring method description

In the measurements of the shape changes of the forgings and the filler, the measuring arm ROMER Absolute ARM 7520si was used (Fig. 4) together with the Polyworks 2014 software and the Real Time Quality Meshing technology. The measuring arm is equipped with seven rotation axes. Each axis includes an absolute encoder which measures the rotation angle of the kinematic pair, which does not require initialization or heating. The arm enables contact

Fig. 4. Measuring station for laboratory tests of forgings with the ROMER Absolute ARM 7520si measuring arm equipped with an integrated laser scanner

measurements as well as optical ones. For the tests, the non-contact measuring method was chosen with the use of the laser scanning system RS2 integrated with the arm. The integrated RS2 scanner characterizes in the possibility to collect up to 50000 points/s for 1000 points on the line at the linear frequency of 50 Hz.

In order to determine the possibility of applying 3D scanning in the tests of tool shape changes in the analyzed process, it was necessary to determine the optimal surface parameters for the comparison of the measuring data obtained from scanning the filler before and after operation. The selection of the optimal surface fragments aimed at the possibility to obtain the proper data comparison method and thus to assure the most accurate measuring results for the desired shape change.

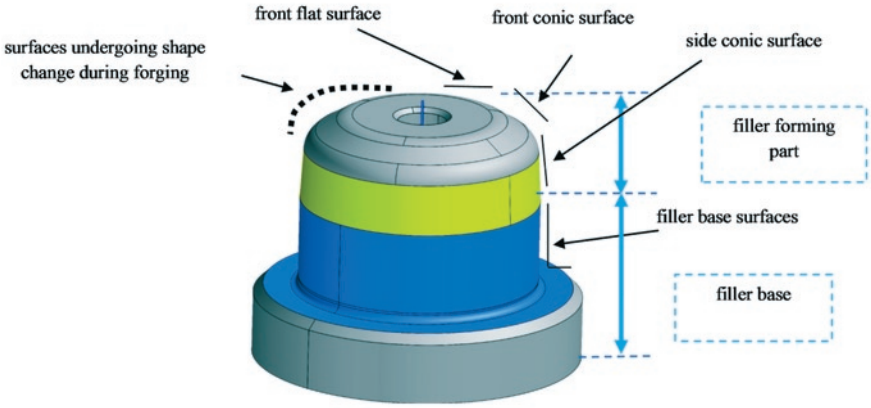


Fig. 5. Functional division of the filler surface with marked surfaces which change their shape during forging

In the data approximation process, the best-fit method is used which applies the Gauss approximation algorithm consisting in calculating the mean element. The calculations involve the use of the principle of the least squares of the deviations of the nominal points from the calculated ones. Such a method can cause the results to be encumbered with a large error, when an inappropriate surface for approximation is selected, especially in the case of significant shape errors. The authors also analyzed the effect of the selected reference surface during scanning on the obtaining of results which most correspond to the actual wear value. To that end, verification of the measurement results was performed by means of the laser scanner (in selected points) by way of validation (measurement) on a coordinate machine. The performed verification confirmed the possibility to obtain the declared accuracy of the RS2 linear scanner at the level below 0.058 [mm].

The analyzed filler can be divided into the forming part and the base (Fig. 5). The filler base is responsible for the proper basing of the surfaces forming the forging in relation to the remaining parts of the built-up forging tool.

The surfaces forming the forging belonging to the forming part of the filler (front flat, front conic, side conic), which are responsible for fulfilling the geometrical properties of the end forging product in the second operation, as a result of wear, change their shape together with the number of the produced forgings.

4. Verification of proposed method of filler wear analysis

First, scanning with the use of the Real Time Quality Meshing technology was performed. The result of the particular measurements by means of the 3D scanning technology is a cloud of points. Next, based on the obtained cloud of points, the program opened a polygonal surface consisting of elementary triangles, representing the shape of the measured object.

Fig. 6 shows a comparison of the shape changes in the forgings after a given number of cycles and after the filler's wear, for which, in order to verify the proposed method, measurements with the use of the measuring arm together with the scanner were performed during the short technological intervals in the forging process. The presented comparison of the forging scans and the corresponding tools (Fig. 6) is made in the middle and towards the end of the forging process (filler operation), in order to exhibit the increasing areas of wear.

In the analysis of the growth of the volume in the case of the forgings as well as its loss, one can assume that they are both at a similar level. For the forging-filler set of 6000 items, the maximal material growth in the normal direction for the forging equals about +1,5mm, whereas for the tool, the loss is at the level of about 1,4mm. For the set of 12500 items, in turn, the maximal differences in the normal direction equal: +2,3mm for the forging and –2,4mm for the filler. Certain differences in the obtained results can be caused by the particular measuring accuracy of the scanner as well as the temperature conditions of the scanned elements (filler temperature: about 120-150ºC, forging temperature: ambient).

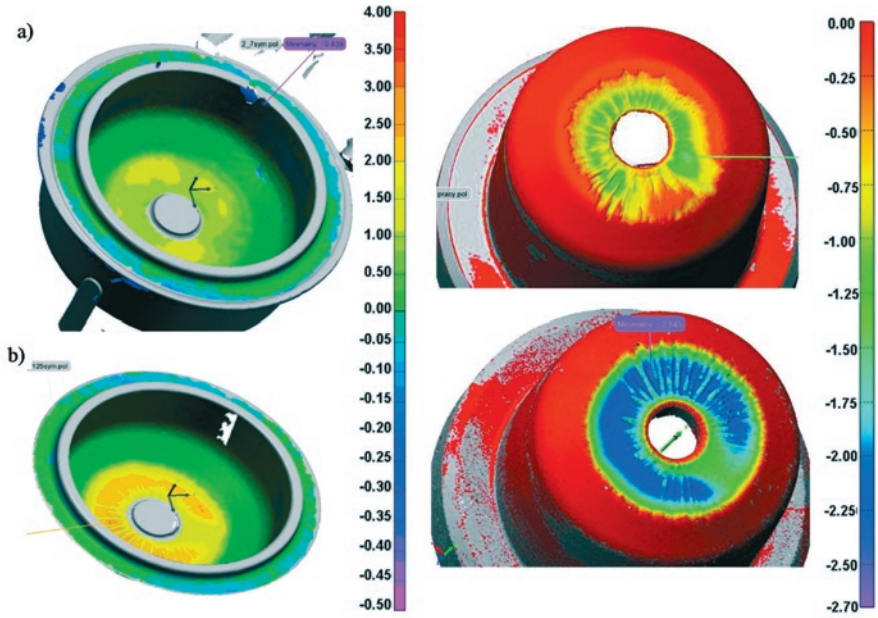


Fig. 6. Comparison of scans of the inner part of the forgings and the corresponding tools, in the form of quantitative shape changes in relation to the CAD models of the nominal forging and the tool, every: a) 6000, b) 12500 items

5. Test results

With the purpose of a more thorough analysis of the filler's wear, on the basis of the measurements of the shape changes of the forgings, scanning was performed for a forging series (every 1000 items), produced by the analyzed filler in the total amount of over 12500 items. Due to the fact that the forging's temperature, after the forging operation, is about 1000ºC, which makes it impossible to tag the forgings with the purpose of further identification, a specially prepared container with partitions was used (Fig. 7), in which the forgings were placed by the operator.

Fig. 8 presets a comparison of selected scans of the inner part of the forgings, in the form of quantitative shape changes in relation to the CAD model of the nominal forging.

As it was to be expected, the images with the scans for the increasing number of forgings point to a proceeding wear of the tool, i.e. the filler. The wear (on the basis of the volume change) is localized in the central part, in the vicinity of the knock-out opening (grey circle) and it is irregular. This probably results from the manner of providing the

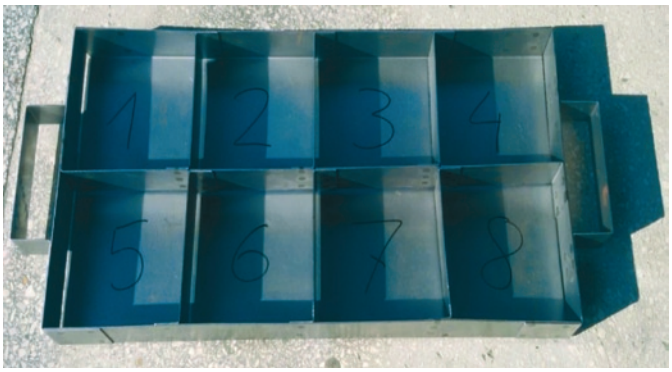


Fig. 7. Box for placing the forgings

cooling and lubricating agent (inappropriate position of the lubricating nozzles). And so, the proposed method would allow for an early detection of this kind of disadvantageous changes in the lubricating process and the appropriate correction of the position of the lubricating nozzles, which would surely prolong the life of the tools.

Fig. 9. Shows a diagram with a comparison of the material loss (volume changes) of the filler on the basis of the volume changes (adequate growth) for the consecutive forgings from a given series.

On the basis of the presented diagram (Fig. 9) resembling the classic wear curve (Lorenz curve), we can observe interesting relations and distinguish a few ranges (periods) of wear. And so, the wear of the analyzed filler based on the forging scan analysis increases very rapidly at the beginning of the forging process up to about 1000 items (period I). This is connected with the approximation of the whole system, in which we observe a transformation of the initial state of the outside layers of the elements of the filler grinding-in with the forging into the **optimal state**. After achieving the optimal state, that is over 2000 items, the state of so-called **normal operation** begins (period II), characterizing in a more or less stabilized level of intensity of the previously mentioned wear phenomena, which, in the analyzed case, goes up to about 8000 items. The volume change for this forging range equals from 900 to 3516mm^3, whereas for 8000 items up to the end of the tool operation (over 12500 items), the volume change equals merely 350mm^3. On this basis, we can conclude that the state of stabilized wear can be assumed to be between 7000 and 8000 forgings, which can be treated as the beginning of period III of the wear period. This state, for the analyzed tool, is present up to the moment of maximal exploitation, that is 12500 items, and it finishes as a result of exceeding the acceptable shape change of the tool, due to its elimination from further production. The classic Lorenz curve, towards the end of the period of normal operation, usually turns into the state of accelerated wear, which, unfortunately, cannot be observed in the analyzed case. This can be explained by the reduced pressure exerted on the tool as a result of the proceeding wear of the contact surfaces of the tool and the formed forging.

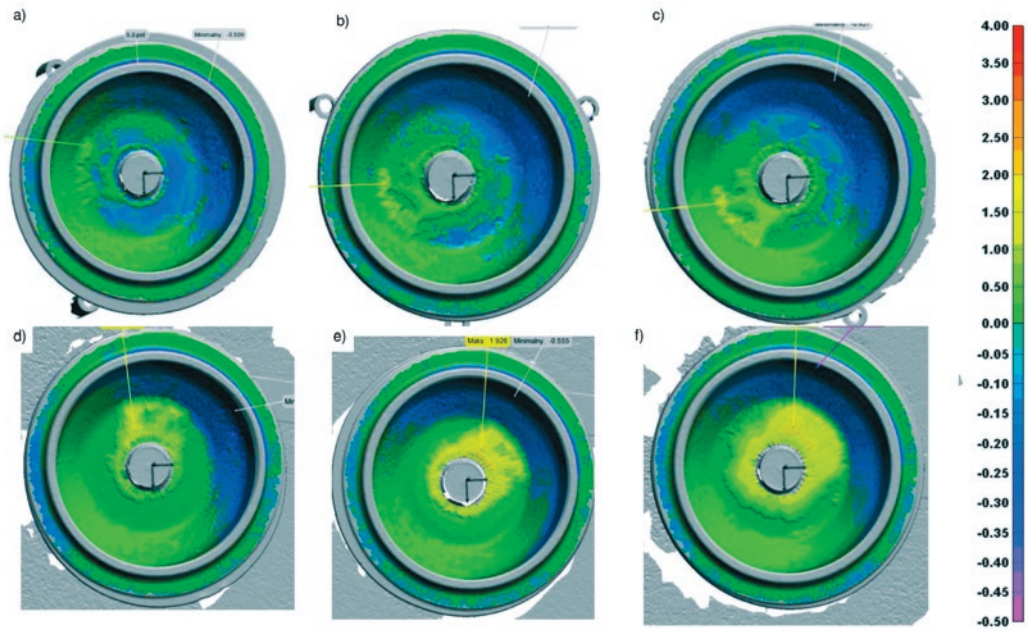


Fig. 8. Comparison of scans of the inner part of the forgings, in the form of quantitative shape changes in relation to the CAD model of the nominal forging, every: a) 2000, b) 3000, c) 4000, d)5000, e) 7000, f) 8000 items

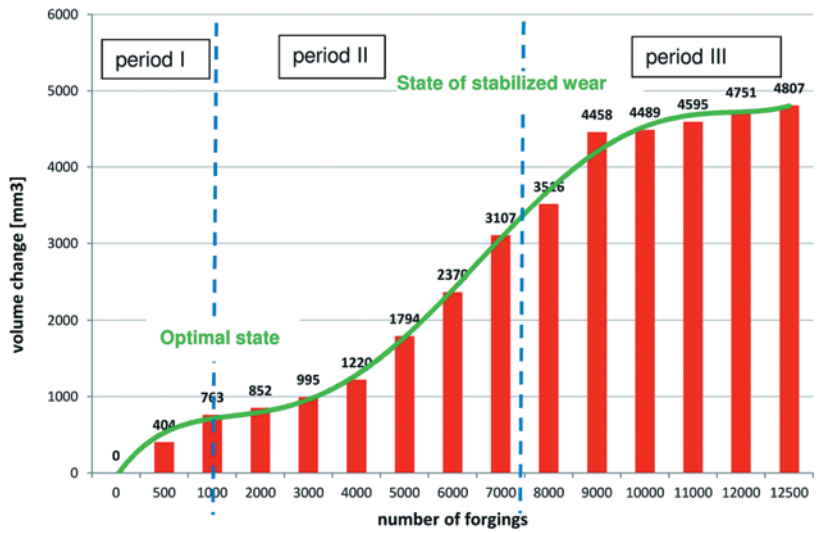


Fig. 9. Comparison of the material loss (volume changes) in the filler based on the volume changes on the surface of consecutive forgings

In order to identify the destruction phenomena and mechanisms concerning the forging tools, the authors performed a series of laboratory tests. As a result of a cyclic operation of high thermo-mechanical loads in the whole exploitation period, the tools undergo many destructive phenomena, occurring with different intensity and frequency. The tests revealed that, in the initial period of the forging process (about 400-1000 items.), on the analyzed tools, as a result of thermal fatigue, we can observe a primary, and together with the increasing number of forgings, secondary fatigue crack network (Fig. 10). The effect of this is a relatively

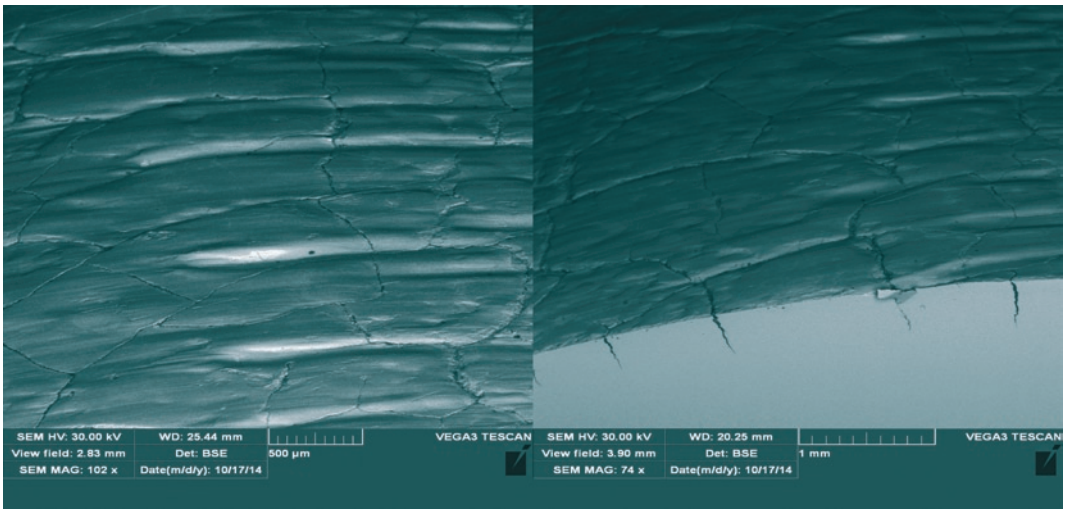


Fig. 10. Filler area (after 500 forgings) with primary and secondary crack network resulting from thermal fatigue, magnification 60x, scanning electron microscope TESCAN VEGA3

large loss (volume change at the level of 404mm^3, after about 400-500 forgings) of the tool material. What is more, the test results concerning the properties of the nitrided layer for this type of tools showed that, in the case of abrasive wear as the dominating destructive mechanism, in the initial period (400-500 forgings), we can observe a very rapid „rubbing off" of the nitrided layer.

In turn, with a larger number of forgings (Fig. 11), after the mentioned stabilized wear has been reached, we can see numerous grooves (in the vicinity of the filler opening) as well as smaller and larger cracks, while it is difficult to observe the characteristic network of cracks originating from thermal fatigue.

The value of stabilized wear for the analyzed tool (up to 7000 items) is probably connected with the intensified abrasive wear, additionally supported by the detaching hard particles (nitrides from the nitrided layer and oxides of the tool material) in the surface layer of the tool, working as abradant.

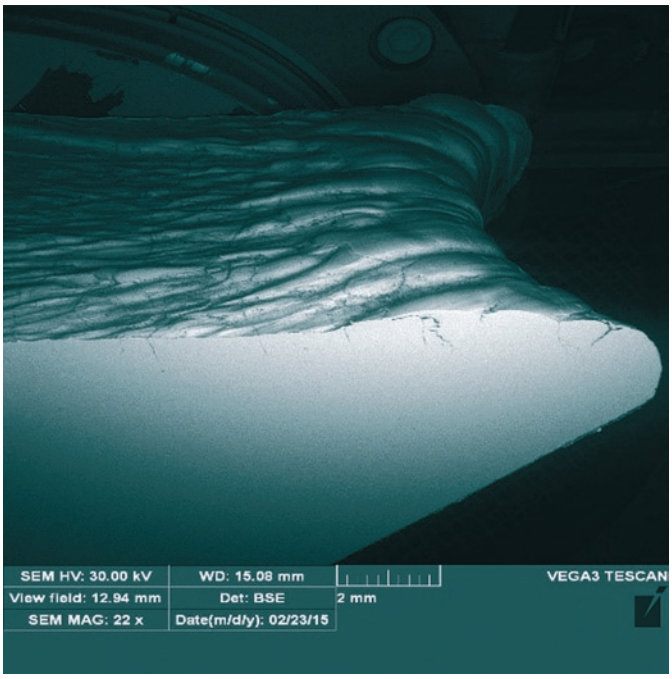


Fig. 11. Filler area in the vicinity of the opening (after 9000 forgings) with visible grooves ("washing out" of material) and microcracks, magnification 22x, scanning electron microscope TESCAN VEGA3

As regards the range from 8000 to 12500 forgings, the almost constant wear value at the level of about 4440 to 4800mm3 is connected with the reduced pressure on the worn tools (as compared to new ones), which was verified by way of numerical modeling of the forging process.

At present, the authors are conducting research concerning the analysis of the occurrence of the mentioned saturation level with the particular number of forgings for forging tools used in other hot die forging processes.

6. Summary

The performed research with the use of a measuring arm together with an integrated laser scanner for the analysis of the filler wear, on the basis of the measurements of the shape changes of consecutive forgings (directly on the production line) proved the validity of applying new measuring technologies in order to directly analyze the quality and change of the tool shape (without disassembling the instrumentation from the forging unit). Owing to this, such an analysis was possible directly during the production process.

The analysis of the volume growth of consecutive forgings based on the measurements makes it possible to precisely determine the material loss of the forging tool in the consecutive stages of its operation. This is proved by the full correlation between the results of the measurements of the volume changes of the series of the increasing number of produced forgings and the tool in the middle and towards the end of its work.

The innovative approach to evaluating the current state of the forging tool proposed by the authors makes it possible to make decisions about a prolongation or shortening of the time of its operation based on the actual (current) wear, and not on the basis of the strictly set tool life (maximal number of produced forgings). This allows for an optimal use of the given tool, with the preservation of the possibly highest quality of the produced forgings.

This method makes it possible to eliminate the human factor from the process of making decisions by way of determining the geometrical tolerance for the changing shape of the tools correlated with the tolerance and shape of the forgings.

What is more, the proposed method would allow for an early detection of the occurrence of disadvantageous changes during the forging process, such as inappropriate lubrication or a premature fracture of the tool and thus it would enable a fast reaction, that is a proper correction of the position of the lubricating nozzles or an immediate removal of the tool from the production, and retooling.

The exhibited advantages of the proposed new approach to the analysis of the state of the tool based on the measurements of the shape changes of the forgings, with the use of a laser scanner, would surely translate to a prolongation the tool life of the forging instrumentation and a significant reduction of the production costs.

References

1. Bills P.J, Racasan R, Underwood R.J, Cann P, Skinner J, Hart A.J, Jiang X, Blunt L. Volumetric wear assessment of retrieved metal-on-metal hip prostheses and the impact of measurement uncertainty. Wear 2012; 274–275: 212-219, http://dx.doi.org/10.1016/j.wear.2011.08.030.
2. Carmignato S, Spinelli M, Affatato S, Savio E. Uncertainty evaluation of volumetric wear assessment from coordinate measurements of ceramic hip joint prostheses. Wear 2011; 270(9-10): 584-590, http://dx.doi.org/10.1016/j.wear.2011.01.012.
3. Farooqui S.A, Morse E.P.: Methods and artifacts for comparison of scanning CMM performance. Journal of Computing and Information Science in Engineering 2007; 7(1): 72-80, http://dx.doi.org/10.1115/1.2709928.
4. Gąska A, Olszewska M. Ocena odwzorowania kształtu za pomocą współrzędnościowego ramienia pomiarowego wyposażonego w głowicę optyczną. Postępy Nauki i Techniki 2011; 7: 37-43.
5. Gronostajski Z, Hawryluk M, Kaszuba M, Niechajowicz A, Polak S, Walczak S, Jabłoński D. Die profile optimization for forging constant velocity joint casings. Archives of Metallurgy and Materials 2011; 56(2): 551-558, http://dx.doi.org/10.2478/v10172-011-0059-z.
6. Gronostajski Z, Hawryluk M, Niechajowicz A, Zwierzchowski M, Kaszuba M, Będza T. Application of the scanning laser system for the wear estimation of forging tools. Computer Methods in Materials Science 2011; 11(2): 425-431.
7. Gronostajski Z, Hawryluk M. The main aspects of precision forging. Archives of Civil and Mechanical Engineering 2008; 8(2): 39-55, http://dx.doi.org/10.1016/S1644-9665(12)60192-7.
8. Gronostajski Z, Kaszuba M, Hawryluk M, Marciniak M, Zwierzchowski M, Mazurkiewicz A, Smolik J. Improving durability of hot forging tools by applying hybrid layers. Metallurgy 2015; 54(4): 687-690.
9. Gronostajski Z, Kaszuba M, Hawryluk, M, Zwierzchowski M. A review of the degradation mechanisms of the hot forging tools. Archives of Civil and Mechanical Engineering 2014; 14(4): 528-539, http://dx.doi.org/10.1016/j.acme.2014.07.002.
10. ISO GPS 10360-4:2000 Geometrical Product Specifications (GPS) – Acceptance and Reverification Tests for Coordinate Measuring Machines (CMM) – Part 4: CMMs used in Scanning Measuring Mode. Norma.
11. Kawalec A, Magdziak M. Metoda obliczania krzywej offset. Przegląd Mechaniczny 2011; 70(7-8): 26-30.
12. Kawalec A, Magdziak M. Wpływ metody dopasowania na wyniki pomiarów pióra łopatki, XI Forum Inżynierskie ProCAx, 2012; 2.
13. Kontrola jakości odkuwek i matryc - archiwizacja i regeneracja matryc. Materiały informacyjne: ITA-polska; http://www.ita-polska.com.pl.
14. Kuş A. Implementation of 3d optical scanning technology for automotive applications. Sensors 2009; 9: 1967-1979, http://dx.doi.org/10.3390/s90301967.
15. Langton D.J, Sidaginamale R.P, Holland J.P, Deehan D, Joyce T.J, Argol A.F, Meek R.D, Lord J.K. Practical considerations for volumetric wear analysis of explanted hip arthroplasties. Bone & Joint Research 2014; 3: 60-68, http://dx.doi.org/10.1302/2046-3758.33.2000249.
16. Li F. X, Longstaff A, Fletcher S, Myers S. Integrated tactile and optical measuring systems in three dimensional metrology. Computing and Engineering Researchers' Conference, University of Huddersfield, Mar 2012; 1-6.
17. Lord J.K, Langton D.J, Nargol A.V.F, Joyce T.J. Volumetric wear assessment of failed metal-on-metal hip resurfacing prostheses. Wear 201; 272(1): 79-87.
18. Lu Z1, McKellop HA. Accuracy of methods for calculating volumetric wear from coordinate measuring machine data of retrieved metal-on-metal hip joint implants. Proceedings of the Institution of Mechanical Engineers, part H 2014; 228(3): 237-49, http://dx.doi.org/10.1177/0954411914524188.
19. Lulkiewicz J, Szkudelski S, Pachutko B. Kompleksowe badania matryc kuźniczych. Projektowanie i Konstrukcje Inżynierskie 2015; 3(90):16-19.
20. Machácek, P, Tomícek J. Application of laser scanning in reverse engineering and prototype manufacturing. WTP 2010; 1(21): 35-44.
21. Magdziak M. Metoda pomiaru powierzchni swobodnych na obrabiarce sterowanej numerycznie. Praca doktorska, Wydział Budowy Maszyn i Lotnictwa, Politechnika Rzeszowska, Rzeszów 2012.
22. Marton E, Pizzolon F. Dimensional control during forging. Innovative 3d laser measuring systems, Forge Applications 2010; (1).
23. Measurement of a forging die for tooling corrections. Materiały informacyjne: AICON_3D_SYSTEM; http://www.aicon3d.com/fileadmin/user_upload/produkte/ en/breuckmann_Scanner/ 01_PDF_IuT/Forging_die_measurement_Web.pdf.
24. Pachutko B, Ziółkiewicz S. Investigation of the wear processes of dies for forging building anchors basing on metallographic examinations. Obróbka Plastyczna Metali 2012; 23(4): 277-293.
25. Peng X., Wang L., Yang W., Zhao Y., Li D., Jin Z., Fan B., Dong S., Zhang S.: Optimized adaptive control for evaluation of artificial hip joint volumetric wear by coordinate measuring. Hsi-An Chiao Tung Ta Hsueh/Journal of Xi'an Jiaotong University 2014; 48(8): 128-135.
26. Salah Hame, R. A. Influence of fitting algorithm and scanning speed on roundness error for 50 mm standard ring measurement using CMM. Metrology and Measurement Systems 2008; 15(1): 33-53.
27. Ulatowski K. Ramiona pomiarowe – rynek i zastosowania. Magazyn Przemysłowy 2013; http://www.magazynprzemyslowy.pl/zarzadzanie-i-rynek/Ramiona-pomiarowe-rynek-i-zastosowania.
28. Weckenmann A, Weickmann J. Optical inspection of formed sheet metal parts applying fringe projection systems and virtual fixation. Metrology and Measurement Systems 2006; 13(4): 321-334.

Zbigniew GRONOSTAJSKI
Marek HAWRYLUK
Marcin KASZUBA
Jacek ZIEMBA
Department of Mechanical Engineering
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

E-mails: zbigniew.gronostajski@pwr.edu.pl, marek.hawryluk@pwr.edu.pl, marcin.kaszuba@pwr.edu.pl, jacek.ziemba@pwr.edu.pl

Article citation info:
ROMAŃSKI L, BIENIEK J, KOMARNICKI P, DĘBOWSKI M, DETYNA J. Operational tests of a dual-rotor mini wind turbine. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 201–209, http://dx.doi.org/10.17531/ein.2016.2.7.

Leszek ROMAŃSKI
Jerzy BIENIEK
Piotr KOMARNICKI
Marcin DĘBOWSKI
Jerzy DETYNA

OPERATIONAL TESTS OF A DUAL-ROTOR MINI WIND TURBINE

BADANIA EKSPLOATACYJNE DWUWIRNIKOWEJ MINI ELEKTROWNI WIATROWEJ*

The article presents the results of wind-tunnel tests and field studies of a mini dual-rotor wind turbine. The first stage involved testing of an open-circuit wind tunnel built with the aim of performing laboratory tests. The coefficient of uneven air stream distribution at a rated speed was 1.7%, while the index of turbulence intensity in the entire measurement range was between 1.2 and 1.8%. The mini wind turbine was equipped with rotors with new design blades. Compared to the blade designs used in mini wind turbines available on the market, the blades used in the present study were characterized by an efficiency of 0.28. The results of performance tests in the wind tunnel were evaluated statistically using Pearson correlation coefficients and Spearman's rank. We examined the relationship between a dependent variable (power P) and independent variables (average air stream speed V, incidence angle of the blades of the first rotor α_1, incidence angle of the blades of the second rotor α_2, and the distance between the rotors l). The analysis showed, as expected, that the strongest correlation was between power and speed of the air stream. While incidence angles of the two rotors also affected the turbine's power, no such effect was observed for changes in the distance between the rotors. Field tests confirmed the findings and observations made in the wind tunnel.

***Keywords*:** *operation, mini dual-rotor wind turbine, wind tunnel.*

W artykule przedstawiono wyniki badań dwuwirnikowej mini elektrowni wiatrowej przeprowadzone w tunelu aerodynamicznym oraz w terenie. W pierwszym etapie testowano zbudowany w celu przeprowadzenia badań laboratoryjnych tunel aerodynamiczny o konstrukcji otwartej. Wyznaczony współczynnik nierównomierności strugi powietrza przy prędkości nominalnej wynosił 1,7%, natomiast wskaźnik intensywności turbulencji w całym zakresie pomiarowym zawierał się w granicach 1,2-1,8%. Budując mini elektrownię wiatrową wyposażono ją w wirniki w których zastosowano nową konstrukcję łopat. Zastosowane łopaty w porównaniu do zbliżonej konstrukcji łopat stosowanych w mini elektrowniach dostępnych na rynku charakteryzowały się sprawnością wynoszącą 0,28. Po wykonanych badaniach eksploatacyjnych w tunelu aerodynamicznym uzyskane wyniki poddano ocenie statystycznej z wykorzystaniem współczynników korelacji liniowej Pearsona oraz rangi Spearmana. Zbadano zależności między zmienną zależną (moc P) oraz zmiennymi niezależnymi (średnie prędkości strugi powietrza V, kąt zaklinowania łopat pierwszego wirnika α_1, kąt zaklinowania łopat drugiego wirnika α_2, odległości pomiędzy wirnikami l). Na podstawie analizy, zgodnie z oczekiwaniem, stwierdzono, że najsilniejsza korelacja występuje w odniesieniu do prędkości strugi powietrza. Wpływ na moc mają także kąty zaklinowania na obu wirnikach, natomiast nie stwierdzono takiego wpływu w przypadku zmian odległości pomiędzy wirnikami turbiny. Badania w terenie potwierdziły ustalenia i spostrzeżenia poczynione w tunelu aerodynamicznym.

***Słowa kluczowe**: eksploatacja, mini elektrownia wiatrowa dwuśmigłowa, tunel aerodynamiczny.*

1. Introduction and objectives

According to the formula for power generated by wind turbines, the speed of the incoming air stream working on the turbine rotor (speed in the third power) and the diameter of the rotor itself (second power) have the greatest impact on power. In the first case, excluding a situation when the wind turbine is running in a diffuser, we are entirely dependent on the forces of nature, because we have virtually no effect on wind power. In the second case, it seems that all simple solutions have already been exhausted, because enlarging the diameter of rotors above 140 m is connected with great technological problems and a very large, disproportionate increase in manufacturing costs [19]. A cheaper option, which additionally affects the efficiency of the turbine, is to improve the design of the blades themselves. Their geometry and size have to be changed in such a way as to generate on them an increasing lift force during the rotation of the rotor at the same wind speed [3, 10, 16, 26].

Recently many designers have been working on the problem of additional uses of the energy of the air stream when it is already out of the wind turbine rotor in its post-action phase. An example of a design in which this "waste energy" can be utilized is a wind turbine with two rotors situated in the axis of the electric generator [18, 21, 22, 23, 24]. There are two versions of this solution (Figure 1).

In the first design (Fig. 1a), the assembly has two rotors rotating in the same direction. Shaft speed is increased by the use of transmissions. As the electrical generator is propelled by two rotors, the efficiency of the generator increases, which is especially observable in low wind conditions.

In this arrangement, the turbine may also work in the counter-rotating mode, which increases the efficiency of the assembly. The second concept features a structure (Fig. 1b) with two rotors positioned axially before the generator. Rotor assemblies intercept air mass flowing from one direction. The structure of the wind turbine is gearless and the individual components of the generator (stator, rotor) are driven by different turbines. Under the action of flowing air mass, the rotors with different blade pitch rotate in opposite directions. Consequently, the relative rotation speed of the rotor is greater than in

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

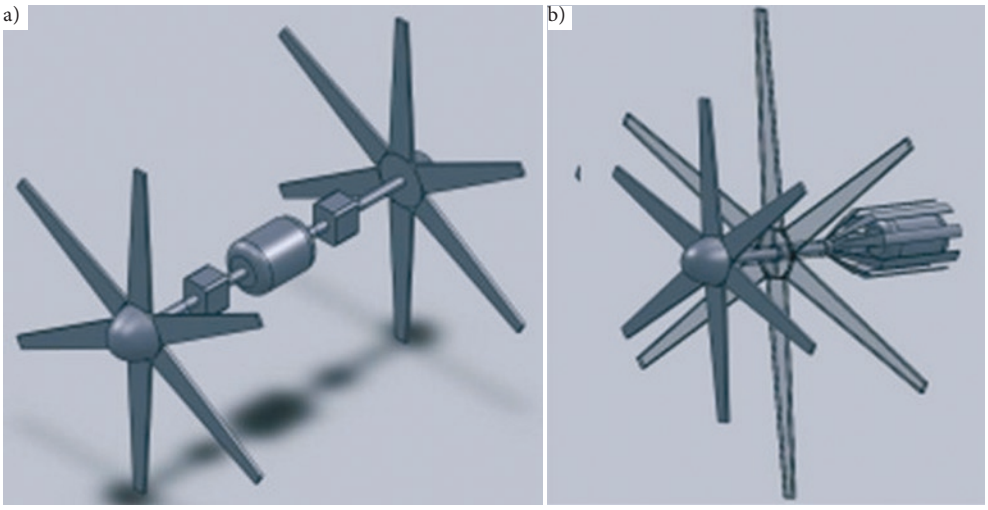


Fig. 1. The relative position of two rotors in wind turbines: a – rotors situated on both sides of the generator, b – rotors situated axially one behind the other [17]

single-rotor structures. As a result, the power generated by the turbine should increase.

Among other solutions, axial flux permanent magnet generators, also known as low-speed generators, are used in the construction of low-power wind turbines. Their use makes it possible to eliminate or reduce the mechanical transmission ratio. This reduces noise and costs of the assembly and increases its efficiency. Low-speed generators are made as cylindrical or disc generators. An example of an axial flux generator with a coreless stator is shown in Figure 2. It has a relatively simple structure; since there is no loss in the stator core, its efficiency is increased [13, 14].

Based on the literature and the authors' experience, a dual-rotor mini wind turbine was designed and built, in which one rotor drove the propeller and the other one drove the stator of the generator in the opposite direction. Operational tests were carried out in the wind tunnel and in natural conditions in the foothill region of the Sudetes.

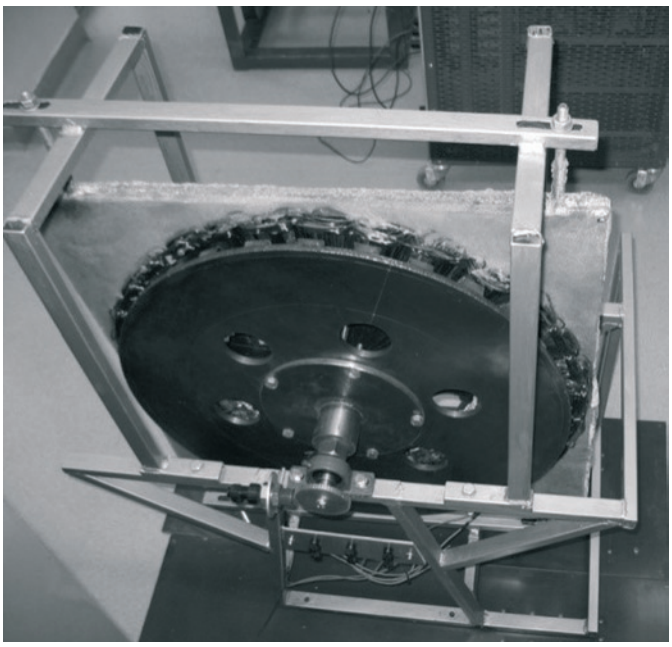


Fig. 2. A model of a low-speed axial flux generator [14]

2. Material and method

In the present study, a wind tunnel was used as the basic test stand for evaluating the performance of different aerodynamic objects, including mini wind turbines. The tunnel was designed and built on the basis of information available in the literature [4, 5, 7, 8, 9] and counsel provided by specialists in the field from several research centres. The shape of the Witoszyński confusor used in constructing the tunnel was determined numerically [1, 6, 11]. The combined use of a stream straightener and the Witoszyński confusor yielded an increase in air velocity in the measuring chamber, giving a more concentrated and steady flow. The basic geometric dimensions of the tunnel were as follows: square inlet 2.5 x 2.5m large,

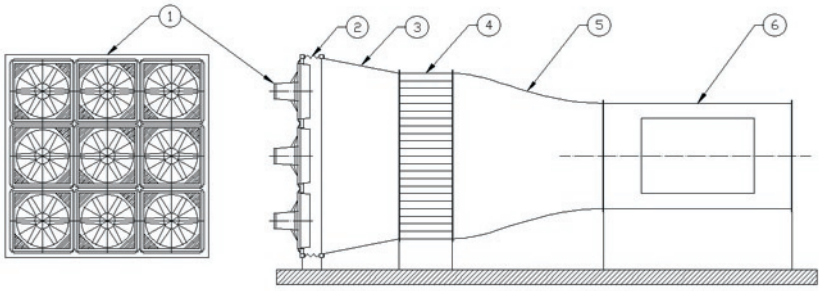


Fig. 3. Diagram of an open circuit wind tunnel: 1 – frame with fan section, 2 – vibration damper, 3 – square-section symmetrical contraction, 4 – honeycomb screen, 5 – the Witoszyński confusor, 6 – test chamber with an observation window [15]

total length 7m, and outlet – a measurement chamber with a diameter of 1.4 m (Fig. 3). The use of nine independently controlled axial fans, 2.2 kW power units, made it possible to achieve an air stream velocity in the measuring space of up to 17.5 $m \cdot s^{-1}$ at a dynamic pressure of approximately 200 Pa and to obtain a balanced stream velocity gradient in the cross-section of the measuring chamber. The fans were controlled through changes in the rotational speed of the rotors made by adjusting the frequency of power inverters in the range of 15–50Hz in 0.01 Hz increments [15].

The wind tunnel was tested when the test chamber was empty. Air flow testing in the tunnel was performed in two stages. In the first stage, the basic parameters of the stream were determined including pressure, velocity and qualitative indices of the tunnel such as the uniformity of velocity distribution in the test chamber and turbulence intensity indicators, which were considered as a function of the arithmetic mean of velocity. In the second stage, we investigated the effect of adjusting the operation of the individual fans on the distribution of velocity fields in the test section of the wind tunnel. Measurements of uniformity of stream velocity distribution in the test chamber were done using the traversing method according to the Polish standard PN-ISO 5221. The measurement points were arranged in the circular cross-section measuring channel on the basis of the Log-Chebyshev method recommended in the standard. In accordance with the aforementioned method, the channel was divided into concentric rings. Because the channel had a diameter greater than 0.25m, it was divided into five rings. The centre of the channel was measured in relation to

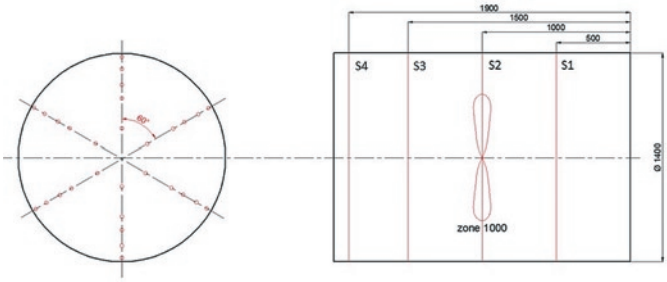


Fig. 4. Schematic layout of measurement points when traversing the cross section of the channel and the location of the measurement zones over the entire length of the chamber [15]

the axis of symmetry of the duct. In order to obtain a proper mean, an equal number of measurements was performed on each of the rings. 10 measurement points were located on 3 axes inclined with respect to each other at an angle of 60° (Fig. 4). Tests of the field distribution of flow rate were performed in 4 selected measurement areas: S1, S2, S3, and S4, whose distance from the edge of the outlet was 0.5, 1.0, 1.5, and 1.9 m, respectively [15].

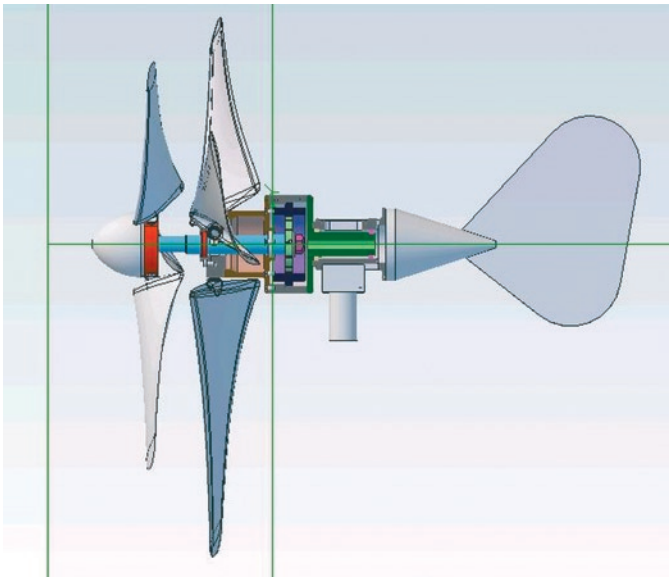


Fig. 5. Schematic diagram of the dual-rotor wind turbine built at the Institute of Agricultural Engineering of the Wrocław University of Environmental and Life Sciences

A diagram of the tested dual-rotor turbine is shown in Fig. 5. The functional model of the turbine was designed for quick installation of different shapes of blades, setting of their angle of attack and changing the position of the rear rotor relative to the front rotor. The incidence angle of the first blade (parameter α_1) could take six values between 125° and 150° in 5° increments. The incidence angle of the second blade (parameter α_2) could also take six values between 30° and 55° in 5° increments. The distance between the rotors was adjusted from 0.105 m to 0.14 m, in 5 mm increments.

As part of the research, new types of wind turbine blades were designed and built, which proved to be highly efficient aerodynamically. Appropriate airfoils of the blades were designed after a comparative analysis of numerous different shapes discussed in the literature [2, 12, 20, 23, 25], development of numerical models, and the authors' model-based tests. The blades were made using CNC technology and built with several layers of material, i.e., epoxy-glass composite, and their structures were based on two efficient aerodynamic airfoils.

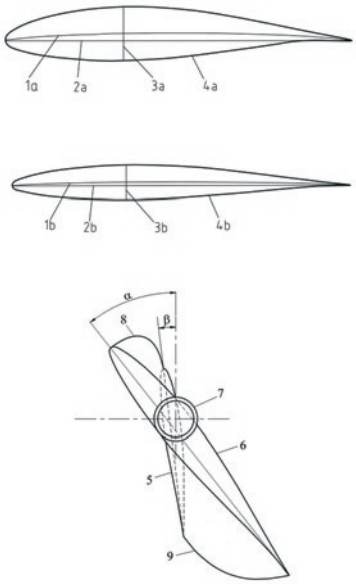


Fig. 6. The outlines of the initial airfoil NACA 63-415 and the end airfoil NACA 63-210 and view from the blade's mounting root: α - incidence angle of the initial airfoil, β - incidence angle of the end airfoil, 1a, 1b - camber line, 2a, 2b - chord line, 3a, 3b - airfoil thickness, 4a, 4b - airfoil curve, 5 - plane of the end airfoil NACA 63-210, 6 - plane of the initial airfoil NACA 63-415, 7 - mounting root, 8 - leading edge, 9 - trailing edge

Figures 6 present an exemplary airfoil of a twisted blade NACA 63-415-NACA 63-210.

An important element of any wind turbine is the generator, whose task is to produce electricity as a result of rotation of the blades. The tested mini-turbine used a solution in which the rotor and the stator rotated independently in opposite directions.

It was assumed that the generator rotors would rotate in opposite directions. Figure 7 shows a model of a prototype generator constituting an integral whole of the mini-wind turbine using a synchronous permanent magnet generator.

Operational tests of the mini dual-rotor wind turbine were carried out in two stages (Figure 8): the first stage was conducted in a wind tunnel (Figure 8a) with assumed constant operating parameters and the second stage featured field conditions (Figure 8b) of the village Wiry near Sobótka. Long-term observations of the Institute of Meteorology and Water Management show that the average wind speed in the region of Lower Silesia is 3.5 m·s^{-1} and the annual wind energy

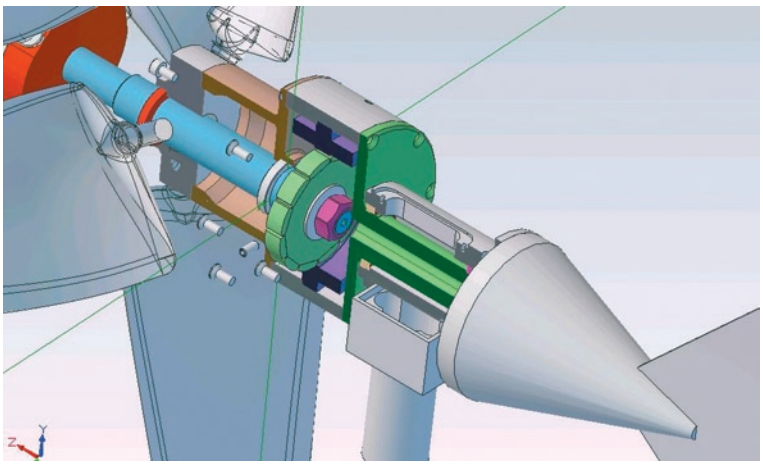


Fig. 7. Schematic diagram of the dual-rotor wind turbine built at the Institute of Agricultural Engineering of the Wrocław University of Environmental and Life Sciences

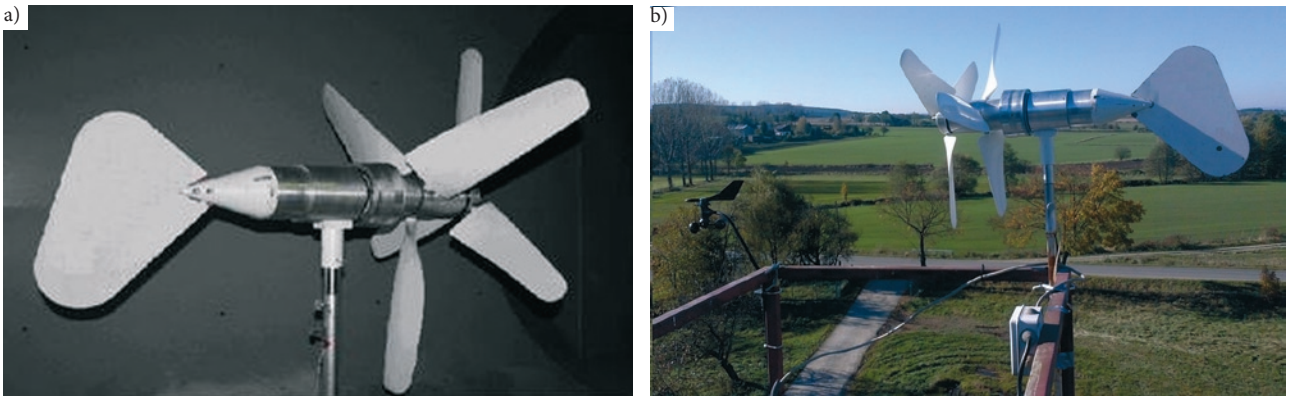


Fig. 8. View of the tested mini turbine in operation: a- in the wind tunnel, b- during the field tests

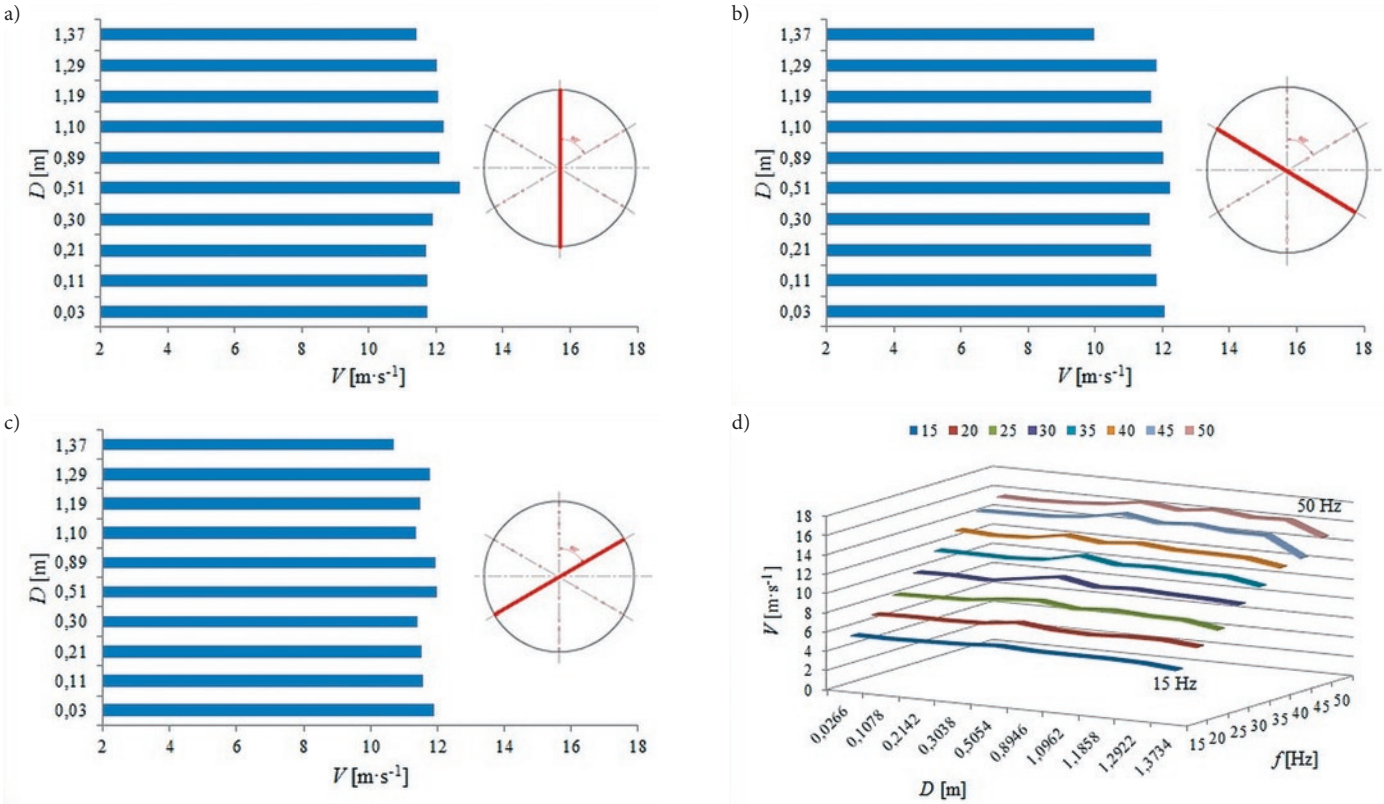


Fig. 9. Sample distributions of air stream speed in the measuring chamber for the setting of fan frequency at 35 Hz in the zone S2: a - measurements for the axis 90°, b - measurements for the axis 150°, c - measurements for the axis 210°, d - characteristics of speed distributions for the full range of fan settings

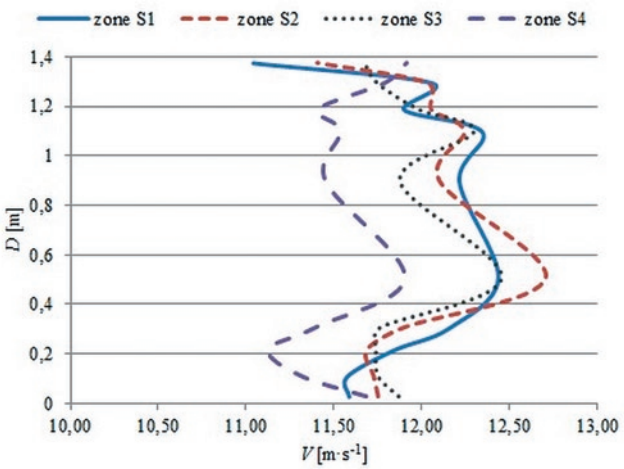


Fig. 10. Base air stream speed profiles for 4 different measurement zones for the setting of fan frequency at 35 Hz

is approx. 1,000 kWh·m $^{-2}$. Given the distribution of wind speeds in different months of the year, as well as wind energy conversion efficiency into electricity of approx. 20%, it was assumed that 1 m^2 of wind stream would yield only 0.53 kWh per day. The wind turbine was connected to a voltage regulator, a battery and a transmitter which collected data on the energy generated in the turbine (voltage, amperage) as well as data from a weather station installed near the wind turbine. For statistical analysis, the statistical software package Statistica v.10 by StatSoft was used.

3. Results

Due to the fact that wind turbines achieve their rated power at wind speeds of around 10-12 m · s^{-1}, the article provides an example of the air flow profiles measured in the measuring chamber corresponding to these speeds (Figure 9).

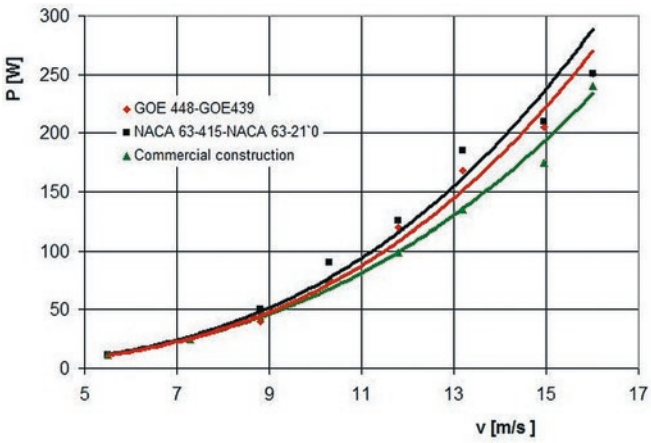


Fig .11. Power curve of the tested mini wind turbine using various blade designs

Table 1. Efficiency coefficients of the tested turbines

Turbine or airfoil name	Rotor diameter (m)	Wind power (W) at V = 12 m·s^{-1}	Power supplied by the rotor (W)	Efficiency reached (-)
JSW-750-12	0.75	471.4	97	0.21
GOE – 448/439	0.75	471.4	120	0.25
NACA 63-415/210	0.75	471.4	131	0.28

Table 2. Pearson and Spearman coefficients determining the statistically significant correlation between the dependent variable and the independent variables for the dual-rotor wind turbine

Independent variable	Pearson correlation coefficients	Spearman's rank correlation coefficient
V	0.90	0.95
α_1	–0.16	–0.14
α_2	0.17	0.13
l	0.02	0.02

The presented distributions corresponded to frequency settings of fans at 35 Hz. Based on the obtained basic distributions, it was observed that the air stream was characterized by slight irregularity and increased speed values in the vicinity of the axis of the measuring chamber.

The maximum local value of dynamic pressure was 198 Pa and was obtained by setting the frequency of the current supplying the fan motors at 50 Hz. Air stream velocity was calculated at 17.55 m·s^{-1}. Local minimum dynamic pressure values for the setting of 15 Hz amounted to 17 Pa, while the stream velocity was calculated at 5.14 m·s^{-1}. The impact of the distance of measuring areas S1-S4 on stream velocity profile is shown in Figure 10.

The designated index of turbulence intensity ranged between 1.2-1.8% [15]. The comparison of the performance of the mini wind turbine with removable rotors consisting of blades with different airfoils was based on a compilation of the plotted power characteristics.

Test results of the new design of the blades in relation to factory models are shown in Figure 11. Testing was performed in each case for 5 different incidence angles. The graph shows power curves which characterise the given blade at an "optimal" incidence angle. Both of the new designs of rotor blades proved to be more efficient than the factory rotors (Table 1).

The resulting efficiency of the mini wind turbine relative to the power of wind stream at the inlet to the rotor was 0.21 for the factory

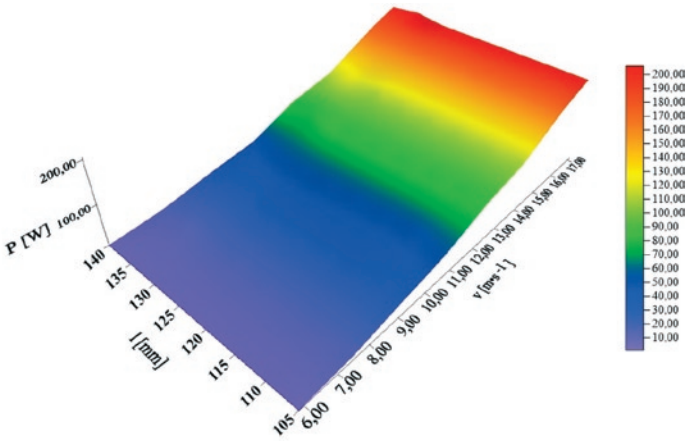


Fig. 12. Dependence of generated power on the distance between the rotors and the average air stream velocity for blade incidence angles α1 = 125° and α2 = 50°

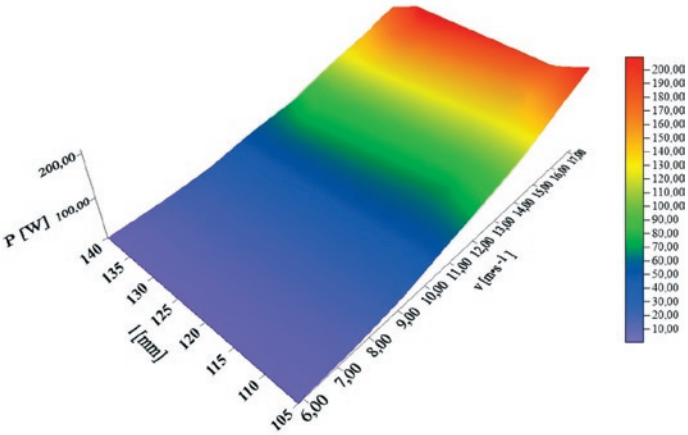


Fig. 13. Dependence of generated power on the distance between the rotors and the average air stream velocity for blade incidence angles α_1 = 130° and α_2 = 55°

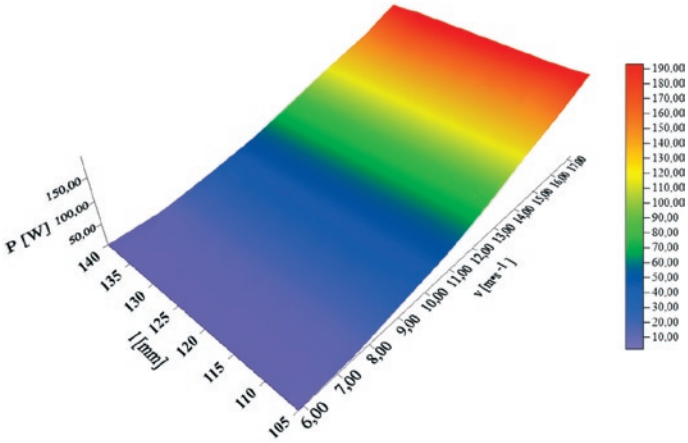


Fig. 14. Dependence of generated power on the distance between the rotors and the average air stream velocity for blade incidence angles α_1 = 125° and α_2 = 40°

turbine with 0.25 declared in the device's manual. In the case of the wind turbine with the GOE airfoil rotor blades, it was 0.25. When comparing

the power values of all the tested rotors at 12m·s^{-1}, which is recognized as the rated wind speed, it was observed that both designs generated more electricity: 19.5% for the GOE blade airfoil and 23.7% for

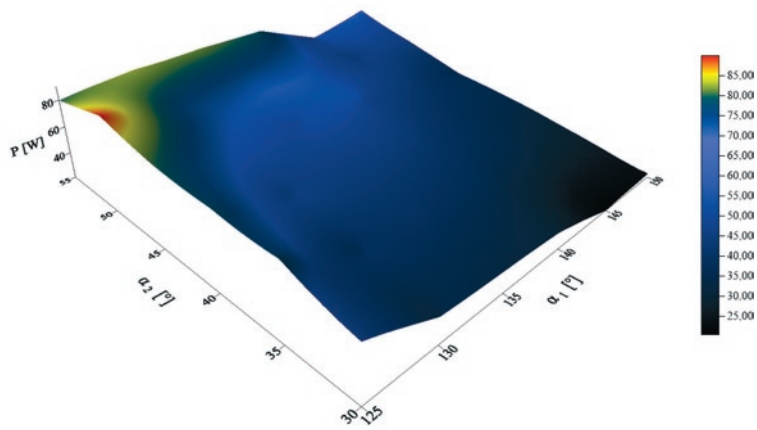


Fig. 15. Power characteristics as a function of blade incidence angles in particular rotors for the air stream velocity of 12.41m·s^{-1}

NACA, respectively. It is worth emphasizing that the new rotors have higher generated power values in the entire tested speed range. The rotor with NACA blades reached the efficiency of 0.28.

In order to obtain optimum settings of the rotors for field research, it was necessary to conduct a statistical analysis based on the results obtained in the wind tunnel tests. It was mainly aimed at determining the statistically significant relationship between the dependent variable (power *P*) and independent variables (average air stream velocity *V*, incidence angle of the blades of the first rotor α_1, incidence angle of the blades of the second rotor α_2, the distance between the rotors *l*) with the use of Pearson correlation coefficients and Spearman's rank. The results are shown in Table 2.

Based on the analysis, it can be concluded that the strongest correlation (in both cases coefficients ≥ 0.90) occurs for the air stream velocity, which is logical and confirmed by tests. Weak negative correlation occurs for the incidence angle of the blades of the first rotor. The negative coefficients suggest that reducing the incidence angle should result in an increase in power generated. Also a weak but positive correlation occurs for the incidence angle of the blades of the second rotor. Positive values of coefficients describe the increase in power generated along with increasing value of the incidence angle of the blades. Statistically significant lack of correlation for the two indices occurs in relation to the distance between the rotors. The absence of such correlations is illustrated by three exemplary three-dimensional graphs shown in the figures: 12 – 14. They were made as $P = f(l; v)$ for:

α_1=125° and α_2=50°; α_1=130° and α_2=55°; α_1=125° and α_2=40°.

The analysis of the data in these figures indicates that speed has a significant impact on the power generated by the wind turbine, as opposed to the distance between the rotors which has no such effect. Due to the compactness of the structure, this distance should be as small as possible while maintaining a safe enough distance between the rotors in the event of hurricane winds; when blades deviate from the perpendicular there can be no contact between the blades as it would result in their mutual damage. In order to determine the optimal incidence angle of the blades from the point of view of power delivered by the wind turbine, the dependencies resulting from the function $P=f(\alpha_1; \alpha_2)$ were presented in a graphic form. The figures were made for eight air stream speeds within the measuring range from 5.58 to 17.17 m·s^{-1}. Figure 15 shows an example of the stream speed graph for 12.41 m·s^{-1}. Selected optimal values of incidence angles of blades for specific values of air stream speed are summarized in Table 3.

A histogram (Fig. 16) of the sets of incidence angles makes it much easier to draw conclusions due to the fact that some of these same sets of blade incidence angles (α_1, α_2) allow the turbine to generate the most power at given air stream speeds.

After analysing the data relating to the power generated at a particular wind speed for the given incidence angles α_1 and α_2, it can be concluded that the optimal configuration values are 125° for the first rotor blades and 50° for the second rotor blades of the wind turbine. Other acceptable configurations:

α_1 = 130° and α_2 = 55°; α_1 = 125° and α_2 = 40°.

Table 3. Selected optimal incidence angles of blades for the air stream speed of 12.41 m·s^{-1}

v_{avg}	α_1	α_2
m·s^{-1}	°	°
5.85	125	40
7.53	125	50
9.17	125	50
10.82	125	50
10.82	130	55
12.41	125	50
14.08	125	40
14.08	125	45
14.08	125	50
14.08	130	55
14.08	135	55
14.08	140	55
15.82	125	50
15.82	130	55
17.17	130	55

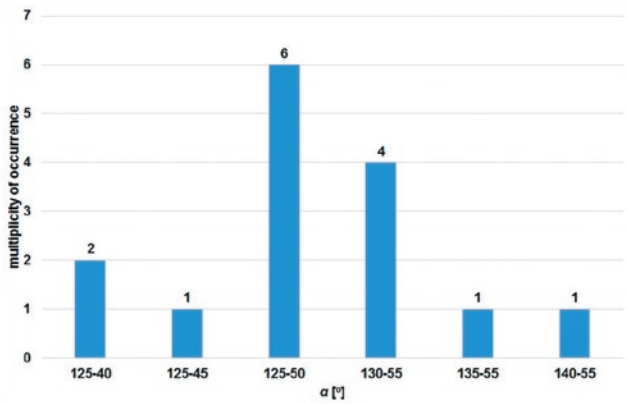


Fig. 16. The number of times a setting of incidence angles of the blades of the first and second rotor occurs after the dual-rotor wind turbine has reached its maximum power capacity for a given air stream speed

Specifications of the dual-rotor mini wind turbine are summarized in Table 4.

Figure 17 shows the results of simulations of the performance of the wind turbine in the wind tunnel under changing working conditions reflecting the conditions that occur during wind gusts. After analysing the electrical power generated by the plant it can be concluded that the moment of power capacity increase is recorded only after a 1-3 s delay relative to the moment the air stream (gust) speed begins to increase.

Table 4. Specifications of the dual-rotor wind turbine

No.	Turbine parameters	Wind turbine
1	Rated wind speed ($m \cdot s^{-1}$)	12.5
2	Maximum wind speed ($m \cdot s^{-1}$)	35
3	Minimum wind speed ($m \cdot s^{-1}$)	2.5
4	Rotor diameter (m)	0.75/0.75
5	Number of rotor blades (pcs.)	3/3
	Generator	
6	Generator Type	Synchronous
7	Excitation	Self-excited with permanent magnets
8	Voltage supply	Brush
9	Maximum electrical power (W)	400
10	Rated electrical power (W)	300
11	Voltage [V]	12
12	Voltage control system	built-in
13	Output voltage	AC
14	Stator diameter (m) ext./int.	0.15/0.092
15	Width of the stator (m)	0.055
16	Rotor diameter (m)	0.09
17	Rotor width (m)	0.02

The wind turbine, as previously mentioned, worked under varying wind conditions and was situated on a platform 10 meters above the ground. A record of the changes in power capacity for the period 5-6 November 2014 is given in Figure 18. In this period, approx. 12 distinct wind gusts shown in the Figure as peaks were recorded. A detailed record of the increase in the power delivered by the turbine generator during a gust of 23:36 hours, on 5 November 2014, is shown in Figure 19. In the analysed case, there was a problem concerning the recording as power was recorded at 1 s intervals while speed could only be recorded at 3 s intervals. Observations made during the wind tunnel tests were consistent with those from field studies, as the power increase followed an upward trend in wind speed and was delayed by 1-2 seconds. Registered wind speed of a 15-second gust was 9 $m \cdot s^{-1}$. It should be presumed that the actual speed was slightly higher, since the value of 9 $m \cdot s^{-1}$ is the average for 3 seconds, which should also explain the fact that the maximum of the power generated occurred a second earlier than the recorded maximum gust speed. During the field performance tests, wind speed rarely exceeded 7 $m \cdot s^{-1}$ and the average calculated for the examined period was 2.21 $m \cdot s^{-1}$ (Fig. 20). Its direction and repeatability is illustrated in Figure 21 featuring the wind rose. In accordance with the prevailing tendency in Lower Silesia these were mainly winds from the west or from its north-western

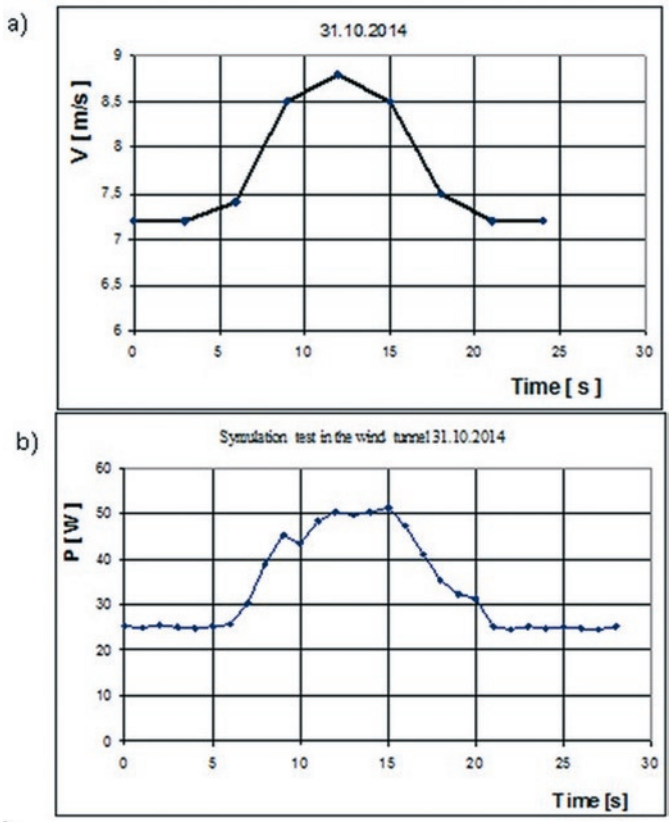


Fig. 17. Simulation of the performance of the wind turbine in the wind tunnel: a- air stream speed curve, b- generated power curve

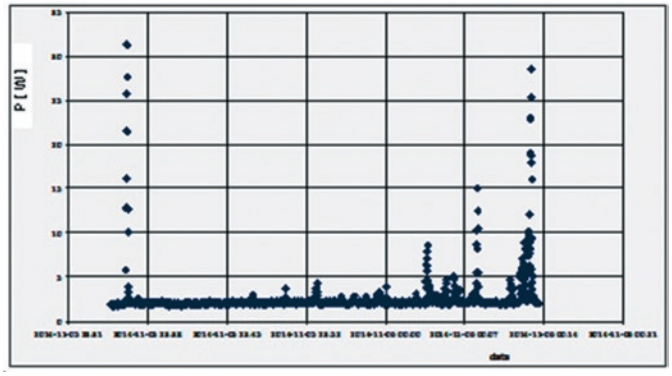


Fig. 18. Distribution of power generated by the dual-rotor wind turbine in the period: 5-6 November 2014

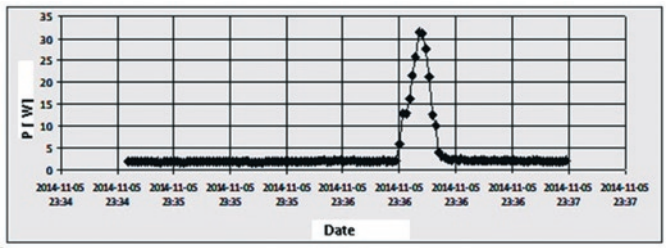


Fig. 19. Distribution of power generated by the dual-rotor wind turbine on 5 November 2014

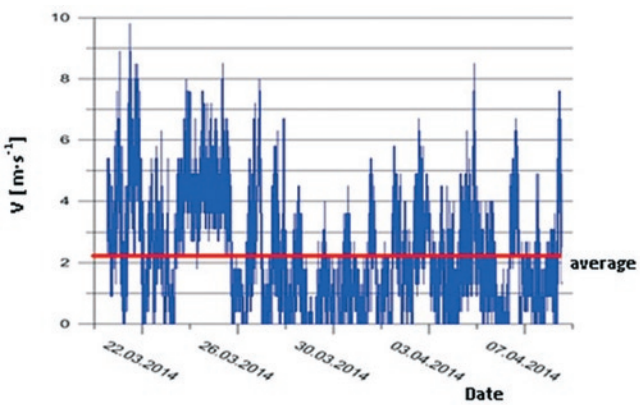


Fig. 20. Distribution of wind speed in the area where the wind turbine operated

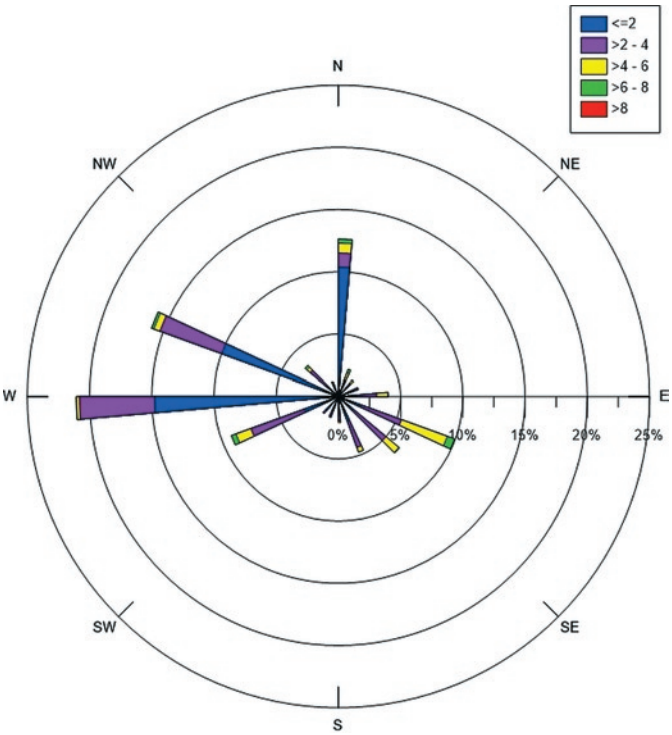


Fig. 21. Wind rose of the area of operation of the wind turbine

area. Over 80% of these winds was blowing at a speed of less than 4 $m \cdot s^{-1}$.

Electrical current delivered by the wind turbine very rarely exceeded the value of 1 A, and the average for the whole testing period was 50 mA, which of course affected the value of delivered power the average of which was 0.6 W. Distribution of power generated by the wind turbine is shown in Figure 22. It has been calculated that for an average wind speed of 2.21 $m \cdot s^{-1}$, the power of the wind working against the rotor of a wind turbine is approx. 6.6 W. In such poor wind conditions, the conversion efficiency of wind energy into electrical energy was only 9%.

Figures 23 and 24 show examples of the results of tests carried out in October 2014. When analysing the data from these graphs, one can observe the correlation between wind speed and power generated by the plant. For example, until 13.30 hours wind speed was increasing,

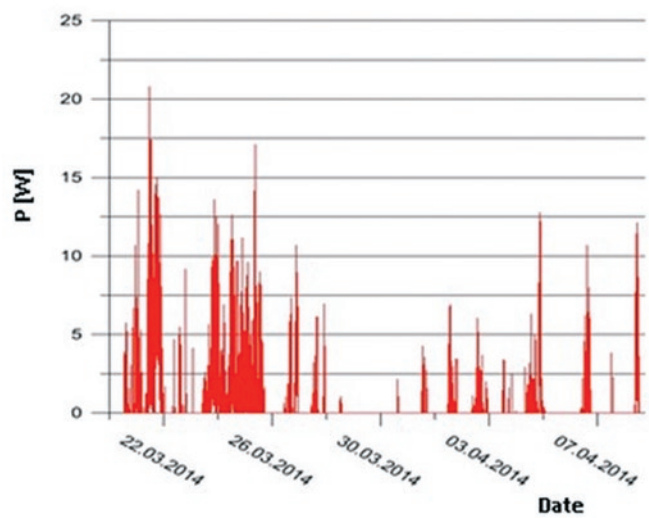


Fig. 22. Distribution of power generated by the dual-rotor wind turbine

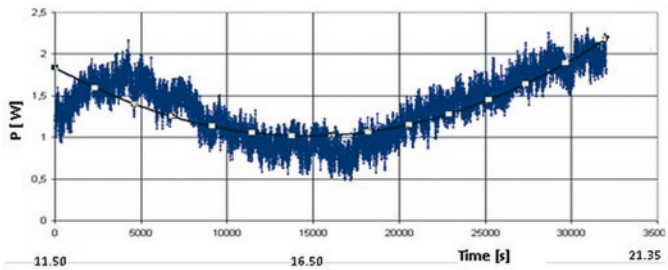


Fig. 23. Distribution of power delivered by the wind turbine on 29 October 2014

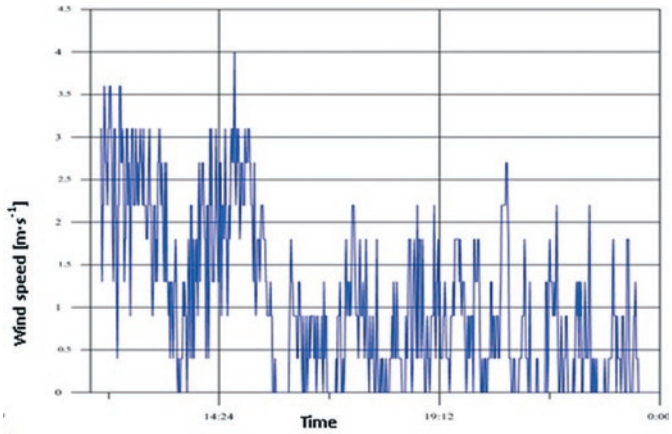


Fig. 24. Distribution of wind speed in the area where the wind turbine operated on 29 October 2014

which resulted in a increase in power generated, while after 14.30 with decreasing wind speed the lowest power generated on that day was recorded. In the period when speed was increasing, the average power of the wind was 4.3 W (the maximum absolute error of measurement 1,17W); compared to the registered electrical power generated, it would mean that the conversion efficiency of the turbine was 36%, which appears to be a fairly optimistic result.

4. Conclusions

1. The erected open circuit wind tunnel was characterized by the coefficient of uneven air stream distribution of 1.7% at the rated speed, while the index of turbulence intensity in the entire measurement range was between 1.2 and 1.8%.
2. The NACA-type rotor blades designed for the model of a mini wind turbine increased power generated by the tested plant by 33%.
3. Statistical analysis showed that the power of the dual-rotor mini wind turbine substantially depended (at a significance level of 0.05) on air stream speed and incidence angles of the rotor blades while it did not depend on the distance between the rotors. Effect of other factors was not determined.
4. During gusts of wind, power generated by the wind turbine increased at a 1-3 s delay relative to the recorded increase in wind speed.

*This research work was financed by the National Centre for Science as the research project N313 789940 entitled „**Analysis of the operation of a dual-rotor mini wind turbine**”.*

References

1. Arżanikow N S, Malcew W N. Aerodynamika. Warszawa: PWN, 1959.
2. Bak Ch, Fuglsang P, Johansen J, Antoniou I. Wind Tunnel Tests of the NACA 63-415 and a Modified NACA 63-415 Airfoil, Risø-R-1193(EN). 2000.
3. Capuzzi M, Pirrera A, Weaver PM. Structural design of a novel aeroelastically tailored wind turbine blade. Thin-Walled Structures 2015; 95: 7-15, http://dx.doi.org/10.1016/j.tws.2015.06.006.
4. Chong TP, Joseph PF, Davies POAL. Design and performance of an open jet wind tunnel for aero-acoustic measurement. Applied Acoustics 2009; 70: 605-614, http://dx.doi.org/10.1016/j.apacoust.2008.06.011.
5. Flaga A. Inżynieria wiatrowa. Warszawa: Wydawnictwo Arkady, 2008.
6. Goliński J A, Troskolański A T. Strumienice: teoria i konstrukcja. Warszawa: WNT, 1979.
7. Gumuła S, Knap T, Strzelczyk P, Szczerba Z. Energetyka wiatrowa. Kraków: Wyd. AGH, 2006.
8. Gumuła S, Mikoś M, Pytel K, Stępniewski Ł. Kształtowanie pola prędkości strumienia powietrza wypływającego z wentylatora osiowego. Mechanika/AGH 1999; 18(4): 499-507.
9. Gumuła S, Pytel K. Kształtowanie parametrów strugi powietrza za wentylatorem osiowym w kanale kwadratowym i w przestrzeni otwartej za kanałem. Mechanics 2005; 24(5): 246-251.
10. Hossein H, Liang C, Haitao Z, Vassilios K, Cem S, Tat-Hean G. A dual de-icing system for wind turbine blades combining high-power ultrasonic guided waves and low-frequency forced vibrations. Renewable Energy 2015; 83: 859-870, http://dx.doi.org/10.1016/j.renene.2015.05.025.
11. Idelchik I E. Handbook of Hydraulic Resistance. Jaico Publishing House 2005.
12. Jagodziński W. Silniki wiatrowe. Warszawa: PWT, 1959.
13. Karolewski B, Ligocki P. Rodzaje prądnic tarczowych. Wiad. Elektrotechniczne 2008: 8.
14. Karolewski B. Parametry modeli bezrdzeniowych prądnic tarczowych. Elektro.inf. 2011: 6.
15. Komarnicki P, Romański L, Bieniek J, Dębowski M, Pałka K. Wpływ regulacji pracy wentylatorów na równomierność rozkładu strumienia powietrza w tunelu aerodynamicznym. Inżynieria Rolnicza 2013; 2(143): 155-165.
16. Lee S, Son E, Lee S. Velocity interference in the rear rotor of a counter-rotating wind turbine. Renewable Energy 2013: 54: 235-240, http://dx.doi.org/10.1016/j.renene.2012.08.003.
17. Miąskowski W, Nalepa K, Pietkiewicz P, Komar W. Dwuwirnikowa mikrosiłownia wiatrowa z osią poziomą. Olsztyn: Wyd. UWM, 2012; 485-492.
18. Mitelet L, Oprina G, Chichaia R, Nicokaie S, Nedelcu A, Popescu M. Wind Tunnel Testing for a New Experimental Model of Counter-Rotating Wind Turbine. Procedia Engineering 2015; 100: 1141-1149, http://dx.doi.org/10.1016/j.proeng.2015.01.477.
19. Pagnini LC, Burlando M, Repetto MP. Experimental power curve of small-size wind turbines in turbulent urban environment. Applied Energy 2015; 154: 112–121, http://dx.doi.org/10.1016/j.apenergy.2015.04.117.
20. Selig M S, Guglielmo J J, Broeren A P, Giguere P. Summary of Low-Speed Airfoil Data. Virginia: SoarTech Publications, 1995; 1.
21. Seungmin L, Eunkuk S, Soogab L. Velocity interference in the rear rotor of a counter-rotating wind turbine. Renewable Energy 2013; 54: 235-240. http://dx.doi.org/10.1016/j.renene.2012.08.003
22. Seungmin L, Hogeon K, Soogab L. Analysis of aerodynamic characteristics on a counter-rotating wind turbine. Current Applied Physics 2010; 10: 1567-1739.
23. Somers D M. S822 and S823 Airfoils. October 1992--December 1993. NREL Report No. SR-500-36342, 2005; 37.
24. Sung Nam J, Tae-Soo N, Ki-Wahn R. Aerodynamic performance prediction of a 30kW counter-rotating wind turbine system. Renewable Energy 2005; 30: 631-644, http://dx.doi.org/10.1016/j.renene.2004.07.005.
25. Świtoński E, Jureczko M, Mężyk A. Optymalne projektowanie kompozytowych łopat elektrowni wiatrowej. Acta Mechanica et Automatica 2007; 1.1(1).
26. Wei X, Pan Z, Liping L. A novel folding blade of wind turbine rotor for effective power control. Energy Conversion and Management 2015; 101: 52-65, http://dx.doi.org/10.1016/j.enconman.2015.05.037.

Leszek ROMAŃSKI
Jerzy BIENIEK
Piotr KOMARNICKI
Institute of Agricultural Engineering
Wrocław University of Environmental and Life Sciences
ul. Chełmońskiego 37a, 61-530 Wrocław, Poland

E-mails: leszek.romanski@up.wroc.pl,
jerzy.bieniek@up.wroc.pl, piotr.komarnicki@up.wroc.pl

Marcin DĘBOWSKI
Faculty of Mechanical and Power Engineering
Wrocław University of Technology
ul. Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

Jerzy DETYNA
Department of Mechanics, Materials Science and Engineering
Wroclaw University of Technology
ul. Smoluchowskiego 25, 50-370 Wroclaw, Poland

Emails: marcin.debowski@pwr.edu.pl,
jerzy.detyna@pwr.edu.pl

Article citation info:
SHEN Q, QIU J, LIU G, LV K. Intermittent fault's parameter framework and stochastic petri net based formalization model. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 210–217, http://dx.doi.org/10.17531/ein.2016.2.8.

Qinmu SHEN
Jing QIU
Guanjun LIU
Kehong LV

INTERMITTENT FAULT'S PARAMETER FRAMEWORK AND STOCHASTIC PETRI NET BASED FORMALIZATION MODEL

MODEL PARAMETRYCZNY NIEZDATNOŚCI PRZEJŚCIOWEJ ORAZ MODEL FORMALNY OPARTY NA STOCHASTYCZNEJ SIECI PETRIEGO

The intermittent fault widely exists in many products and brings high safety risk and maintenance cost. At present there are some different opinions on the notion of intermittent fault and there is no comprehensive parameter framework for fully describing intermittent fault. Also the formalization model which can mathematically describe intermittent fault hasn't been constructed. In this paper, the conception of intermittent fault is discussed. A new definition of intermittent fault is put forward. Then the intermittent fault's parameter framework is presented. After that, the Stochastic Petri Net (SPN) based formalization model for intermittent fault is constructed. Finally an application of the SPN formalization model is shown. The parameters for intermittent fault are computed based on the proposed model and a case study is presented. The result shows the validity of the model. The model could assist the further research such as intermittent fault diagnosis and prognostic of remaining life.

***Keywords**: intermittent fault, parameter framework, stochastic Petri net, formalization model.*

Niezdatność przejściowa charakteryzuje wiele produktów i pociąga za sobą wysokie zagrożenie bezpieczeństwa oraz wysokie koszty eksploatacji. Obecnie istnieje wiele poglądów na temat pojęcia niezdatności przejściowej; nie stworzono jednak kompleksowego modelu parametrycznego pozwalającego w pełni opisać zjawisko niezdatności przejściowej. Nie skonstuowano także modelu formalnego, za pomocą którego można by opisać niezdatność przejściową w kategoriach matematycznych. W pracy omówiono koncepcję niezdatności przejściowej. Zaproponowano nową definicję tego pojęcia a następnie przedstawiono model parametryczny niezdatności przejściowej. Skonstruowano także model formalny niezdatności przejściowej oparty na stochastycznej sieci Petriego (SPN). Wreszcie, pokazano zastosowanie formalizacji SPN. Na podstawie zaproponowanego modelu obliczono parametry dla niezdatności przejściowej. Przedstawiono także studium przypadku. Otrzymane wyniki potwierdzają wiarygodność modelu. Opracowany model może być pomocny w dalszych badaniach dotyczących problemów, takich jak diagnozowanie niezdatności przejściowej czy prognozowanie pozostałego okresu użytkowania produktu.

***Słowa kluczowe**: niezdatność przejściowa, model parametryczny, stochastyczna sieć Petriego, model formalny.*

1. Introduction

Intermittent fault (IF) exists in many products, including from small elements to huge complicated equipment. The frequent occurrence of intermittent fault brings on serious troubles and results in high maintenance cost and safety risk. Early in the late 1960s, Hardie[1, 8] had indicated that IFs comprised over 30% of pre-delivery failures and almost 90% of field failures in computer systems. Roberts[17] figured out that up to 80 to 90% of system faults was arisen by IF in some situations. Banerjee[2] indicated that in wireless sensor networks IF was the most frequently occurring. Intermittent faults bring on many maintenance problems, such as No Fault Found (NFF), Can Not Duplicate (CND) and so on[20]. In 2012 a survey among 80 aerospace organizations ranked IF as the highest perceived cause of NFF[23]. NFF problem has been the highest cost source in aerospace maintenance. For example, the annual NFF exchange cost of the F-16 avionics boxes due to the IFs was over $20,000,000[21, 22].

Many issues of IF have been studied, including fault mechanism[4, 16], depicting parameters[5, 14], fault influence[7, 24], fault model, fault diagnosis[19, 22] and fault tolerance[3, 11, 12]. Sorensen[20] defined IF as any temporary deviation from nominal operating conditions of a circuit or device. Syed[23] defined IF as temporary malfunction of a device. Upon these definitions the environment induced disturbance may also be regarded as IF. Pan[14] regarded IF as a hardware error which occurs frequently and irregularly for a period of time. Upon the definition the IF could be a hardware fault, but in the early stage, IF may not occur frequently. Prasad[15]proposed three-state Markov model of IF, including normal, IF and permanent fault state. Sedighi[18] constructed an analytical state-space model of a robot-arm and diagnosed it by the residual between the measured output and estimated output of the model. Masson[13] modelled the interconnect system with the undirected graph and analysed its sufficient and necessary conditions for diagnosis. Singh[19] constructed Factorial Hidden Markov Model (FHMM) to diagnose multiple IFs.

By far there are some confusions on understanding IF. In addition, despite some researchers partially characterize the IF, there is no comprehensive framework of parameters to fully depict it. Finally the proposed IF models and diagnosis methods are merely applicable in particular situation. There is no formalization model which can generally and roundly express the IF problem.

In view of the above problems, the conception of IF is discussed in section 2. The systemic intermittent fault event and IF are distinguished. A more appropriate definition of IF is given. In section 3 a comprehensive parameter framework of IF is presented, which can

exhibit the temporal and probabilistic characteristic of IF. Further in section 4, the Stochastic Petri Net (SPN) based formalization model for IF is constructed. Also the model solution is given briefly. After that, in section 5 an application of the SPN formalization model is shown. The parameters proposed before are computed based on the SPN model. The conclusion is in Section 6.

2. Conception of intermittent fault

At present the difference between the phenomena of systemic intermittent fault and corresponding cause has not been distinguished. This brings on some confusion. For example, hash electromagnetism may induce an instantaneous pulse and then cause an error. Then it could be inferred as an IF. But in fact the hardware is fault free. In this section, the causes of systemic intermittent fault events are discussed, one of which is IF. Then a new definition of IF is put forward.

2.1. Cause of systemic intermittent fault event

When there is an observation that a product intermittently loses its given function, a judgment that IF is occurring may be made. But in fact the intuitionistic observation of IF is just a superficies which can be called systemic intermittent fault event. As shown in Fig. 1, the reasonable causes for systemic intermittent fault event include as below.

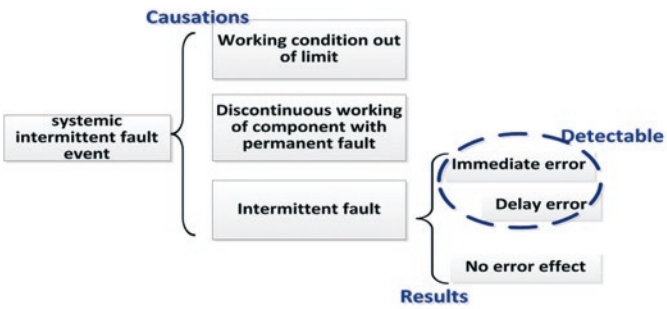


Fig. 1. Corelation between systemic intermittent fault event, IF and IF result

(1) Working condition of beyond limitation. The limitation here isn't the rated working environment, but the practical environment where components can work well. Due to the design defect and process variation, the appropriate working condition may not be consistent with the designed. The reasons of working condition going beyond limitation include variations of exterior environment and interior variations induced by components' working.

(2) Discontinuous activity of one component with permanent fault. As Kleer[6] indicated, there is a kind of IF which can disappear if it is modelled in a more detailed level. For example, when two wires are short-circuited, it looks like that the upper level gate has an intermittent fault. But in fact there is an un-modelled and unwanted connection. Practically in a piece of main equipment, the elements may not be working at the same time. If an element has a permanent fault, the fault appears only when it works. On the contrary, the fault would be temporarily masked if it doesn't work. Thereby the upper level function of this element manifests an intermittent off work.

(3) Intermittent fault. The occurrence of intermittent fault results from the essential physical degradation in products. This causes an intermittent interruption to normal function which will repeatedly manifest in a same characteristic.

2.2. Definition of intermittent fault

As shown in Fig. 1, intermittent fault will result in three cases, i.e. no error effect, immediate error and delay error. If there is no error when the IF occurs, it is no error effect. If the error occurs as soon as the IF occurs, it is immediate error. If the error comes into being after a certain time of IF occurring, then it is delay error.

There are two situations when the IF result is no error effect. When the IF duration is temporary or its induced abnormal signal is slight, then it will not disrupt system's normal performance. In addition, if the product itself has recovery mechanism, then the IF induced error is masked. For example, the network communication protocol supports error detection and retransmission. When the network connector has a slight poor contact, some data packages would be lost and then re-transmitted. Thus the communication function is still accomplished.

Thereby it can be found that IF has three typical characters. First, its occurring moment and duration are stochastic. IF may be in active state when it occurs or inactive state when it is temporarily suspended. Also it can recover without intervention. Second, IF occurs due to the physical injury. Third, when an intermittent fault occurs, an error may be induced or not. Generally only the IF which can induce error is paid attention to.

In summary, IF could be defined as fault that occurs irregularly and repeatedly for a certain time. The definition is formally consistent with the definition of fault[10]. Upon this definition, the IF is a real fault induced by a physical injury. It indicates the temporal intermittent character of IF. So the IF differs from permanent fault. Also the repetitive character distinguishes the IF from transient fault. It may only last for a period of time, as it can be recovered without intervention.

3. Intermittent fault's parameter framework

The IF occurs randomly, and its duration is not deterministic. As Guilhemsang[7] indicated, IF can randomly occur for a few times or more and continue from a few nano-seconds to seconds. Wells[25] indicated IFs can hold on from a few cycles to seconds or more, even as long as some days. When the IF occurs, it can be called as an activity. And when the IF temporarily disappears, it can be called as inactivity.

Considering the temporal exhibition and stochastic characteristic of IF, a parameter framework with respect to time and statistical domain is required to fully depict IF.

3.1. Temporal parameters of intermittent fault

As below, there are eight parameters in time domain for IF.

(1) IF activity time T_A. It denotes the duration of an IF activity.

(2) IF inactivity time T_I. It denotes the interval time between two activities.

(3) IF activity number N. It denotes the number of activities in specified time length T_s.

(4) IF activity frequency f_N. It denotes the IF activity number in unit time. It can be calculated as the rate of IF activity number N and time length T_s

(5) IF lasting time T_L. It denotes the total time in a burst of continuous activities. It can be calculated as Eq. (1):

$$T_L = \sum_{i=1}^{N} T_A^i + \sum_{i=1}^{N-1} T_I^i \tag{1}$$

Where T_A^i is the duration of ith activity, T_I^i is the interval time between i^{th} activity and its next.

(6) IF pseudo period T_P. It denotes the average time of a cycle of IF activities. Correcher[5] computes it as the rate in time window of all activities' time and activity number:

$$T_P = \frac{\sum_{i=j}^{i=k} t_{i+1} - t_i}{k - j + 1} \tag{2}$$

Where j and k respectively denotes the index of first and last activity in time window. t_i is the moment when the ith activity occurs.

(7) Error delay time T_d. It denotes the time IF has been lasting for when the error occurs.

(8) Error duration time T_c. It denotes the lasting time of an error. It should be noted that even if the IF had turn into inactivity, the error can still continue for a certain time.

3.2. Probabilistic parameters of intermittent fault

Let $\{IF_n \mid n = 1,2,\cdots,M\}$ denotes IF mode set, there are five probabilistic parameters of IF.

(1) IF existing probability p^I. It is the probability of existing IF_n and satisfies with $p_n^I + p_n^N = 1$, where p_n^N is the no fault probability.

(2) IF occurring condition probability p. It denotes the condition probability of given IF occurring when the product is faulty. That is:

$$\begin{aligned} & p_n = \Pr\{IF_n \mid faulty\} > 0 \\ & s.t. \quad n = 1,2,\cdots,M, 0 < \sum_{n=1}^{M} p_n = 1 \end{aligned} \tag{3}$$

It is computed as:

$$p_n = \frac{p_n^I}{\sum_{k=1}^{M} p_k^I} \tag{4}$$

(3) Fault activity probability p^A. It is the probability of IF in active state when IF is existing. That is:

$$\begin{aligned} & p_n^A = \Pr\{IF_n \text{ is active} \mid IF_n\} > 0 \\ & s.t. \quad n = 1,2,\cdots,M, 0 < \sum_{n=1}^{M} p_n^A \le M \end{aligned} \tag{5}$$

The temporal failure density (TFD) proposed by Correcher[5] denotes the IF average active time in a sliding time window with window width W. TFD is computed as below:

$$D = \frac{T_C + \sum_{i=j}^{N_W} T_A^i}{W} \tag{6}$$

Where N_W is the number of activities in the window. j is the index of first activity. T_C is the remaining time in window of the fault which occurs before the window.

In fact, there is the relation between p_i^A and D:

$$p^A = \lim_{W \to 0} D \tag{7}$$

The parameter p^A will increase with time. This corresponds to the fact that IF is due to the physical degradation. p^A will grow with the degradation process, so it can be used to prognosticate the remaining life and determine the optimum time for maintenance or exchange.

(4) Fault inactivity probability p^{IA}. It is the probability of IF in inactive state when it exists. That is:

$$p_n^{IA} = \Pr\{IF_n \text{ is inactive} \mid IF_n\} > 0 \tag{8}$$

It satisfies with $p_n^A + p_n^{IA} = 1$

(5) Causing error probability p^E. It is the probability of inducing an error when IF is active.

4. Stochastic Petri Net based formalization model for intermittent fault

The existing IF models don't cover all of the IF states and characters, or they can only be applied in particular situations. So a formalization model which is more general for different fields and more properly expresses the different states should be constructed. Petri Net (PN) can suitably depict complicate dynamic system[9]. Compared to Markov model, PN can model the transition with temporal character. In Stochastic Petri Net (SPN) the time between enable and firing of a transition is a stochastic variable which submits to a random distribution. So it is well consistent with the state transition of IF. That is the reason for adopting SPN to model IF.

In this section, the SPN of IF and corresponding Markov chain are drawn up. Then the SPN model is solved to obtain the transition probability matrix, the state probability distribution with time and the steady probability distribution.

4.1. Construction of SPN formalization model

Only single IF is considered. The SPN model is shown in Fig. 2. The inhibitor arc and condition arc are introduced into the model to extend its expression efficiency on state transition. The arc with a

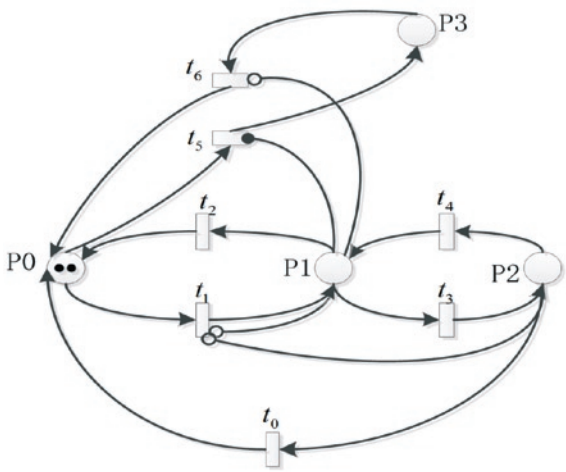


Fig. 2. SPN formaliation model for single intermittent fault

white circle end is the inhibitor arc. It is enabled when its input place has zero token. The arc with a black dot end is the condition arc. It is enabled when its input place has defined tokens. The physical meanings of different places and transitions are listed in Table 1. The transition firing rate denotes the average firing times in unit time when it is enable. According to the physical meaning, there will be $a_0 = a_2$.

As shown in Fig. 2, it can be seen that the SPN model can effectively exhibit the IF characters such as state transition, temporal randomness and fault influence.

The SPN model for IF can be expressed as a septuplet.

$$SPN =< P,T,F,E,W,M_0,R > \tag{9}$$

Where $P = \{P0,\cdots P3\}$ represents the set of places, $T = \{t_0,\cdots t_6\}$ represents the set of transitions, F represents the arcs, $E = \{E_0,\cdots E_6\}$ represents the enable function of transitions, W represents the weight of arcs. $M_0 = \{m_0,\cdots m_3\}$ represents the initial number of tokens in the places. $R = \{a_0,\cdots a_6\}$ represents the set of transition firing rates.

Assume that the transition firing rate is subjected to negative exponential distribution. The SPN model will be homogeneous with finite Markov chain. After analysing, the IF SPN model is converted into a Markov chain, as shown in Fig. 3. The physical meanings of different states and the corresponding tokens in different places is shown in Table 2.

From Fig. 3, we can obtain the transition rate matrix Q:

Table 2. Markov states' meanings and token numbers in different places

M	P0	P1	P2	P3	Meanings
M0	2	0	0	0	NO IF, NO Error
M1	1	1	0	0	IF active, NO Error
M2	1	0	1	0	IF inactive, NO Error
M3	0	1	0	1	IF active, Error
M4	0	0	1	1	IF inactive, Error
M5	1	0	0	1	NO IF, Error

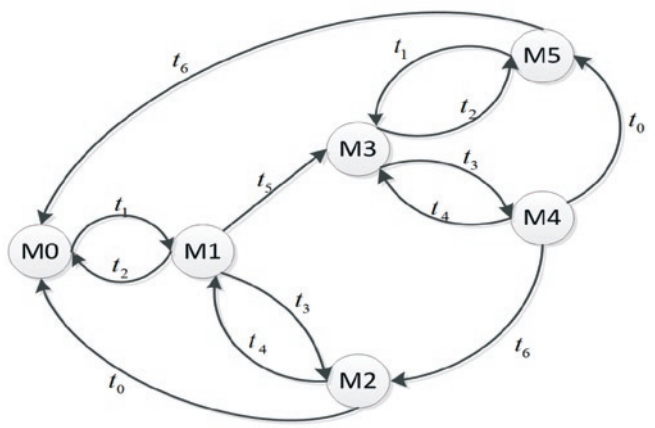


Fig. 3. Isomorphic Markov chain of SPN model

$$Q = \left[q_{ij}\right]_{6\times6} =$$

$$\begin{bmatrix} -a_1 & a_1 & 0 & 0 & 0 & 0 \\ a_2 & -(a_2+a_3+a_5) & a_3 & a_5 & 0 & 0 \\ a_0 & a_4 & -(a_0+a_4) & 0 & 0 & 0 \\ 0 & 0 & 0 & -(a_2+a_3)- & a_3 & a_2 \\ 0 & 0 & a_6 & a_4 & -(a_0+a_4+a_6) & a_0 \\ a_6 & 0 & 0 & a_1 & 0 & -(a_1+a_6) \end{bmatrix}$$

where $\{q_{ij} \mid i,j = 0,1,\cdots,5\}$ is the transition rate from state i to state j .

4.2. Probabilities solution of SPN formalization model

(1) Solution of transition probability

The transition probability matrix $P(t)$ is:

$$P(t) = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{05} \\ p_{10} & p_{11} & \cdots & p_{15} \\ \vdots & \vdots & \ddots & \vdots \\ p_{50} & p_{51} & \cdots & p_{55} \end{bmatrix}$$

where $\{p_{ij} \mid i,j = 0,\cdots,5\}$ is the transition probability from state i to state j . Based on the Kolmogorov forward equation, the time derivative of $P(t)$ is:

$$\dot{P}(t) = P(t)\times Q \tag{10}$$

Eq. (10) is solved to obtain:

$$P(t) = e^{Qt} = \sum_{k=0}^{\infty} \frac{(Qt)^k}{k!} \tag{11}$$

(2) Solution of state probability distribution

The probability of residing in state i at time t is $p_i(t)$, and then the state probability distribution at time t is:

$$F(t) = [p_0(t) \quad p_1(t) \quad \cdots \quad p_5(t)] \tag{12}$$

According Fokker-Planck equation, the time derivative of $F(t)$ is:

$$\dot{F}(t) = F(t)Q \tag{13}$$

Assume that the initial distribution is $F(0) = \begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0\end{bmatrix}$, and then it can obtain via the Laplace transform as bellow:

$$F(s) = \begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0\end{bmatrix}\times\left[s\mathbf{I} - Q\right]^{-1} \tag{14}$$

So the state probability distribution at time t can be obtain via the inverse Laplace transform of Eq. (14).It satisfies with:

$$F(t) = \begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0\end{bmatrix}P(t) \tag{15}$$

(3) Solution of steady state probability distribution

$\{P_i \mid i = 0,1\cdots 5\}$ is the steady probability of residing in state i. The steady state probability distribution is:

$$P_s = [P_0 \quad P_1 \quad \cdots \quad P_5] \tag{16}$$

Then Fokker-Planck equation is:

$$P_s Q = 0 \tag{17}$$

Solve Eq. (17) to obtain P_s. It is a linear equation set.

5. Application of the SPN formalization model

The SPN formalization model can mathematically express IF well. In this section the IF parameters defined previously are calculated based on the model.

5.1. Computation of intermittent fault parameters

As can be seen, the fault activity number N depends on the evaluating time length. Activity frequency f_N depends on the activity number. IF lasting time T_L depends on the variation of environment such as vibration and temperature. IF occurring condition probability p_i is only considered when there are multiple IF modes. Fault inactivity probability p^{IA} can be calculated by subtracting fault activity probability from 1. So these parameters are not analysed by the SPN model.

(1) The expectation time of residing in different states can be calculated as:

$$T_i^E = E[T_i] = -\frac{1}{q_{ii}} \quad i = 0,\cdots,5 \tag{18}$$

(2) The expectation of IF activity time is:

$$\begin{aligned} T_A^E &= E[T_A] = E[T_1] + E[T_3] \\ &= -\frac{1}{q_{11}} - \frac{1}{q_{33}} = \frac{1}{a_2 + a_3 + a_5} + \frac{1}{a_2 + a_3} \end{aligned} \tag{19}$$

(3) The expectation of IF inactivity time is:

$$\begin{aligned} T_I^E &= E[T_I] = E[T_2] + E[T_4] \\ &= \frac{1}{a_0 + a_4} + \frac{1}{a_0 + a_4 + a_6} \end{aligned} \tag{20}$$

(4) The expectation of IF pseudo period is:

$$T_P^E = T_A^E + T_I^E \tag{21}$$

(5) The expectation of error delay time is:

$$T_d^E = E[T_d] = \frac{1}{q_{13}} = \frac{1}{a_5} \tag{22}$$

(6) The expectation of error duration time:

The stepping route of error states is shown in Fig. 4. There are two cyclic return paths. The expectation of error duration time is the summation of time in all the states.

When it is in state M4, the expectation time of one-step transition is:

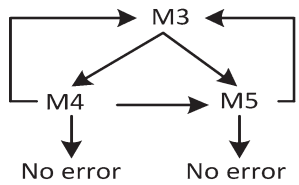


Fig. 4. Stepping route of error states in one cycle

$$T_c^{M4} = T_4^E + (p_{43}T_3^E + p_{45}T_5^E + p_{44}T_4^E) \tag{23}$$

When it is in state M5, the expectation time of one-step transition is:

$$T_c^{M5} = T_5^E + (p_{53}T_3^E + p_{55}T_5^E) \tag{24}$$

When it is in state M3, the expectation time of one-step transition is:

$$T_c^{M3} = T_3^E + (p_{34}T_4^E + p_{35}T_5^E + p_{33}T_3^E) \tag{25}$$

Because M3 is the initial state of the error cyclic route, so substitute T_4^E in Eq. (25) as T_c^{M4} and T_5^E as T_c^{M5}, the expectation time in one cycle is:

$$\begin{aligned} T_{c1} = T_3^E &+ (p_{34}p_{43} + p_{35}p_{53} + p_{33})T_3^E \\ &+ (p_{34} + p_{34}p_{44})T_4^E \\ &+ (p_{35} + p_{34}p_{45} + p_{35}p_{55})T_5^E \end{aligned} \tag{26}$$

So the total time of error states after n cycles will be:

$$\begin{aligned} T_{cn} = T_3^E &+ A(A+1)^{n-1}T_3^E \\ &+ B(A+1)^{n-1}T_4^E \\ &+ C(A+1)^{n-1}T_5^E \end{aligned} \tag{27}$$

Where $A = p_{34}p_{43} + p_{35}p_{53} + p_{33}$, $B = p_{34} + p_{34}p_{44}$, $C = p_{35} + p_{34}p_{45} + p_{35}p_{55}$.

As $0 < A << 1$, $(A+1)^{n-1}$ can be neglected when n is not large, so the expectation of error duration time can be calculated as:

$$T_c = T_3^E + AT_3^E + BT_4^E + CT_5^E \tag{28}$$

(7) IF existing probability is calculated as:

$$p^I = \frac{P_1 + P_2 + P_3 + P_4}{\sum_{s=0}^{5} P_s} \tag{29}$$

(8) Fault activity probability is calculated as:

$$p^A = \frac{P_1 + P_3}{P_1 + P_2 + P_3 + P_4} \tag{30}$$

(9) Causing error probability is calculated as:

$$p^E = p_{13} \tag{31}$$

5.2. Case study

Take a connector's intermittent contact fault as an example. The parameters are calculated based on the SPN model. Fault occurring rate is 1e-7, other rates is shown in Group 1 of Table 3. As shown in Fig. 5(a) and (b), resident time and steady probability of state M0 is very large, while in other states they are almost negligible. The existing probability of IF is just 9.9999e−6. It is because that the IF occurring rate is very small. As shown in Fig. 5(c) and (d), when IF occurs, fault activity time is greater than fault inactivity time, and fault activity probability is greater than fault inactivity probability. It is because that IF turning into inactivity rate is smaller than IF turning into activity rate. Since IF causing error rate is a little larger than sum of IF recovering rate and IF turning into activity rate, the IF causing error probability is 0.5 plus. The long error duration time is due to the low error recovering rate.

To compare with the generic situation, i.e. Group 1, set the rate of IF recovering , turning into inactivity and turning into activity an extreme value respectively, as shown in sets of Group 2-4 in Table 3. In the table the altered rates compared to Group 1 are marked with bold Italic font. It should be noted that the rate 100 is large enough to exhibit the margin result. The computation results are shown in Fig. 6. Compared to Group 1, in Group 2 the recovering rate is larger, so the IF existing probability is almost zero. IF activity time and inactivity time decreases, a Xiaomi YI nd it is almost always in active state when IF occurs. Compared to Group 1, in Group 3 the rate of IF turning into inactivity is so large that the fault is almost always inactive, so the inactivity time is greater than activity time and the fault activity probability decreases. Compared to Group 1, in Group 4 the rate of IF turning into activity is very large, thus the fault activity time is larger. Meanwhile both of the causing error probability and error duration time increase. In all the four groups, the causing error rate hasn't been altered, so the error delay times are the same.

To examine the influence of continuous variety of one rate, set the fault occurring rate comparative with others. It is considered that the rate of IF turning to inactivity is varying from zero to seven and other rates are shown in Group 5 of Table 3. The result is shown in Fig. 7 and Fig. 8. It can be observed that when the rate of fault turning into inactivity increases, the resident time of state M1 and M3 decrease (curve 2 and 4 in Fig. 7(a)) as others keep in constant. It is due to the decrease of activity time (curve 1 in Fig. 8 (a)). Consequently both of the fault causing error probability (curve 3 in Fig. 8(b)) and error duration time (curve 5 in Fig. 8(a)) decrease.

As shown in Fig. 7(b), with the increasing of rate a_3 , the probabilities in inactive state increase (curve 3 and 5), and the probability in

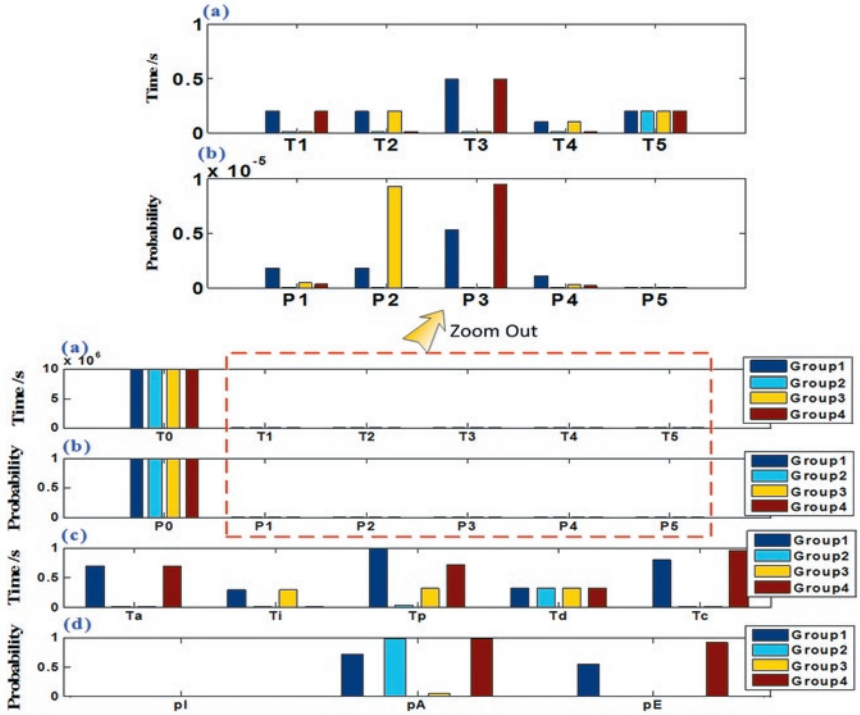


Fig. 6. Parameter solutions for different setting groups (a) resident time of states, (b) state steady probability, (c) different times, (d) different probabilities

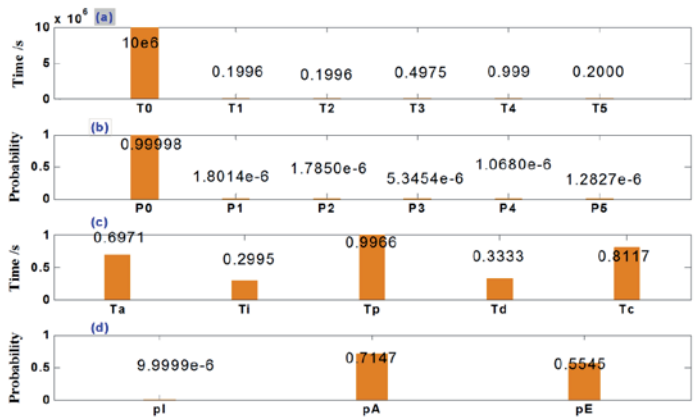


Fig. 5. Parameter solutions of connector intermittent contact fault (a) resident time of states, (b) state steady probability, (c) different times, (d) different probabilities

Table 3. Transition firing rate sets

Meanings	Rate	Group 1	Group 2	Group 3	Group 4	Group 5
IF recovering	a_0	0.01	***100***	0.01	0.01	***2***
IF occurring	a_1	1e-7	1e-7	1e-7	1e-7	***4***
IF recovering	a_2	0.01	***100***	0.01	0.01	***2***
IF turn into inactivity	a_3	2	2	***100***	2	***0~7***
IF turn into activity	a_4	5	5	5	***100***	5
IF causing error	a_5	3	3	3	3	3
Error eliminating	a_6	5	5	5	5	5

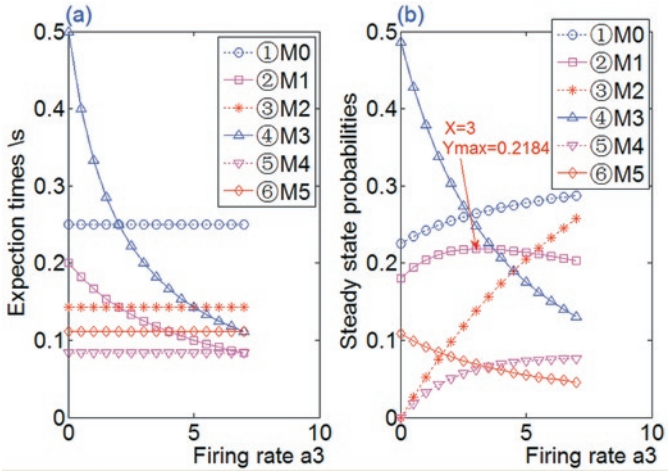


Fig. 7. Solutions of different states varying with IF turning into inactivity rate (a) expectation time, (b) steady state probability

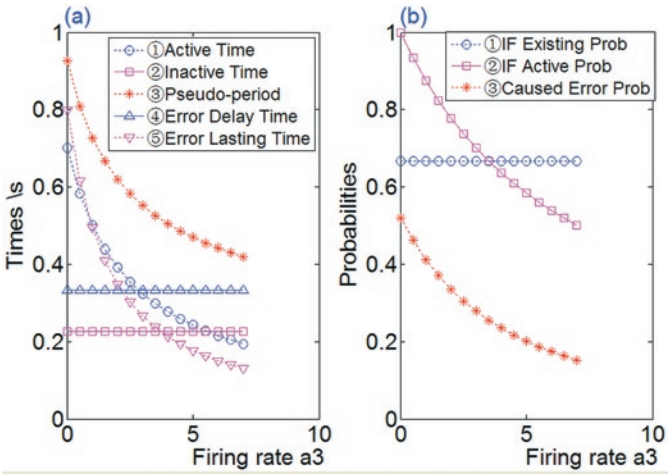


Fig. 8. Solutions of different states varying with IF turning into inactivity rate (a) times, (b) probabilities

error state decreases (curve 6). The reason of steady probability in state M0 increasing (curve 1) is the increasing opportunity of returning M0 from M2 via transition t_0 when the steady probability in M2 increases rapidly (curve 3). The steady probability in state M1 first increases, reaching maximum value when a_3 is three, and then decreases. As shown in Fig. 3, when a_3 is zero, state M1 will turn into state M3 or M0. State M0 can return to M1, but state M3 can't return to M1. With the increasing of a_3, the probability of residing in state M2 increases, and thus the opportunity of returning from state M2 to M1 increases. Consequently the steady probability of M1 increases. But when the rate a_3 is up to three, it is equivalent to the rate of transiting from state M1 to M3, and then the effect of increasing steady probability of M1 which comes from the return transition from M2 to M1 is no longer predominant, thus the continuous rising of a_3 will result in the increase of fault inactivity probability and decrease of activity probability.

The computation examples above show the availability of the SPN model. The parameters of IF can be analysed and calculated based on the SPN model. So the complicated IF problem is reduced to the study of transition firing rates.

6. Conclusions

In this paper the conception of IF is analysed, so as that the confusions on it are clarified. The systemic intermittent fault event and IF are distinguished, thus a more appropriate definition of IF is given out. Considering the statistical and temporal characters of IF, the parameter framework of IF is constructed, which includes the parameters in statistical domain and time domain. These parameters can more fully characterize the IF. And then the SPN formalization model for IF is proposed. This model can mathematically express the IF problem and reduce the complex problem into studying the seven parameters of transition firing rate. As an application of the SPN formalization model, the IF parameters can be calculated based on it. The computation method is given out and a case of calculating these parameters shows the availability of the SPN formalization model.

During the further research, the solution of transition firing rate should be studied, and then the IF diagnosis and prognostic of remaining life could be studied based on the SPN model.

Acknowledgement

This study is supported by National Natural Science Foundation of China (No. 61403408).

References

1. Ball M, Hardie F. Effects and detection of intermittent failures in digital systems. Proceedings of the fall joint computer conference, Montvale, N. J, 1969: 329-335, http://dx.doi.org/10.1145/1478559.1478597.
2. Banerjee N, P.M.Khilar. Performance analysis of distributed intermittent fault diagnosis in wireless sensor networks using clustering. 5th International Conference on Industrial and Information Systems, Mangalore, India, 2010: 13-18, http://dx.doi.org/10.1109/iciinfs.2010.5578744.
3. Constantinescu C. Impact of intermittent faults on nanocomputing devices. The 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, Edinburgh, UK, 2007: 238-241.
4. Constantinescu C. Intermittent faults and effects on reliability of integrated circuits. Annual Reliability and Maintainability Symposium, Las Vegas, USA, 2008: 370-374, http://dx.doi.org/10.1109/rams.2008.4925824.
5. Correcher A, Garcia E, Morant F, Blasco-Gimenez R, and et al. Diagnosis of intermittent fault dynamics. International Conference on Emerging Technologies and Factory Automation, Hamburg, 2008: 559-566, http://dx.doi.org/10.1109/etfa.2008.4638454.
6. De Kleer J. Diagnosing intermittent faults. 18th International Workshop on Principles of Diagnosis, Nashville, TN, UA, 2007: 45-51.
7. Guilhemsang J, Héron O, Giulieri A. Impact of the application activity on intermittent faults in embedded systems. 29th VLSI Test Symposium, Dana Point, CA, 2011: 191-196, http://dx.doi.org/10.1109/VTS.2011.5783782.
8. Hardie F H, Suhocki R J. Design and use of fault simulation for saturn computer design. IEEE Transactions on Electronic Computers 1967; EC-16(4): 412-429, http://dx.doi.org/10.1109/PGEC.1967.264644.
9. IEC 62551:2012. Analysis techniques for dependability - Petri net techniques.

10. IEC 60050-192:2014. International electrotechnical vocabulary – part 192: Dependability.
11. Kandasamy N, Hayes J P, Murray B T. Transparent recovery from intermittent faults in time-triggered distributed systems. IEEE Transactions on Computers 2003; 52(2): 113-125, http://dx.doi.org/10.1109/TC.2003.1176980.
12. Kothawade S, Chakraborty K, Roy S, Han Y. Analysis of intermittent timing fault vulnerability. Microelectronics Reliability 2012; 52(7): 1515-1522, http://dx.doi.org/10.1016/j.microrel.2012.03.003.
13. Mallela S, Masson G M. Diagnosable systems for intermittent faults. IEEE Transactions on Computers 1978; 100(6): 560-566, http://dx.doi.org/10.1109/TC.1978.1675147.
14. Pan S, Hu Y, Li X. Ivf:Characterizing the vulnerability of microprocessor structures to intermittent faults. IEEE Transactions on Very Large Scale Integration (VLSI) Systems 2012; 20(5): 777-789, http://dx.doi.org/10.1109/TVLSI.2011.2134115.
15. Prasad V B. Computer networks reliability evaluations and intermittent faults. IEEE Proceedings of the 33rd Midwest Symposium on Circuits and Systems, Calgary, Alta, 1990; 327-330.
16. Qi H, Ganesan S, Pecht M. No-fault-found and intermittent failures in electronic products. Microelectronics Reliability 2008; 48(5): 663-674, http://dx.doi.org/10.1016/j.microrel.2008.02.003.
17. Roberts M W. A fault-tolerant scheme that copes with intermittent and transient faults in sequential circuits. Proceedings of the 32nd Midwest Symposium on Circuits and Systems, Champaign, IL, 1989: 36-39.
18. Sedighi T, Phillips P, D Foote P. Model-based intermittent fault detection. Procedia CIRP 2013; 11: 68-73, http://dx.doi.org/10.1016/j.procir.2013.07.065.
19. Singh S, Kodali A, Pattipati K. A factorial hidden markov model (FHMM)-based reasoner for diagnosing multiple intermittent faults. International Conference on Automation Science and Engineering, Bangalore, India, 2009: 146-151, http://dx.doi.org/10.1109/coase.2009.5234134.
20. Sorensen B A, Kelly G, Sajecki A, Sorensen P W. An analyzer for detecting intermittent faults in electronic devices. Systems Readiness Technology Conference(AUTOTESTCON '94), Anaheim, CA, 1994: 417–421.
21. Steadman B, Sievert L S, Sorensen B, Berghout F. Attacking "bad actor" and "no fault found" electronic boxes. AUTOTESTCON, Orland, FL, UA, 2005: 821-824.
22. Steadman B, Berghout F, Olsen N. Intermittent fault detection and isolation system. AUTOTESTCON, Salt Lake City, 2008: 37-40, http://dx.doi.org/10.1109/autest.2008.4662580.
23. Syed W A, Khan S, Phillips P, Perinpanayagam S. Intermittent fault finding strategies. 2nd International Through-life Engineering Services Conference, Cranfield, Cranfield University, UK, 2013: 74-79, http://dx.doi.org/10.1016/j.procir.2013.07.062.
24. Wei J, Rashid L, Pattabiraman K, Gopalakrishnan S. Comparing the effects of intermittent and transient hardware faults on programs. 41st International Conference on Dependable Systems and Networks Workshops, Hong Kong, 2011: 53-58, http://dx.doi.org/10.1109/dsnw.2011.5958835.
25. Wells P M, Chakraborty K, Sohi G S. Adapting to intermittent faults in multicore systems. 13th International Conference on Architectural Support for Programming Languages and Operating Systems, Seattle, Washington, UA, 2008: 255-264.

Qinmu SHEN
Jing QIU
Guanjun LIU
Kehong LV

Science and Technology on Integrated Logistics Support Laboratory
National University of Defense Technology

College of Mechatronic Engineering and Automation
National University of Defense Technology
Changsha, Hunan, 410073, P. R. China

E-mails: qinmu2004@sina.com, qiujing16@sina.com,
gjliu342@qq.com, fhrlkh@163.com

Article citation info:
KOZIELSKI M, SIKORA M, WRÓBEL Ł. Decision support and maintenance system for natural hazards, processes and equipment monitoring. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 218–228, http://dx.doi.org/10.17531/ein.2016.2.9.

Michał KOZIELSKI
Marek SIKORA
Łukasz WRÓBEL

DECISION SUPPORT AND MAINTENANCE SYSTEM FOR NATURAL HAZARDS, PROCESSES AND EQUIPMENT MONITORING

SYSTEM WSPOMAGANIA DECYZJI DLA MONITOROWANIA ZAGROŻEŃ NATURALNYCH, PROCESÓW I URZĄDZEŃ

This paper presents the DISESOR integrated decision support system and its applications. The system integrates data from different monitoring and dispatching systems and contains such modules as data preparation and cleaning, analytical, prediction and expert system. Architecture of the system is presented in the paper and a special focus is put on the presentation of two issues: data integration and cleaning, and creation of prediction model. The work contains also two case studies presenting the examples of the system application.

Keywords: *decision support system, prediction, expert system, data cleaning, process monitoring, device monitoring, hazard.*

W pracy przedstawiono zintegrowany system wspomagania decyzji DISESOR oraz jego zastosowania. System pozwala na integrację danych pochodzących z różnych systemów monitorowania i systemów dyspozytorskich. Struktura systemu DISESOR składa się z modułów realizujących: przygotowanie i czyszczenie danych, analizę danych, zadania predykcyjne oraz zadania systemu ekspertowego. W pracy przedstawiono architekturę systemu DISESOR, a szczególny nacisk został położony na zagadnienia związane z integracją i czyszczeniem danych oraz tworzeniem modeli predykcyjnych. Działanie systemu przedstawione zostało na dwóch przykładach analizy dla danych rzeczywistych.

Słowa kluczowe: *system wspomagania decyzji, czyszczenie danych, predykcja, system ekspertowy, monitorowanie procesów, monitorowanie urządzeń, monitorowanie zagrożeń.*

1. Introduction

Coal mining is a heavy industry that plays an important role on an energy market and employs hundreds of thousands of people. Coal mining is also an industry, where large amount of data is produced but little is done to utilise them in further analysis. Besides, there is a justified need to integrate different aspects of coal mine operation in order to maintain continuity of mining what can be done by introduction of a decision support system (DSS).

Currently coal mines are well equipped with the monitoring, supervising and dispatching systems connected with machines, devices and transport facilities. Additionally, there are the systems for monitoring natural hazards (methane-, seismic- and fire hazards) operating in the coal mines. All these systems are provided by many different companies, what causes problems with quality, integration and proper interpretation of the collected data. Another issue is that the collected data are used chiefly for current (temporary) visualisation on boards which display certain places in the mine. Whereas, application of domain knowledge and the results of historical data analysis can improve the operator's and supervisor's work significantly.

For example, due to the short-term prognoses about methane concentration, linked with the information about the location and work intensity of the cutter loader, it is possible to prevent emergency energy shutdowns and maintain continuity of mining (the research on this methodology was discussed in [27]). This will enable to increase the production volume and to reduce the wear of electrical elements whose exploitation time depends on the number of switch-ons and switch-offs.

It is possible to see the rising awareness of monitoring systems suppliers who has started to understand the necessity to make the next step in these systems development. Therefore, the companies providing monitoring systems seek their competitive advantage in equipping their systems with knowledge engineering, modelling and data analysis methods. This is a strong motivation to consider a DSS presented in this paper.

The goal of this paper is to present an architecture of the DISESOR integrated decision support system. The system integrates data from different monitoring systems and contains an expert system module, that can utilise domain expert knowledge, and analytical module, that can be applied to diagnosis of the processes and devices and to prediction of natural hazards. Special focus of the paper is put on the data integration and data cleaning issues, such as outlier detection, realised by means of the data warehouse and the ETL process. The work also contains a more detailed presentation of the prediction module and two case studies showing real applications of the system.

The contribution of the paper consists of the architecture of the DISESOR integrated decision support system, its data repository and prediction module. Additionally, it covers the presentation of the issues connected with the preparation and cleaning of the data collected by monitoring systems, especially outlier detection. Finally, the contribution covers case studies presenting application of the described

system to abyssal mining pump stations diagnostics and methane concentration prediction in a coal mine.

The structure of the paper is as follows. Section 2 presents the works related to the presented topic. The architecture of the DISESOR system and its data repository are presented in section 3. The more detailed descriptions of the data preparation and cleaning and prediction modules are presented in sections 4 and 5 respectively. The case studies of abyssal mining pump stations diagnostics and methane concentration prediction task are presented in sections 6 and 7 respectively. Section 8 presents the final conclusions.

2. Related work

The typical environments deployed in a coal mine are monitoring and dispatching systems. These systems collect a large number of data which can be utilised in further analysis, e.g., on-line prediction of the sensor measurements, which area was surveyed in [11]. Such analysis can address different aspects of coal mine operation such as, e.g., equipment failure or natural hazards.

The examples of the research in the field of natural hazards in an underground coal mine cover, e.g., methane concentration prediction and seismic hazard analysis. The research on the prediction of the methane concentrations was presented in [26, 27, 28]. Application of data clustering techniques to seismic hazard assessment was presented in [15]. There are also approaches to prediction of seismic tremors by means of artificial neural networks [8] and rule-based systems [9]. Each research listed above is a standalone approach not incorporated into any integrated system.

Analytical methods that were mentioned require the data which are extracted, cleaned, transformed and integrated. Decision support systems utilise a data repository of some kind, e.g., a data warehouse [13]. The critical dependence of the decision support system on a data warehouse implementation and an impact of the data quality on decision support is discussed in [17].

There are applications of machine learning methods to diagnostics of mining equipment and machinery presented in literature [4, 5, 12, 19, 30]. The issue of mining industry devices diagnostics was raised among others in the works [7, 10, 18, 25, 31]. Besides, some initial concepts of the system that processes data streams delivered by the monitoring systems were presented in [6].

However, to the best of the authors knowledge there is no example of the integrated decision support system for monitoring processes, devices and hazards in a coal mine (except the work dealing with DSS for coal transportation [14], which loosely corresponds to the given topic).

3. System architecture

The general architecture of the DISESOR integrated decision support system is presented in Fig. 1. The architecture of the system consists of: Data repository, Data preparation and cleaning module, Prediction module (that are presented in more detail in the following sections), Analytical module and Expert system module (shortly presented below, as they are not the main focus of the paper).

3.1. Decision support system

The core of analytical, prediction and expert system modules is based on the RapidMiner [22] platform. The RapidMiner environment was customised to the requirements of the non-advanced user by disabling unnecessary options and views. Therefore, an advanced user can use the whole functionality of RapidMiner, whereas the non-advanced user can use such thematic operators as e.g., "Solve a methane concentration prediction issue" or "Solve a seismic hazard issue". Additionally, due to the target application of the system in Polish coal mines the RapidMiner environment was translated into Polish (for this reason several figures in this work contain Polish names). Finally, RapidMiner was extended in the created application by additional operators wrapping R [21] and MOA (Massive On-line Analysis) [1] environments.

The goal of the Data preparation and cleaning module, which is referred further as ETL2, is to integrate the data stored in data warehouse and process them to the form acceptable by the methods creating prediction and classification models. In other words the ETL2 module prepares the training sets.

Prediction module is aimed to perform incremental (on-line) learning of predictive models or apply classification and prediction models created in analytical module for a given time horizon and frequency of the values measured by the chosen sensors. This module also tracks the trends in the incoming measurements. The created predictive models are adapted to the analysed process on the basis of the incoming data stream and the models learnt on historical data (within the analytical module). The module provides the interfaces that enable the choice of quality indices and their thresholds that ensure the minimal prediction quality. If the quality of predictions meets the conditions set by a user, the predictions will be treated as the values provided by a soft sensor. They can be further utilised by e.g., expert system but also they can be presented to a dispatcher of a monitoring system.

Expert system module is aimed to perform on-line and off-line diagnosis of machines and other technical equipment. It is also aimed to supervise the processes and to support the dispatcher or expert decision-making with respect to both technical condition of the equipment and improper execution of the process. The inference process is performed by means of classical inference based on stringent rules and facts, fuzzy inference system or probabilistic inference based on belief networks. Additionally, the system contains a knowledge base editor that allows a user to define such rules and networks.

Analytical module is aimed to perform analysis of historical data (off-line) and to report the identified significant dependencies and trends. The results generated by this module are stored in the repository only when accepted by a user. Therefore, this module supports a user in decision-making of what is interesting from monitoring and

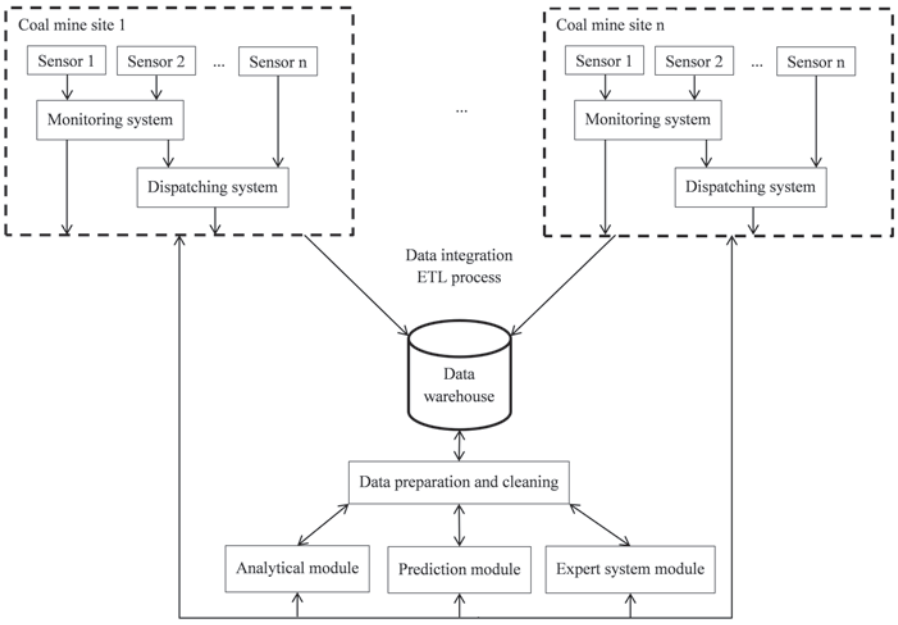


Fig. 1. Architecture of the DISESOR integrated decision support system

prediction point of view. Besides, it provides additional information that can be utilised to enrich the knowledge of expert system or that can be utilised to comparative analysis. The module supports identification of changes and trends in the monitored processes and tools and it also enables to compare the operator's and dispatcher's work.

3.2. Data repository

Data repository was designed as a data warehouse of a snowflake structure, that is presented in a general form in Fig. 2. The structure of a data warehouse results from the analysis of databases of the existing monitoring systems and the characteristics of the known sensors. The full list of tables with their description is presented in Table 1.

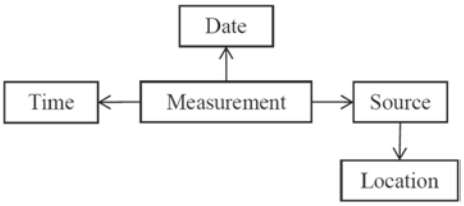


Fig. 2. Simplified schema of data repository

Table 1. Tables creating a data warehouse structure

Measurement	Value of a measurement
State	State of a measurement, e.g., *alarm*, *calibration*, *breakdown*
Discretisation	The measured values can be of discrete type
Time	Time of a measurement, range [00:00:00, 23:59:59], 1 second resolution
Time_category	Category, e.g., *mining* or *no mining*
Date	Date of a measurement
Location	Location of the measurement source
Location_attribute	Characteristics of the given location
Location_hierarchy	Hierarchical structure of location
Source	Measurement source, e.g., sensor or device
Source_attribute	Characteristics of the given source

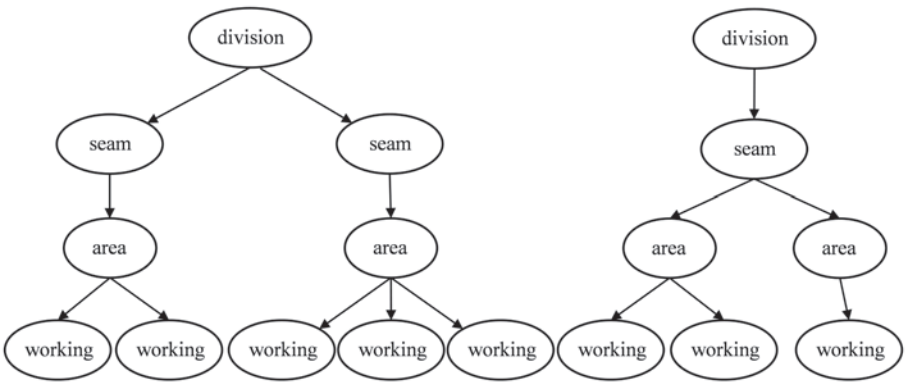


Fig. 3. Location hierarchy in a coal mine

The central table of the data repository is *Measurement* where all the measurements are stored. The dimensions related to the *Measurement* table are *Date*, *Time* and *Source*. *Date* and *Time* describe when the measurement was registered, whereas *Source* describes what registered the given measurement. The *Source* table contains among others such information about sensors/devices as:

- name (e.g. MM256),
- description (e.g. methane meter number 256),
- type name (e.g. methane meter),
- measured quantity (e.g. methane concentration),
- measurement unit (e.g. %CH4),
- name of a system that collects the data (e.g. THOR),
- range of measurements.

The *Source* table is described by means of *Location* dimension, that describes where in a coal mine it is located. The location has hierarchical structure, some sample hierarchy is presented in Fig. 3. The top-most level of the hierarchy is formed by coal mine divisions. Divisions consist of seams, which are divided into mining areas. At the bottom of the hierarchy there are mining workings.

The data warehouse is loaded with data by means of the ETL process designed for the main monitoring and dispatching systems for coal mining, which are deployed in Poland, Ukraine and China, e.g., THOR dispatching system [24] or Hestia natural hazards assessment system [9]. The ETL process was designed by means of Open Talend Studio [29].

During the tests of the created solution the data warehouse was loaded with 800 million records what resulted in 200 GB of data. It enabled the performance tests and optimisation of both the logical data warehouse structure and database management system (PostgreSQL [20]). As a result the *Measurement* data table was partitioned according to the months of measurements and the indices for foreign keys in this table were created. On the DBMS side several configuration parameters were adjusted, e.g., shared_buffers, work_mem, checkpoint_segments, effective_cache_size.

4. Data preparation and cleaning

The goal of ETL2 (Data cleaning and preparation) module is to deliver integrated data (in a form of a uniform data set) coming from chosen sources (especially sensors) in a chosen time range. Therefore, in this section the issues of frequency adjustment, aggregation and missing values imputation are presented. The outlier detection issues are extended in the subsection 4.1.

Measurements can be collected with different frequencies. Additionally, some systems collect a new measurement only after significant (defined in a monitoring system) change of the measured value. Table 2 presents how the measurements of two methanometers can look like, when collected directly from the data warehouse. The ETL2 process uniforms the data to the form, where each recorded measurement represents the time period defined by a user, e.g., 1 second (Table 3).

Within the ETL2 module there are also executed procedures of data cleaning, that identify outlier values and impute the missing values. These tasks are realised both by means of the simple functions presented below and by means of operators available in RapidMiner environment.

Another operations performed by means of the methods included in the ETL2 module are data aggregation (e.g., 10 measurements are replaced with 1 measurement) and manually performed definition of derived variables (e.g., a new variable can be calculated as a sum of the values of two other variables). The general scheme of data processing within ETL2 module is presented in Fig. 4.

Table 2. Data collected directly from data warehouse (- means that the measurement value does not change, ? means a missing value)

MN234 [%CH_4]	MN345 [%CH_4]	T [s]
0.1	0.1	0
0.2	-	1
-	0.2	4
0.5	?	7
0.3	0.3	9

Table 3. Data prepared to the further transformation, cleaning, etc.

MN234 [%CH_4]	MN345 [%CH_4]	T [s]
0.1	0.1	0
0.2	0.1	1
0.2	0.1	2
0.2	0.1	3
0.2	0.2	4
0.2	0.2	5
0.2	0.2	6
0.5	?	7
0.5	?	8
0.3	0.3	9

All the phases of processing presented in Fig. 4 are performed as separate RapidMiner operators. As a result of the processing performed by means of the ETL2 module we receive a data set that can be either analysed (by means of analytical module), or utilised to prediction model creation (by means of prediction module), or utilised within diagnostic process (by means of expert system module).

In order to select the variables that should be analysed a user can utilise THOR dispatching system, where each sensor (and attributes) are presented on a map of the region of interest. An exemplary screen of the THOR system is presented in Fig. 5. The system that is being created enables in turn, data (time-series) visualisation in order to select the time periods, that are the most interesting from the analyst point of view. Fig. 6 presents the visualisation of time-series consisting of several thousands of records. The developed operator creating such visualisation utilises R environment [21].

Aggregation of the measurements replaces several values with a single one. The period of aggregation is chosen by a user, who sets a number of measurements that should be aggregated or a time unit defining the windows containing measurements to be aggregated. The following aggregation operators are available for each attribute: average, minimum, maximum, median, dominant, the number of occurrences. For each record being the result of the aggregation there is calculated a weight, that is inversely proportional to the number of missing values existing in the aggregated data. The weight calculation is also based on a weighted average for all the attributes. This approach enables us to reduce the number of missing values in data and introduce weights that can be utilised by the chosen methods (e.g. rule induction).

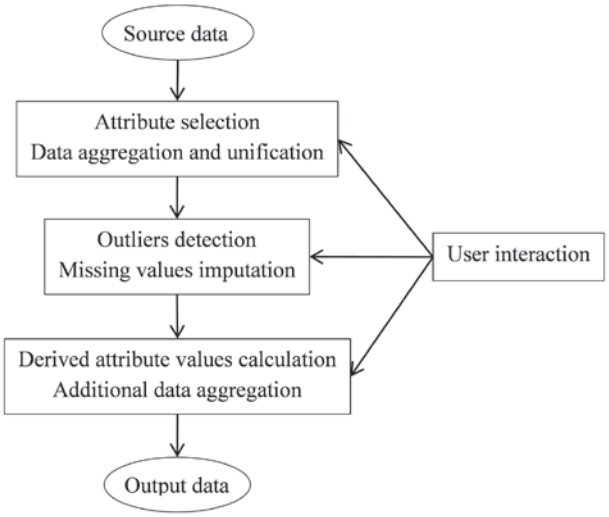


Fig. 4. General characteristics of the data processing in ETL2 module

The operator that imputes missing values performs the analysis of each attribute separately. The following methods that change the value or imputing the missing value can be utilised:

- a logical expression defining the replacing values (e.g. replace each value <1 with „low state”),
- the way how to receive the replacing values:
  - the value set by a user,
  - the last valid measurement,
  - average of the neighbouring measurements (with the parameter defining the number of neighbours),
  - linear regression of the two points (the last one before missing values section and the first one after this section),
  - linear regression of the data preceding missing values (with the parameter defining the window size).

The maximal number of consecutive missing values that can be imputed is defined as a separate parameter, as imputing the values for the long breaks in the measurements has no practical meaning. If the

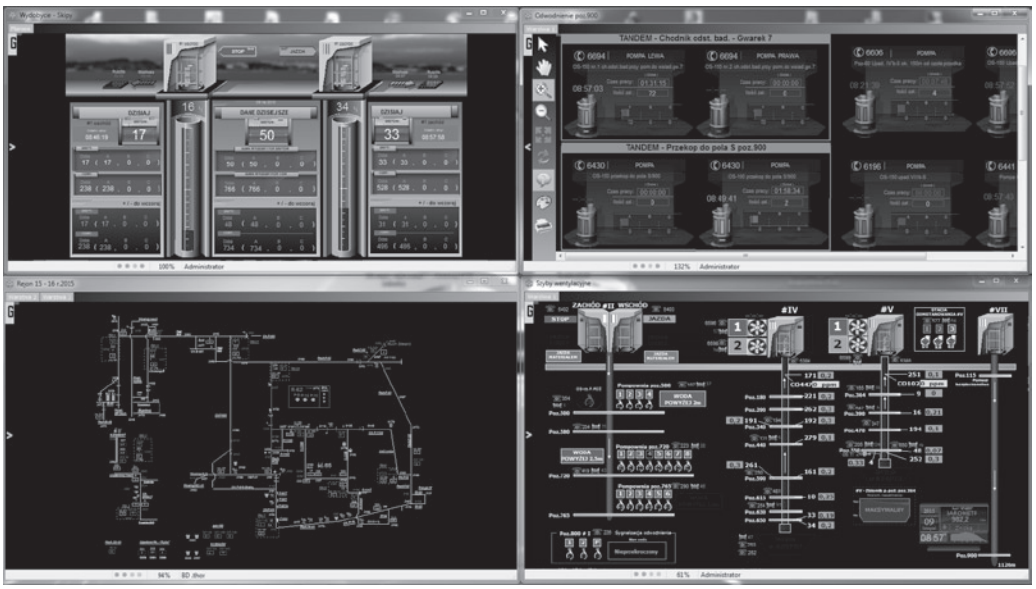


Fig. 5. Visualisation available in THOR dispatching system presenting a topology of the sensors

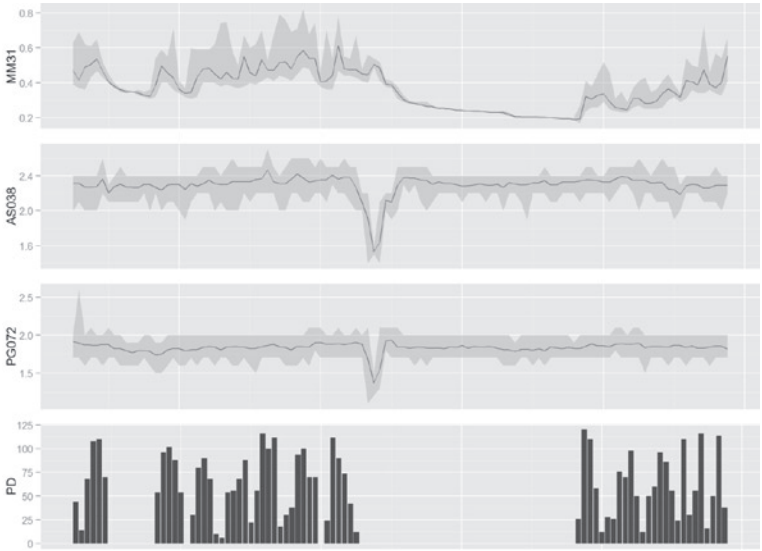


Fig. 6. Visualisation of exemplary time-series: methane concentration, air flow and mining cycle on a chosen longwall

resulting data set still contains missing values, the analyst can use a number of methods that are able to analyse data with missing values.

Introduction of a new derived variable can cover, among others, introduction of delays (the values of the previous measurements) or calculation of increments and trends (e.g. as an ordinal - increases, decreases). Another operator enables data smoothing by means of different filters (e.g. average, median). Finally, the last operator enables creation of dependent variable (decision variable). Typically, this variable contains the moved forward values of the chosen attribute, what enables to receive a proper prediction horizon. The operator defining the dependent variable has expanded functionality what enables e.g. to define the dependent variable as a maximal value of a given attribute in a defined time interval (e.g. 3 to 6 minutes in advance).

It is also important that within the developed framework the operators can be applied multiple times and in unrestricted order. Moreover, it is possible to pre-process data by means of the operators delivered by RapidMiner, that are dedicated to multidimensional analysis/identification of outliers and missing values (e.g. the operator applying local k-NN to missing values imputation).

When data preprocessing is finished, the whole process is saved according to XML-based RapidMiner standard, that was created for the needs of the system. Thereby, the prediction module and expert system module are able to transform the incoming data to the form that is acceptable by prediction and inference solutions. The incoming data in this case are collected on-line directly from the monitoring systems.

4.1. Outlier detection methods

Analysis of data coming from several underground coal mines showed that the missing values are relatively rare because most of the monitoring systems are the safety ones, where undisturbed data transfer is of the high importance. The methods that are based on linear interpolation or the last measured value approach fit well to the imputation of missing value task. Among 800 million data measurements that were loaded to DISESOR data repository only 0.5% contained missing values. These missing values consisted of single missing measurements, tens of missing measurements or longer periods of missing values being a result of transmission break (in this last case there is no effective method of missing value imputation).

The issue that is much more complex is detection of outlier values that can be a result of measurement interference. RapidMiner environment, except manual (expert) elimination of missing values that were mentioned above, offers several methods of automatic outlier detection. Such analysis is multidimensional, what means that impact of each variable of a given record is verified. Four methods of this type were evaluated during this research. These methods are characterised by high effectiveness in outlier detection and efficiency, as they do not require extensive computations. The methods that were chosen are the following [22]:

- Detect Outliers – Density (CDODe) – the method identifying the outliers on the basis of their density. The method requires two parameters. A record is identified as an outlier if there is at least the defined ratio of other records (where the ratio is given

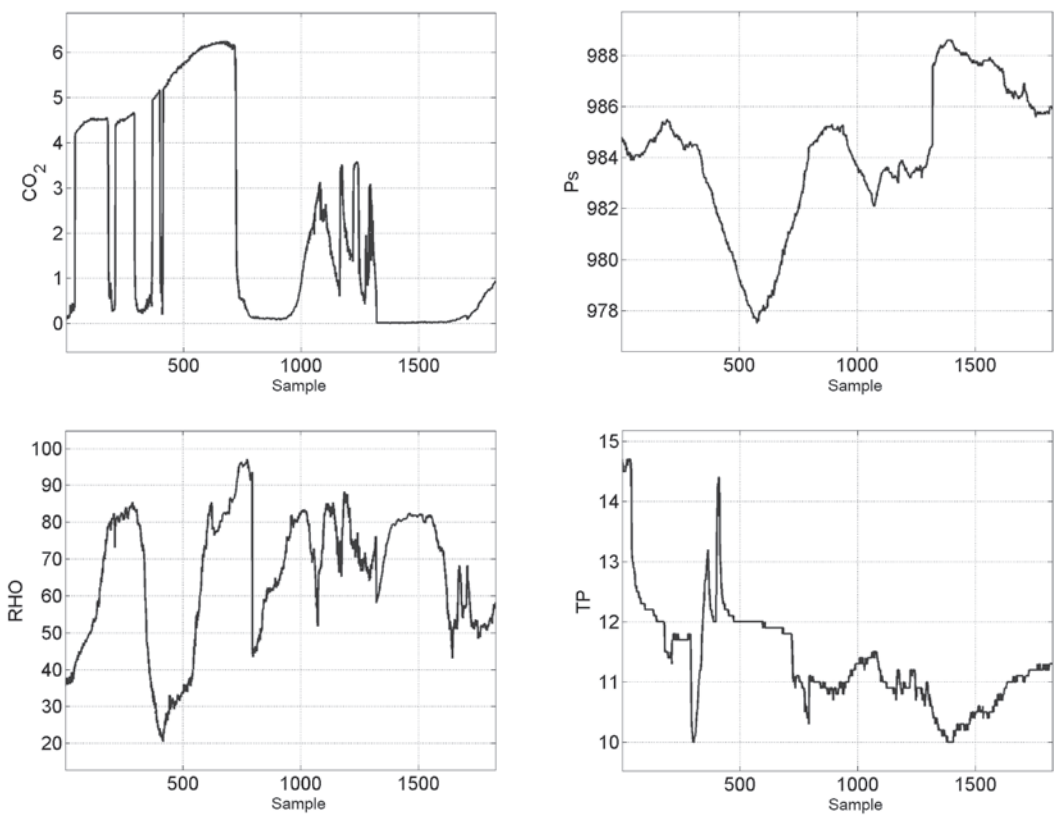


Fig. 6. Visualisation of exemplary time-series: methane concentration, air flow and mining cycle on a chosen longwall

as a parameter p) being more distant from this record then defined parameter d.

- k-NN Global Anomaly Score (GAS) – the method based on kNN approach. Each record is associated with its average distance to the rest of the records (by means of kNN method). Next, the record is identified as an outlier (or not) on the basis of interquartile range analysis.
- Local Density Cluster-Based Outlier Factor (LDCOF) – the method utilizing cluster analysis. A record is identified as an outlier on the basis of its distance to the centroid of the nearest large cluster. The distance is normalised by average distance to centroid among the members of this cluster. This method identifies small clusters as outliers.
- Histogram-based Outlier Score (HBOS) – the method based on a frequency histogram. The histogram can be created for the number of bins defined by a user or derived dynamically. The records belonging to the bin of the smaller size are labeled as outliers.

A more detailed description of the methods presented above can be found in RapidMiner documentation [22].

The methods listed above were applied to the analysis of time series of measurements registered on the operator platform in mine dewatering station [24]. Each record is characterised by the following variables (see Fig. 7): CO2 – CO2 concentration on the operator platform, Ps – atmospheric pressure, RHO – humidity on the operator platform, TP – temperature on the operator platform.

In order to verify the efficiency of outlier detection methods the outlier values in quantity 0.5%, 1%, 3% of original datasets were introduced to them. The outlier values were generated with use of noise with normal distribution.

The datasets were divided into training (2/3 of original time series – initial part) and test (1/3 of original time series – last part) datasets. The task was defined as a classification one, where two classes were defined – outlier values and correct values. Due to the imbalanced distribution of the examples from the two classes the results are presented as balanced accuracy reflecting average classification accuracy in each of the classes. The value 50 means that all the examples were classified to one class what makes the method useless.

During the first phase, where training data were analysed, the optimal parameters of the outlier detection algorithms were searched. The parameters were searched for each of the three experiments (0.5%, 1%, 3%). When the parameters were calculated, they were applied to test data analysis. The results of the analysis are presented in Table 4.

The second experiment was designed in such way that the outlier values were generated randomly from a given range encompassing the original measurements. Fig. 8 presents CO_2 time series containing 3% of outlier values.

The results of the analysis are presented in Table 5. It is clear that this task is much more difficult than the previous

Table 5. Evaluation of outlier detection methods – random values within a given range

Algorithm	Parameters	CO_2 data					
		0.5%	1%	3%	0.5%	1%	3%
		Training			Testing		
GAS	0.5%	-	81.51	71.25	75.93	81.39	71.25
	1%	78.49	-	83.58	78.21	78.30	78.31
	3%	63.77	78.81	-	76.04	74.79	72.80
HBOS	0.5%	-	85.40	84.93	80.49	80.96	77.45
	1%	85.23	-	82.44	84.18	84.65	81.41
	3%	81.40	85.95	-	81.40	81.13	74.87
LDCOF	0.5%	-	81.51	76.11	75.93	81.39	71.25
	1%	78.49	-	83.58	78.21	78.30	78.31
	3%	63.77	78.81	-	76.04	74.79	72.80
CDODe	**0.5%**	-	**86.19**	**86.39**	**83.71**	**83.55**	**80.58**
	1%	**77.69**	-	**81.14**	**83.16**	**80.52**	**75.61**
	3%	**81.46**	**81.39**	-	**81.48**	**81.53**	**80.34**

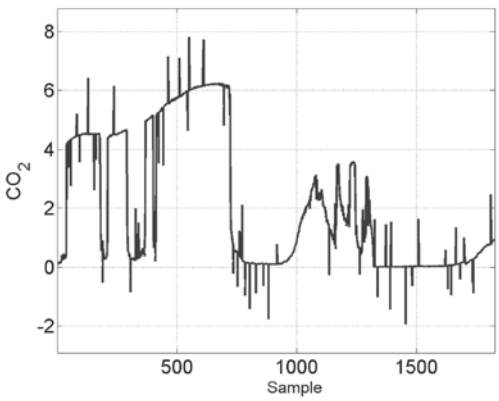


Fig. 8. CO_2 time series, with generated outlier values

Table 4. Evaluation of outlier detection methods – noise with normal distribution

Algorithm	Parameters	Balanced accuracy					
		0.5%	1%	3%	0.5%	1%	3%
		Training data			Test data		
GAS	0.5%	-	94.44	65.74	100	100	71.30
	1%	100	-	77.78	100	100	90.74
	3%	100	100	-	100	100	100
HBOS	0.5%	-	99.50	99.58	99.53	99.50	98.84
	1%	98.29	-	97.47	98.49	100	98.59
	3%	100	97.22	-	100	80.56	100
LDCOF	0.5%	-	100	100	100	100	100
	1%	100	-	100	99.94	99.94	100
	3%	95.07	95.24	-	95.05	94.99	96.56
CDODe	**0.5%**	-	**100**	**100**	**99.28**	**100**	**100**
	1%	**100**	-	**100**	**99.28**	**100**	**100**
	3%	**100**	**100**	-	**100**	**100**	**100**

one. It can be noticed that the CDODe method is the most stable approach and it gives the best results of outlier identification.

The CDODe algorithm is a default approach to outlier identification in multidimensional time series in the DISESOR system.

5. Prediction module

Prediction module is based on, so called, prediction services. Prediction service is a webservice that predicts values of a variable (discreet or continuous) on the basis of input vector. Prediction service is inseparably connected with a model (regression or classification one) that is the basis of the prediction.

The basic scenario of prediction service application is as follows:

1. Client sends a prediction execution request accompanied by a vector of conditional attributes and a timestamp.
2. Service calculates the prediction delivering a vector of conditional attributes as a model input. The attribute values come directly from a monitoring system, because the data warehouse is not loaded online. The values of the attributes are transformed according to the dedicated ETL2 process to the form acceptable by the prediction model.
3. Service loads the results to a database.

The architecture of the Prediction module is presented in Fig. 9.

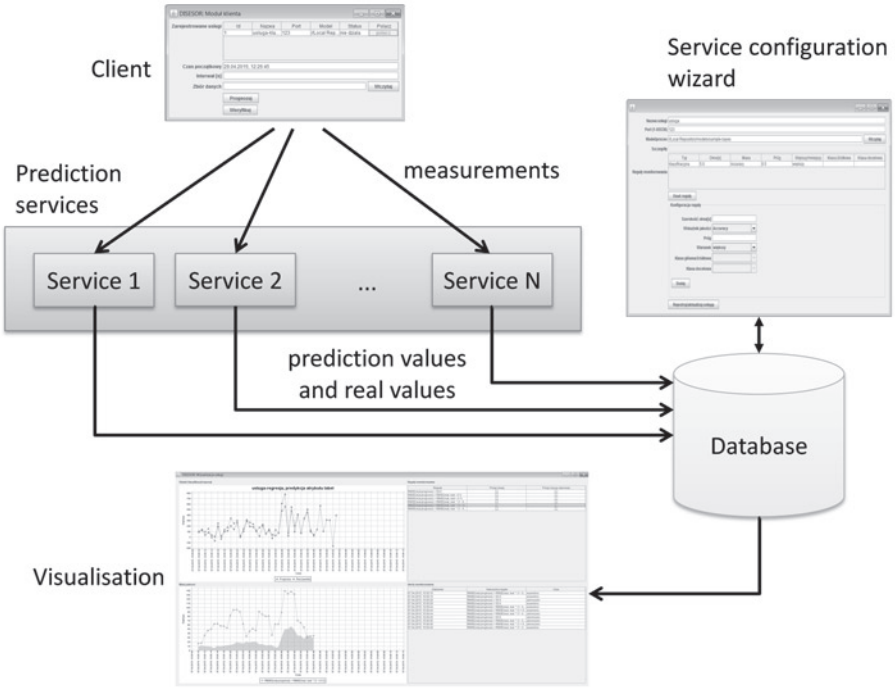


Fig. 9. Architecture and operation of prediction module

Database, which is an internal RapidMiner repository, stores the description of a model and the transformations of the attributes. Additionally, it stores the information about training data, the parameters of the minimal model quality and both predicted and real values of dependent variable. Each model adaptation results in a new database entry what makes the history of the changes available to the users.

Predictions can be visualised and compared on a single chart with the real values that are measured. Such visualisation can be performed by a monitoring or dispatching system (e.g. THOR dispatching system), where predicted values are delivered as measurements of a virtual sensor and the values of both sensors (virtual and real) can be easily compared.

It is assumed for the current module version, that if the quality of the predictions decreases below a given threshold, then a new training set is automatically collected. The size of this new data set is the same as size of the original data. The model adaptation is performed by modifying only the parameters of the existing model (the method and algorithm is not changed). Next, the quality of the model is verified on the same data that triggered the model adaptation (these data are not the part of the new training data set). If the quality of the adapted model is satisfactory, then this new model is applied to prediction. Otherwise, a message is generated stating that prediction cannot be continued and it is needed to come back to analytical module in order to create a new prediction model.

The configuration wizard enables to define the so-called quality monitoring rules. From the practical point of view there is no point in presenting the minimum model quality by means of the measures that are well-known by machine learning community, such as overall classification accuracy, g-mean, specificity, sensitivity, RMSE, MAE etc. Therefore, quality monitoring rules are based on: a sliding time-window (e.g. 1 hour) in which the quality is verified, frequency of the prediction calculation (e.g. 1 minute) and the indicators which are typically called *FalsePositive* and *FalseNegative*. The values of these indicators are explicitly defined by a user for each decision class or only for a target class, e.g. corresponding to "danger". Therefore, knowing the values of *FalsePositive* and *FalseNegative* [3], and a number of predictions that are calculated in a given time-window it is possible to calculate the values of almost all the possible quality measures of prediction model. In case of regression task the module allows so-called insensitivity, what means that the predictions that differ less than the given threshold from the real values are not treated as an error. Additionally, it is possible to define that the values within the given range (e.g. corresponding to the "normal" state) are not counted as errors.

6. Example of the system application to the task of abyssal mining pump stations diagnostics

Abyssal mining pump stations represent a fundamental solution to the problem of a coal mine dewatering. Due to the large responsibility in maintaining the water at a certain level, that guarantees the safe operation of the mine, the systems that oversee the abyssal mining pump stations are safety systems. The pump monitoring systems are installed in several dewatering stations and during the normal operation they register the following pump unit parameters:

- pump unit temperature,
- the power consumed by the motor,
- the current drawn by the motor,
- the productivity of a pump unit.

The values of the parameters listed above are acquired each second. Due to the safety constraints the temperature of the pump motor should not exceed 75 °C and a pump should be turned on when its temperature decreases below 25 °C.

Each underground water well contains four pumping units (see Fig. 10).

Analysis of the collected measurements enables the evaluation of the pump diagnostic states. The following feature vector was used during the analysis of pump diagnostic states:

$$P_i = [T_{U,i}, T_{0,i}\, t_{20\text{-}30,i}, t_{30\text{-}40,i}, t_{40\text{-}50,i}, t_{50\text{-}60,i}, t_{60\text{-}70,i}, P_{U,i}\, Q_{U,i}, L_i, D_{pi}, D_{ki}],$$

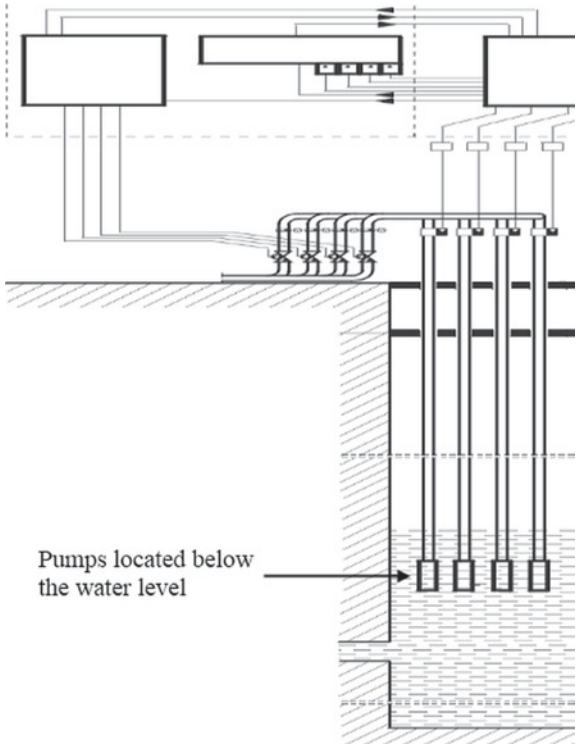


Fig. 10. Abyssal mining pump station

where:

- T_U – temperature of a pump unit in a steady state,
- T_0 – initial temperature of a pump unit,
- $t_{x\text{-}y}$ – time period when the pump temperature changes by 10 degrees (if the pump temperature has not reached a given range, a 0 value was inserted),
- P_U – power of a pump unit in a steady state,
- Q_U – performance of a pump unit in a steady state,
- L – number of starts on the previous day,
- D_p – time and date when the pump was turned on,
- D_k – time and date when the pump was turned off.

Temperature of a pump unit in a steady state (T_U) was calculated as an average value of the last two minutes of operation. Each record P_U reflects a single pumping cycle (starting from the unit turn on to turn off).

Analysis of historical data and interview with the dispatchers of the station (experts) enabled to define three diagnostic states: *a new pump unit* (also after repair), *correct operation* and *suitable for repair*.

The main impact on the diagnostic state of a pump unit have time periods $t_{x\text{-}y}$. Along the pump unit operation, when it becomes exploited, the time periods $t_{x\text{-}y}$ become shorter and the critical temperature when the pump must be turned off is reached faster. This results in pump numerous turn on during the days preceding decision of its repair. Therefore, the number of times the pump was turned on is an important diagnostic indication. It has to be regarded, however, in conjunction with the information about the temperature of a pump unit in a steady state in order to omit other than high temperature turn off reasons.

Pump state diagnostics was based on a Mamdani-type fuzzy system [16] with the following rules (the notation (p1, p2, p3, p4) reflects trapezoidal membership function):

IF T20_30 ∈ (199, 255, 255, 409) THEN a new pump unit
IF T20_30 ∈ (0, 197, 255, 409) and
T30_40 ∈ (245, 246, 256, 362) and
T60_70 ∈ (826, 1159, 1473, 679715) THEN a new pump unit

IF T20_30 = 0 and L∈(1, 1, 1, 2) THEN correct operation
IF T20_30 = 0 and
T40_50 ∈ (0, 0, 387, 727) and
L ∈ (1, 1, 3, 3) THEN correct operation
IF T20_30 ∈ (0, 255, 255, 409) and
T_u ∈ (73.1, 73.41, 74.54, 81.9) THEN correct operation

IF T20_30 = 0 and
T50_60 ∈ (0, 390, 390, 551) and
L ∈ (3, 5, 5, 7) THEN suitable for repair
IF T_0 ∈ (15.14, 19.75, 19.75, 27.26) and
T20_30 = 0 and
T30_40 ∈ (206, 366, 366, 11417) and
L ∈ (3, 5, 7, 7) THEN suitable for repair

Fig. 11 presents the division of attribute L (number of starts on the previous day) into fuzzy sets.

In practice, the state *suitable for repair* does not lead to immediate brake down of a pump unit. Dispatchers suggested that a pump classified to this state is able to (or sometimes has to – waiting for a service) operate up to next 3 months (when a typical operation time lasts 2 years). In order to improve the accuracy of service prediction (pump break down) a decision tree was created by means of a decision tree induction algorithm. Tree induction was performed only on the examples labeled as *suitable for repair*. The resulting tree classifies each vector P_i to one of two decision classes: *less than a month to break down* and *more than a month to break down*. The induced tree utilises only the time periods $t_{x\text{-}y}$. An applied train-and-test method showed the classification accuracy on a level of 90% (the class distribution was balanced, therefore, this measure is appropriate to clas-

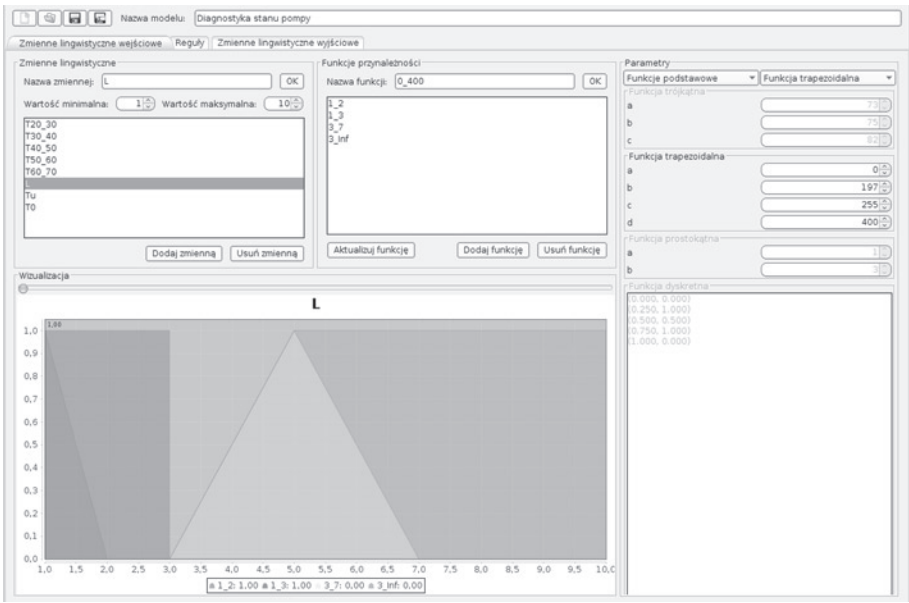


Fig. 11. presents the division of attribute L (number of starts on the previous day) into fuzzy sets.

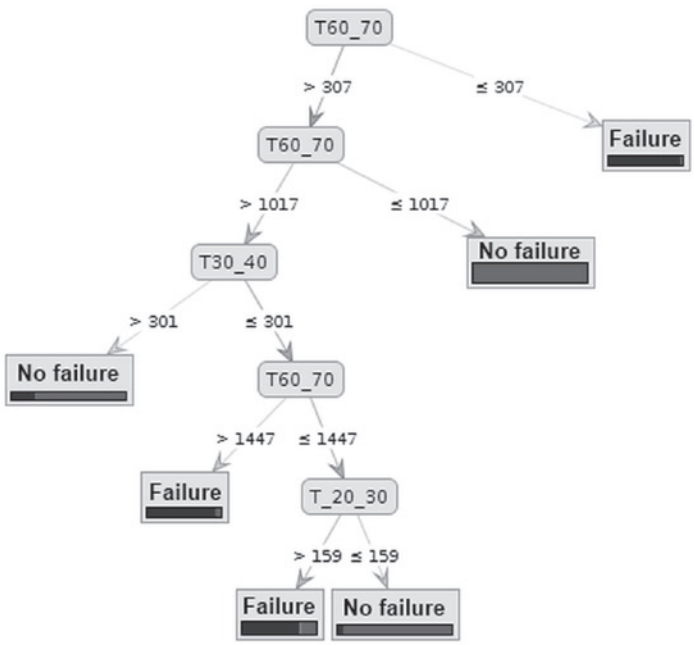


Fig. 12. The decision tree applied to the expert system

sification quality evaluation). The induced tree (the main node) was slightly modified, what increased accuracy by 2%, and it was applied to the expert system. The decision tree that was induced is presented in Fig. 12, where decision classes: *less than a month to break down* and *more than a month to break down* are represented as *Failure* and *No failure* respectively.

The expert system works according to the following steps: after each pumping operation the diagnostic state of a pump unit is evaluated. If a pump is classified as *suitable for repair*, then the decision tree is applied. It identifies the expected time period of the pump operation. If the pump is classified as *less than a month to break down* then the pump should be turned off because the costs of the repair of the broken unit are very high.

7. Example of the system application to the task of methane concentration prediction in mining excavation

The DISESOR system can be applied to solve a variety of tasks. This section presents an example, of the system application to methane concentration prediction.

Methane concentration monitoring is one of the main tasks of the natural hazard monitoring systems in mining industry. Such system is in charge of automatic and immediate shut-down of electricity within a given area, if a methane concentration exceeds a given alarm threshold. The power turn-on is possible after a certain time (from 15 minutes to even several hours), when the methane concentration decreases to the acceptable level. This results in large losses associated with downtime of production. Information from a soft (virtual) sensor presenting to a dispatcher the prediction of the methane concentration with a few minute horizon can prevent electricity shut-down or can allow to lower the mining activity and increase the air flow if possible. Therefore, these actions allow to avoid undesirable situations and unnecessary downtimes.

The task of maximal methane concentration prediction with the horizon from 3 to 6 minutes was realised within the DISESOR system. By means of ETL2 module a set of the following sensors was selected: AN321, AN541, AN547, AN682, BA1000, BA603, BA613, BA623, MM11, MM21, MM25, MM31, MM36, MM38, MM39, MM41, MM45, MM52, MM53, MM54, MM55, MM57, MM58, MM59, MM61, MM81.

The data were aggregated applying minimum operation to anemometer (AN) measurements, average operation to barometer (BA) measurements and maximum operation to methanometer (MM) measurements. The missing values were imputed applying linear regression method. As a dependent variable MM59 sensor was chosen. A map presenting the topology of the mining area and location of the sensors is presented in Fig. 13.

As analytical module is currently being developed, the method of regression tree induction was chosen arbitrarily to create the prediction model. The initial tree was created on the basis of data coming from 1 shift. The model and the list of sensors (variables) together with the defined transformations were forwarded to prediction model running a proper service. The time-window defined for prediction quality monitoring was set to 1 hour and the model adaptation was executed each hour regardless the minimum quality requirements. The adaptation could be executed more often if the minimum quality requirements were not met but there was no such situation. The data that were predicted were delivered on-line by the simulator of THOR system in order to simulate the real stream of measurements.

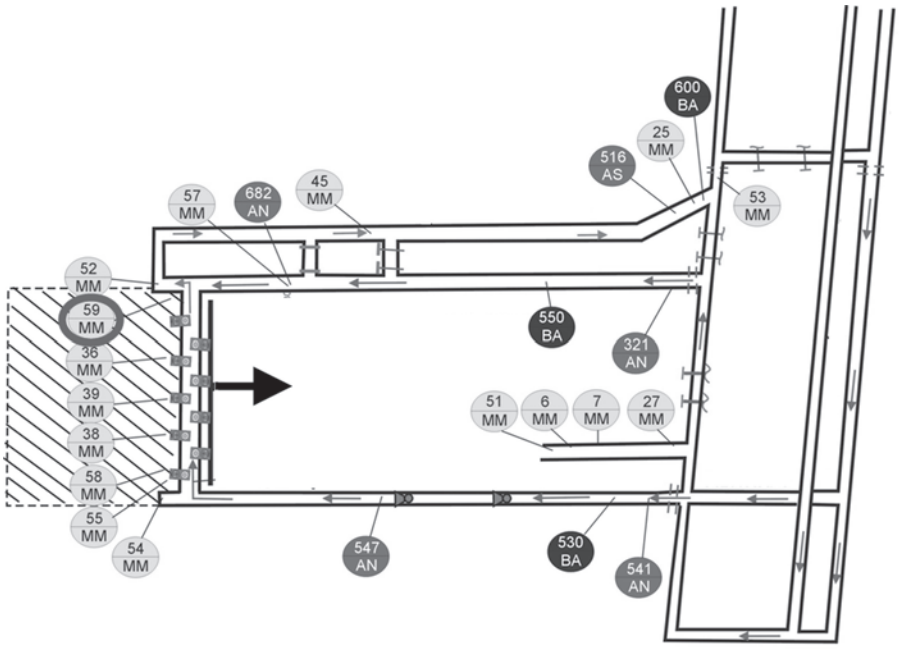


Fig. 13. Topology of the mining area and location of the sensors – MM59 sensor chosen as dependent variable is outlined a thick line

Fig. 14 presents the process of data preparation and the prediction model creation together with the initial regression tree that was created. Whereas, Fig. 15 presents the plot of the real methane concentration and the predicted maximum concentration together with the histogram of errors that are reported to a user. Currently, the user interface is in Polish as the system deployment in Poland was planned

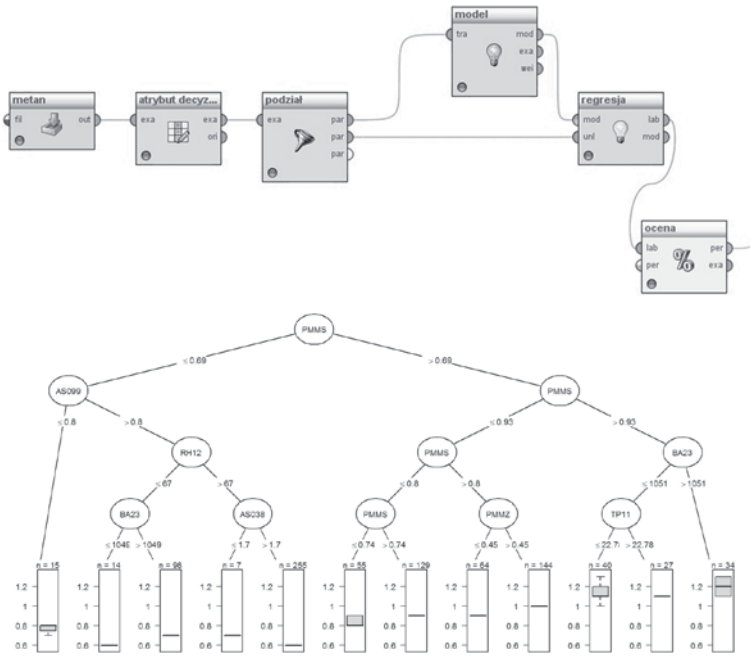


Fig. 14. The process of data preparation and prediction model creation together with the initial regression tree that was created

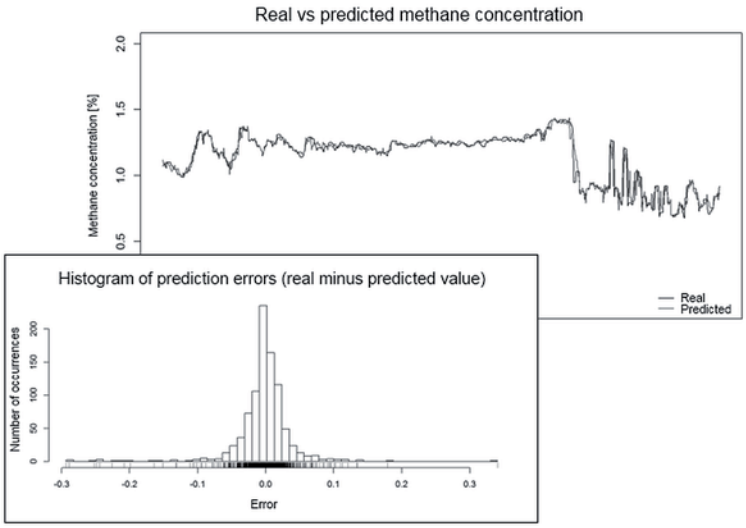


Fig. 15. The plot of the real methane concentration and the predicted maximum concentration together with the histogram of errors that are reported to a user

in the project. However, the English and Chinese versions are also planned.

Fig. 15. The plot of the real methane concentration and the predicted maximum concentration together with the histogram of errors that are reported to a user

8. Conclusion

The system that is being developed delivers the solutions for decision support of a dispatcher and process operator. This system is complete as it delivers the tools that can be applied to data storage, processing and preparation, and also to definition of the models based on expert knowledge (expert system) and the models based on the results of both historical and on-line data analysis. Due to the application and proper customisation of existing tools (RapidMiner, R) and development of the proprietary solutions (e.g. ETL2, rule induction and rough set operators [23] that are not available in RapidMiner) a user receives a broad set of tools that can be applied to different tasks. Finally, the case studies that were presented show that the system can be practically utilised in a coal mine industry.

The DISESOR system provides analytical tools available for advanced users as well as for users who are not data analysts (through many wizards that facilitate the use of the system). However, a routine use of the system requires, in our opinion, a new, gaining popularity, workplace, which is data scientist [2].

Acknowledgement

This research was supported by Polish National Centre for Research and Development (NCBiR) grant PBS2/B9/20/2013 in frame of Applied Research Programme. The work was also financially supported by POIG.02.03.01-24-099/13 grant: GeCONiI Upper Silesian Center for Computational Science and Engineering.

References

1. Bifet A, Holmes G, Kirkby R, Pfahringer B. Moa: Massive online analysis. The Journal of Machine Learning Research 2010; 11: pp. 1601-1604.
2. Dhar V. Data science and prediction. Communications of the ACM 2013; 56(12): 64-73, http://dx.doi.org/10.1145/2500499.
3. Fawcett T. An introduction to ROC analysis. Pattern recognition letters 2006; 27(8): 861-874, http://dx.doi.org/10.1016/j.patrec.2005.10.010.
4. Gąsior S. Diagnosis of longwall chain conveyor. Przegląd Górniczy 2001; 57(7-8): 33-36
5. Głowacz A. Diagnostics of Synchronous Motor Based on Analysis of Acoustic Signals with the use of Line Spectral Frequencies and K-nearest Neighbor Classifier. Archives of Acoustic 2015; 39(2): 189-194, http://dx.doi.org/10.2478/aoa-2014-0022.
6. Grzegorowski M. Scaling of complex calculations over big data-sets. Active Media Technology, Springer 2014; pp. 73-84.
7. Jurdziak L, Zimroz R. Dlaczego diagnostyka maszyn się opłaca i ile można na tym zaoszczędzić? Prace Naukowe Instytutu Górnictwa Politechniki Wrocławskiej 2004; 106: 139-150.

8. Kabiesz J. Effect of the form of data on the quality of mine tremors hazard forecasting using neural networks. Geotechnical & Geological Engineering 2006; 24(5): 1131-1147, http://dx.doi.org/10.1007/s10706-005-1136-8.
9. Kabiesz J, Sikora B, Sikora M, Wróbel Ł. Application of rule-based models for seismic hazard prediction in coal mines. Acta Montanistica Slovaca 2013; 18(4): 262-277.
10. Kacprzak M, Kulinowski P, Wędrychowicz D. Computerized information system used for management of mining belt conveyors operation. Eksploatacja i Niezawodność - Maintenance and Reliability 2011; 2(50): 81-93.
11. Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry. Computers & Chemical Engineering 2009; 33(4): 795-814, http://dx.doi.org/10.1016/j.compchemeng.2008.12.012.
12. Kalisch M, Przystalka P, Timofiejczuk A. Application of selected classification schemes for fault diagnosis of actuator systems. Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on. IEEE 2014; 1381-1390, http://dx.doi.org/10.15439/2014f158.
13. Kimball R, Ross M. The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons, 2011.
14. Kozan E, Liu S Q. A demand-responsive decision support system for coal transportation. Decision Support Systems 2012; 54(1): 665-680, http://dx.doi.org/10.1016/j.dss.2012.08.012.
15. Leśniak A, Isakow Z. Space-time clustering of seismic events and hazard assessment in the Zabrze-Bielszowice coal mine, Poland. International Journal of Rock Mechanics and Mining Sciences 2009; 46(5): 918-928, http://dx.doi.org/10.1016/j.ijrmms.2008.12.003.
16. amdani E H, Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies 1975; 7(1): 1-13, http://dx.doi.org/10.1016/S0020-7373(75)80002-2.
17. March S T, Hevner A R. Integrated decision support systems: A data warehousing perspective. Decision Support Systems 2007; 43(3): 1031-1043, http://dx.doi.org/10.1016/j.dss.2005.05.029.
18. Mazurkiewicz D. Computer-aided maintenance and reliability management systems for conveyor belts. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2014; 16(3): 377-382.
19. Michalak M, Sikora M, Sobczyk J. Analysis of the longwall conveyor chain based on a harmonic analysis. Eksploatacja i Niezawodność - Maintenance and Reliability 2013; 15(4): 332-333.
20. PostgreSQL. Postgresql 2015; Available online: http://www. postgresql.org/.
21. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria 2014; Available online: http://www.R-project.org
22. RapidMiner. RapidMiner 2015; Available online: http://rapidminer.com
23. Riza L S, Janusz A, Bergmeir C, Cornelis C, Herrera F, Ślęzak D, Benitez J M. Implementing algorithms of rough set theory and fuzzy rough set theory in the R package "RoughSets". Information Sciences 2014; 287: 68-89, http://dx.doi.org/10.1016/j.ins.2014.07.029.
24. Sevitel. Thor 2015; Available online: http://www.sevitel.pl/product,25,THOR.html
25. Sikora M, Michalak M. Eksploracja baz danych systemów monitorowania na przykładzie obserwacji pracy kombajnu chodnikowego. Bazy Danych: Rozwój metod i technologii, (Tom 1: Architektura, metody formalne i zaawansowana analiza danych), WKŁ, Warszawa 2008; 429-437.
26. Sikora M, Sikora B. Improving prediction models applied in systems monitoring natural hazards and machinery. International Journal of Applied Mathematics and Computer Science 2012; 22(2): 477-491, http://dx.doi.org/10.2478/v10006-012-0036-3.
27. Sikora M, Sikora B. Rough natural hazards monitoring. Rough Sets: selected Methods and Applications in Management and Engineering, Springer 2012; 163-179, http://dx.doi.org/10.1007/978-1-4471-2760-4_10.
28. Simiński K. Rough subspace neuro-fuzzy system. Fuzzy Sets and Systems 2015; 269: 30-46, http://dx.doi.org/10.1016/j.fss.2014.07.003.
29. Talend. Talend open studio 2015 Available online: https://www.talend.com/ products/talend-open-studio.
30. Wachla D, Moczulski W A. Identification of dynamic diagnostic models with the use of methodology of knowledge discovery in databases. Engineering Applications of Artificial Intelligence 2007; 20(5): 699-707, http://dx.doi.org/10.1016/j.engappai.2006.11.002.
31. Zimroz R. Metody adaptacyjne w diagnostyce układów napędowych maszyn górniczych. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 2010.

Michał KOZIELSKI
Institute of Electronics
Silesian University of Technology
Akademicka 16, 44-100 Gliwice, Poland

Marek SIKORA
Institute of Informatics
Silesian University of Technology
Akademicka 16, 44-100 Gliwice, Poland

Łukasz WRÓBEL
Institute of Innovative Technologies EMAG
Leopolda 31, 40-189 Katowice, Poland
Institute of Informatics, Silesian University of Technology
ul. Akademicka 16, 44-100 Gliwice, Poland

E-mails: michal.kozielski@polsl.pl, marek.sikora@polsl.pl, lukasz.wrobel@ibemag.pl

Article citation info:
LU J-M, LUNDTEIGEN MA, LIU Y, WU X-Y. Flexible truncation method for the reliability assessment of phased mission systems with repairable components. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 229–236, http://dx.doi.org/10.17531/ein.2016.2.10.

Ji-Min LU
Mary Ann LUNDTEIGEN
Yiliu LIU
Xiao-Yue WU

FLEXIBLE TRUNCATION METHOD FOR THE RELIABILITY ASSESSMENT OF PHASED MISSION SYSTEMS WITH REPAIRABLE COMPONENTS

ZASTOSOWANIE METODY ELASTYCZNEGO OBCIĘCIA DO OCENY NIEZAWODNOŚCI SYSTEMÓW O ZADANIACH OKRESOWYCH Z ELEMENTAMI NAPRAWIALNYMI

Phased-mission systems (PMS) are the system in which the component stresses and the system configuration may change over time. Real-world PMS usually consist of a large number of repetitive phases and repairable components. Existing approaches for the reliability analysis of this kind of PMS tend to suffer from the problem of state explosion or binary-decision-diagram (BDD) explosion. This paper presents a truncation method based on the BDD and Markov chains to solve the scaling issue. In our approach, the truncation mitigates the BDD explosion and broadens the applicability of the BDD & Markov method. Different from the classic truncations, our truncation limit is flexible, which ensures that ensure the truncation error is lower than the predefined threshold. The advantages of the proposed method are illustrated through two practical PMS which are challenging to classic non-simulation approaches.

***Keywords**: flexible truncation limit; phased mission systems; reliability evaluation; repairable components.*

Systemy o zadaniach okresowych (phased mission systems, PMS) to takie systemy, w których naprężenia elementów składowych oraz konfiguracja systemu mogą z czasem ulegać zmianie. W warunkach rzeczywistych, PMS zazwyczaj charakteryzują się dużą liczbą powtarzalnych faz zadaniowych i składają się z wielu naprawialnych elementów. Istniejące metody analizy niezawodności tego typu systemów niestety posiadają ograniczenia związane z problemem eksplozji stanów lub eksplozji diagramów binarnych decyzji (binary decision diagram, BDD) Praca przedstawia metodę obcinania opartą na BDD oraz łańcuchach Markowa, która pozwala rozwiązać wspomniane problemy złożoności obliczeniowej. W proponowanym podejściu, obcięcie minimalizuje eksplozję BDD zwiększając możliwości zastosowania metody opartej na BDD oraz łańcuchach Markowa. W odróżnieniu od klasycznego obcinania, w opracowanej przez nas metodzie granica obcięcia jest elastyczna co pozwala zredukować błąd obcięcia poniżej wcześniej określonego progu. Zalety proponowanej metody zilustrowano na przykładzie dwóch stosowanych w praktyce systemów PMS, które stanowią wyzwanie dla klasycznych metod niesymulacyjnych.

***Słowa kluczowe**: elastyczna granica obcięcia; systemy o zadaniach okresowych; ocena niezawodności; elementy naprawialne.*

1. Introduction

In many real-world applications such as airliner, railway, and satellite systems, operation of missions often involves multiple tasks or phases that must be accomplished in sequence. During each phase, some equipments may play different roles and subject to different environmental conditions. Thus, the system configuration, success criteria, and component behavior may change from phase to phase. This kind of systems are commonly termed as the phased-mission systems (PMS) . A typical example of PMS is the monitoring system in the satellite-launching mission which involves the launching, separation, and orbiting phases. Some equipments, such as the ground-station radars, may suffer from increasing stresses in the separation phase and becomes idle in other phases. Because of the dependence problem, the reliability of PMS is not the product of reliabilities of individual phases. Hence, the reliability analysis of PMS is more complex than that of the single-phase systems.

Over the past four decades, substantial progress has been made in the reliability analysis of PMS [3]. Basically, existing methodologies can be categorized into the simulation and the analytical methods. The main advantage of the simulation methods [6, 19, 25] is their wide applicability to a variety of scenarios, while the merit of the analytical approaches lies in the accuracy of algorithm results. The analytical approaches can be further classified into the state-based, the combinatorial, and the modular methods. Typical state-based methods, such as the Markovian models [1, 4, 28, 31], are commonly used to analyze the PMS with repairable components. A commonly known challenge facing the state-based methods is the state explosion problem when the number of components becomes large. On the other hand, the binary decision diagrams (BDD) based methods [2, 15, 16, 18, 22, 27, 29, 30] (belong to combinatorial methods) are efficient for the PMS with many components. However, BDD based methods can suffer a similar explosion problem (node or path explosion) when the number of phases becomes large. This problem is known as the BDD explosion in the PMS analysis.

The modular methods [13, 14, 21, 24, 25], which integrate the state-space and the combinatorial methods, are efficient in analyzing the PMS with a multitude of repairable components. A classic rep-

resentative of the modular methods is the BDD & Markov approach [25] which is designed for the PMS with many exponentially distributed components. As the further work of the BDD & Markov method, Shrestha et al. [24] proposed the multistate multivalued decision diagram for the reliability analysis of PMS with multi-state repairable components. Another further work [13] is the component-behavior model which analyzes the PMS with combinatorial phase requirements and repairable components. However, the BDD & Markov method (and its further works) cannot avoid the BDD explosion problem because the cross-phase BDD has to be assessed from the first phase to the last phase.

In order to overcome the scaling issue, extensive research efforts have been expended in the analysis of approximations, bounds, and truncations. A common practice [5, 11, 12, 17, 19] is to apply truncations to BDD or cut sets, that is, to remove the BDD nodes (or BDD paths) whose probabilities are less than the truncation limit. However, a lot of researchers [5, 7, 12] find that a fixed truncation limit may be the major source of estimation error. In this paper, we implement truncations to reduce the computational complexity of the BDD & Markov method for the analysis of large PMS. Furthermore, our truncation limit will decrease as the truncation proceeds. This truncation strategy can keep the overall error under the user control, and avoids the complex discussion about error estimation in the literature.

The remainder of the paper is organized as follows. Section 2 presents the proposed method which integrates BDD, Markov chains, and truncations. Section 3 illustrates our approach through two real-world PMS. The efficiency of our approach is compared with the conventional BDD & Markov method and the Petri-net simulation. Lastly, Section 4 gives conclusions and future directions of our work.

2. Proposed method for large PMS analysis

2.1. Assumptions & method overview

Throughout the paper, we make the following assumptions for our approach. (1) The life and the repair time of components are independent variables of exponential distributions. (2) The mission is assumed to fail if the system fails in any phase. (3) Repaired components can be reused only in the next phase. (This assumption exists in the BDD & Markov method as well.)

Generally, our approach consists of mainly three steps, that is,

Step 1 - Generate component-behavior vectors (*CBV*) based on BDD.

Step 2 - Compute path vectors (*PV*) phase by phase.

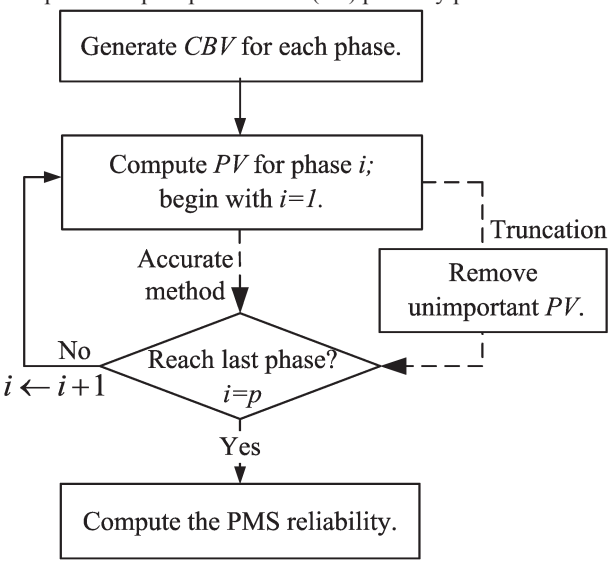


Fig. 1. Flowchart of the proposed method

Step 3 – Sum up the probabilities of the last-phase *PV* to obtain the PMS reliability.

In the following, we first propose our model without the truncation. When the number of phases becomes large, we need to use the truncation step to reduce the time and the space consumption of our model. The flowchart of our approach is given in Fig. 1.

2.2. Model without truncation

The concept of component-behavior vectors (*CBV*) was first proposed in [13]. *CBV*s are in essence equivalent to the BDD paths (readers may refer to [10, 23] for BDD basics). Consider the system whose reliability block diagram (RBD) is shown in Fig. 2. *CBV*s can be obtained through the enumeration of the paths which link the root node and the sink node $\boxed{1}$ in the BDD. In *CBV*s, the matrix $\mathbf{U}_i^{(k)}$ corresponds to the "true" edge of node *k*, while $\mathbf{D}_i^{(k)}$ corresponds to the "false" edge. $\mathbf{E}_i^{(k)}$ is used if the BDD path does not contain the decision node of k. The expressions of $\mathbf{E}_i^{(k)}$, $\mathbf{U}_i^{(k)}$, and $\mathbf{D}_i^{(k)}$ can be find in [13, 25]. For instance, $\mathbf{U}_i^{(k)}$ (for binary-state component) is of the form:

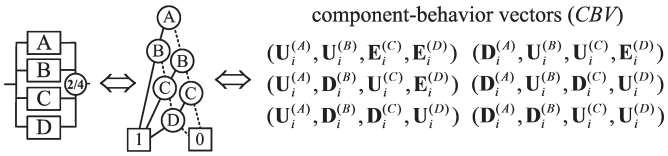


*Fig. 2. Generation of CBV*s

$$\mathbf{E}_i^{(k)} = \exp\left(\begin{pmatrix} -\lambda_i^{(k)} & \lambda_i^{(k)} \\ \mu_i^{(k)} & -\mu_i^{(k)} \end{pmatrix} \cdot T_i\right) \tag{1}$$

where T_i is the duration of phase i. $\lambda_i^{(k)}$ and $\mu_i^{(k)}$ are the failure rates and the repair rates of *k* in phase *i*, respectively.

After we generate *CBV*s for each phase, the reliability of PMS can be assessed through path vectors PV_i. PV_i is defined as:

$$PV_i = \begin{cases} CBV_i, & i=1 \\ PV_{i-1} \circ CBV_i, & i>1 \end{cases} \tag{2}$$

where *i* is the phase index. The operator "$\circ$" represents the element-by-element multiplication, which is also known as the Hadamard product (or the entrywise product) [9]. Take the system in Fig. 3 for instance, there are two *CBV*s in phase 1 and three *CBV*s in phase 2 i.e.,

$$\begin{aligned} CBV_1^{(1)} &= (\mathbf{U}_1^{(A)},\mathbf{E}_1^{(B)},\mathbf{E}_1^{(C)},\mathbf{E}_1^{(D)}) \\ CBV_1^{(2)} &= (\mathbf{D}_1^{(A)},\mathbf{U}_1^{(B)},\mathbf{E}_1^{(C)},\mathbf{E}_1^{(D)}) \end{aligned} \tag{3}$$

$$\begin{aligned} CBV_2^{(1)} &= (\mathbf{U}_2^{(A)},\mathbf{E}_2^{(B)},\mathbf{E}_2^{(C)},\mathbf{E}_2^{(D)}) \\ CBV_2^{(2)} &= (\mathbf{D}_2^{(A)},\mathbf{E}_2^{(B)},\mathbf{U}_2^{(C)},\mathbf{E}_2^{(D)}) \\ CBV_2^{(3)} &= (\mathbf{D}_2^{(A)},\mathbf{E}_2^{(B)},\mathbf{D}_2^{(C)},\mathbf{U}_2^{(D)}) \end{aligned} \tag{4}$$

According to the definition of *PV*s , *PV*s of phase 1 are identical to *CBV*s, i.e.,

$$PV_1^{(q)} = CBV_1^{(q)} \ (q = 1,2,...) \tag{5}$$

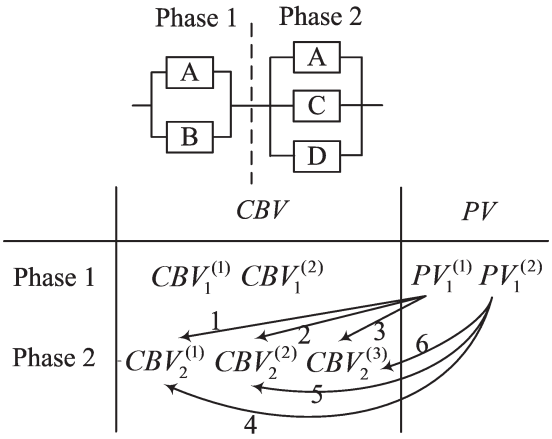


*Fig. 3. Generation of PV*s

*PV*s of phase 2 are the entrywise product of PV_1 and CBV_2. For instance, $PV_2^{(5)}$ (fifth path vector in phase 2) can be expressed as:

$$\begin{aligned} PV_2^{(5)} &= PV_1^{(2)} \circ CBV_2^{(2)} = CBV_1^{(2)} \circ CBV_2^{(2)} \\ &= (\mathbf{D}_1^{(A)}\mathbf{D}_2^{(A)}, \mathbf{U}_1^{(B)}\mathbf{E}_2^{(B)}, \mathbf{E}_1^{(C)}\mathbf{U}_2^{(C)}, \mathbf{E}_1^{(D)}\mathbf{E}_2^{(D)}) \end{aligned} \tag{6}$$

When the final-phase *PV*s are generated, the PMS reliability is the sum of probabilities of the last-phase *PV*s, i.e.,

$$R_{\text{PMS}}(\sum_{j=1}^{p} T_j) = \sum_q \Pr\{PV_p^{(q)}\} \tag{7}$$

where *p* is the index of the last phase. $\Pr\{PV_p^{(j)}\}$ is the probability of $PV_p^{(j)}$. Suppose $PV_p^{(j)} = (\mathbf{a}_1,...,\mathbf{a}_n)$, $\Pr\{PV_p^{(j)}\}$ is defined as:

$$\Pr\{PV_p^{(j)}\} = (\mathbf{v}_0^{(1)} \cdot \mathbf{a}_1 \cdot \mathbf{1}') \cdot (\mathbf{v}_0^{(2)} \cdot \mathbf{a}_2 \cdot \mathbf{1}') \cdot ... \cdot (\mathbf{v}_0^{(n)} \cdot \mathbf{a}_n \cdot \mathbf{1}') \tag{8}$$

where $\mathbf{v}_0^{(k)} = (1,0)$ if the component *k* is initially operational. The column vector $\mathbf{1}' = (1,1)'$ is used to transfer $\mathbf{v}_0^{(k)} \cdot \mathbf{a}_k$ into the scalar. For instance, the probability of $PV_2^{(5)}$ is given by:

$$\begin{aligned} \Pr\{PV_2^{(5)}\} &= \Pr\{(\mathbf{D}_1^{(A)}\mathbf{D}_2^{(A)}, \mathbf{U}_1^{(B)}\mathbf{E}_2^{(B)}, \mathbf{E}_1^{(C)}\mathbf{U}_2^{(C)}, \mathbf{E}_1^{(D)}\mathbf{E}_2^{(D)})\} \\ &= (\mathbf{v}_0^{(A)}\mathbf{D}_1^{(A)}\mathbf{D}_2^{(A)}\mathbf{1}') \cdot (\mathbf{v}_0^{(B)}\mathbf{U}_1^{(B)}\mathbf{E}_2^{(B)}\mathbf{1}') \cdot \\ &\quad (\mathbf{v}_0^{(C)}\mathbf{E}_1^{(C)}\mathbf{U}_2^{(C)}\mathbf{1}') \cdot (\mathbf{v}_0^{(D)}\mathbf{E}_1^{(D)}\mathbf{E}_2^{(D)}\mathbf{1}') \end{aligned}$$

Suppose the PMS contains *p* phases, we can calculate the PMS reliability after we obtain all $\Pr\{PV_p^{(j)}\}$. The proposed algorithm (without truncations) can be summarized by Fig. 4.

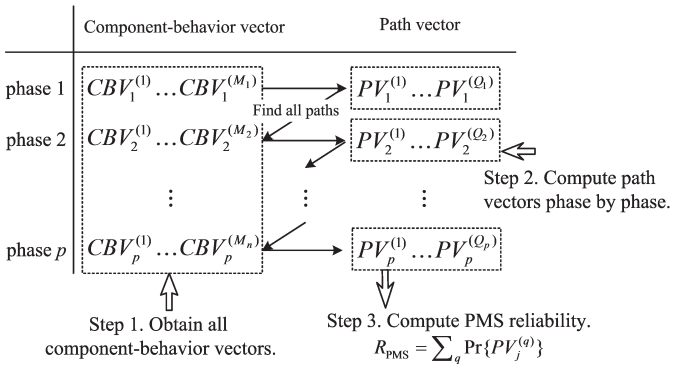


Fig. 4. Proposed algorithm without truncation

2.3. Truncation with flexible threshold

Many real-world PMS may contain thousands of phases which may lead to enormous space consumption in the above algorithm. On this occasion, the truncation step is necessary to reduce the computational complexity. Consider the Fig. 5 PMS which contains several repetitive phases. It is inefficient to calculate all $\Pr\{PV_i\}$ because the number of *PV*s increases exponentially with the phase index *i*. In the BDD & Markov approach [13, 24, 25], a similar problem exists because the top-down algorithm can lead to the BDD path explosion (or bottom-up algorithm leads to BDD node explosion) when the number of phases becomes large.

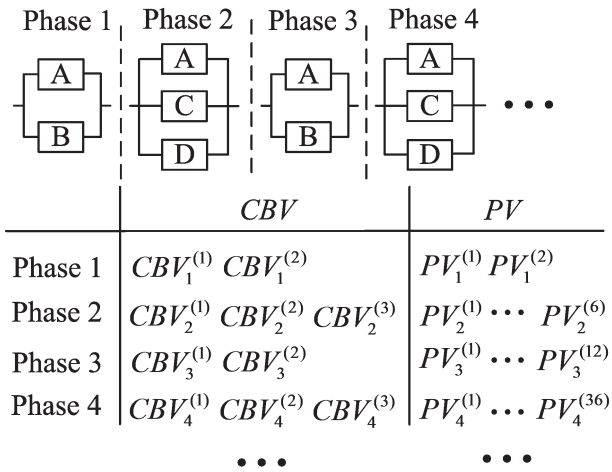


Fig. 5. Exponential increase in the number of PVs

In order to solve the scaling problem, we use the classical truncation methods [5, 11, 12, 17, 19] with fixed truncation limit (i.e. fixed threshold). However, the error analysis associated with a fixed truncation limit may be very complex. In addition, some works [5, 7] show that the truncation error is highly sensitive to the component parameters. Here, we present a decreasing truncation limit to guarantee that the truncation error is less than the predefined "maximum permissible error" (*MPE*). The flexible truncation limit γ_i of phase *i* is defined as:

$$\gamma_i = \frac{MPE - \Pr\{PV\}}{Num(PV_i)} \tag{9}$$

where $Num(PV_i)$ is the number of *PV*s which are removed during phase *i*. With , we remove PV_i if $\Pr\{PV_i\} < \gamma_i$. Every time we de-

lete a path vector, the value of γ_i will decrease. *MPE* is a predefined algorithm parameter.

In the following, we demonstrate that the total truncation error is less than (or equals to) MPE. Let PV_i^{Δ} to be the path vector removed when the algorithm reaches phase *i*. By removing PV_i^{Δ}, we also eliminate some *PV*s in the later phase. Take Fig. 6 for example, suppose $PV_2^{(1)}$ (first *PV* in phase 2) is removed during phase 2. *PV*s stem from $PV_2^{(1)}$ are also removed (including $PV_3^{(1)}$, $PV_3^{(2)}$, and $PV_4^{(1)} \sim PV_4^{(6)}$). Since $\sum_i \Pr\{CBV_i\} < 1$, we can see that $\Pr\{PV_i^{\Delta}\}$ is bigger than (or equals to) the sum of probabilities of *PV*s which stem from PV_i^{Δ}, i.e.,

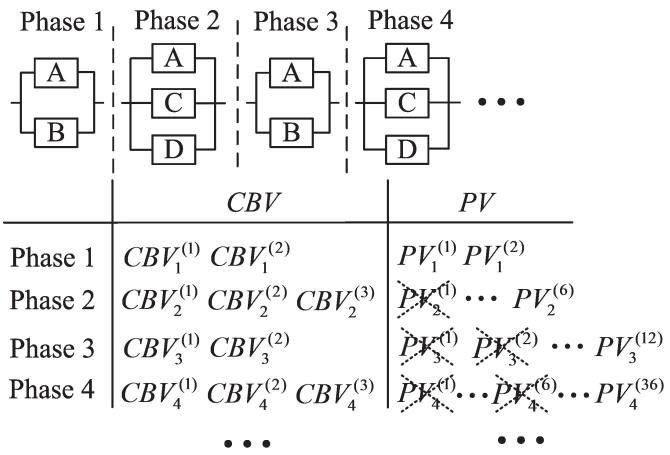


Fig. 6. Truncation of PVs

$$\Pr\{PV_i^{\Delta}\} \geq \sum \Pr\{PV_{i+1}^{*}\};\ (PV_{i+1}^{*} \text{ stems from } PV_i^{\Delta}) \qquad (10)$$

By repeating Eq. (10), we have:

$$\Pr\{PV_i^{\Delta}\} \geq \sum \Pr\{PV_{i+1}^{*}\} \geq \ldots \geq \sum_s \Pr\{PV_q^{(s)}\} \qquad (11)$$

where $PV_q^{(s)}$ is the last-phase *PV*s which stem from PV_i^{Δ}. When the algorithm reaches the last phase, we sum up the probabilities of all removed PV_i^{Δ}, and then:

$$MPE \geq \sum_i \Pr\{PV_i^{\Delta}\} \geq \sum_i \sum_s \Pr\{PV_q^{(s)}\} = \left|R_{\text{PMS}} - R'_{\text{PMS}}\right| = error \qquad (12)$$

where $MPE \geq \sum_i \Pr\{PV_i^{\Delta}\}$ holds because of the definition of γ_i. The inequality (12) shows that the total truncation error is less than the user-defined *MPE*. Overall, the proposed method with truncations can be summarized by Fig. 7.

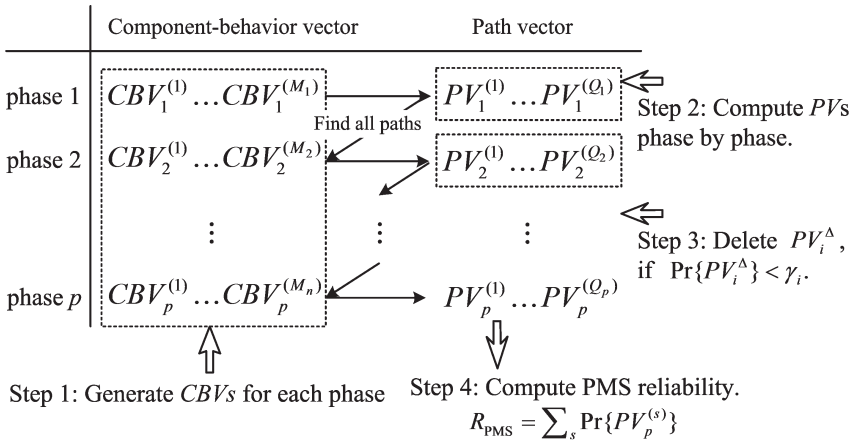


Fig. 7. Proposed algorithm with truncations

2.4. Determination of *MPE*

From above algorithm, we find that a smaller *MPE* parameter leads to a more accurate reliability result, but at the cost of more memory and time expense. Generally, *MPE* is determined by the following steps:

Step 1 – Set *MPE*=0 and check whether the algorithm is able to generate a reliability result. Empirically, if the algorithm cannot generate a result within 3 min, it suggests that the memory space is not sufficient (computing with virtual memory is unacceptably time-consuming). In this case, *MPE* should increase (see Step 2).

Step 2 – Let *MPE*=10^{-9}, and run the algorithm to see the algorithm performance. If the algorithm still encounter the problem of insufficient memory, move to Step 3.

Step 3 – Increase *MPE* by 10 times, and run the truncation algorithm with new *MPE*. Repete Step 3 until the truncation algorithm is able to generate a result. The appropriate *MPE* is the first value which guarantees the algorithm can generates a result.

For most PMS cases, our truncation algorithm is able to generate a result with *MPE*=0.1. Above 3-step procedure can be summarized as Fig. 8. In summery, *MPE* is determined by the case concerned and the performance of user's computer. We need to test different *MPE* sample to determine the appropriate *MPE*.

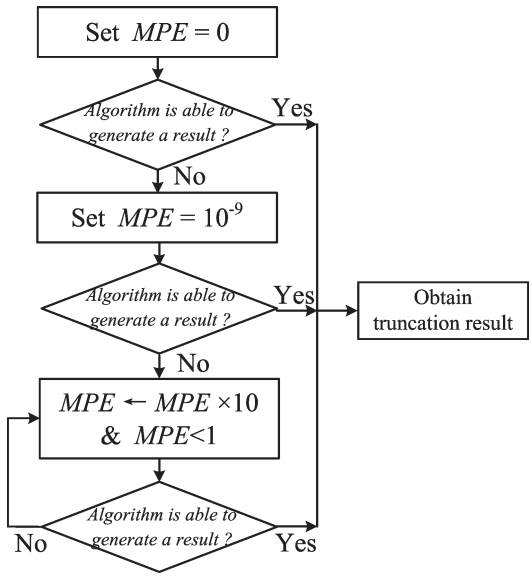


Fig. 8. Procedure to determine maximum permissible error

3. Case study

3.1. Train-speed monitoring mission

The surveillance of the train speed is an important mission in railway administration. The speed of the train can be measured by the sensors installed near railway stations, or by the radars installed along the railway. Consider a train which travels across three cities for two laps in a day, as shown in Fig. 9. The monitoring system is operational as long as two sets of data can be simultaneously recorded. This system can be modeled as the PMS whose RBD is shown in Fig. 10.

Suppose all equipments in Fig. 10 are repairable, there are three kinds of methods available for the reliability analysis of this PMS. Apparently, the conventional Markovian method [1] is not efficient because of the enumeration of 2^{11} states. The BDD & Markov method [13, 24, 25] and our approach can generate the results using less space than the Markovian approach. In comparison with the BDD & Markov method, the main advantage of our approach is the truncation step which significantly reduces the computation time (see the experimental results in Table 2). In this PMS, we assume the equipments may fail even if they are idle. The system parameters are presented in Table 1.

In our approach, we first generate *CBV*s of phases 1-6 (see Fig. 11). *CBV*s of phases 7-12 are the same as that of phases 1-6. Secondly, we calculate PVs phase by phase using. For instance, $PV_3^{(1)}$ is of the form:

$$PV_3^{(1)} = (\mathbf{U}_1^{(S1)}\mathbf{E}_2^{(S1)}\mathbf{E}_3^{(S1)}, \mathbf{E}_1^{(S4)}\mathbf{E}_2^{(S4)}\mathbf{U}_3^{(S4)}) \tag{13}$$

Finally, given $MPE=10^{-6}$, the PMS reliability is given by:

$$R_{\mathrm{PMS}}(\sum_{j=1}^{18} T_j) = \sum_s \Pr\{PV_{18}^{(s)}\} \tag{14}$$

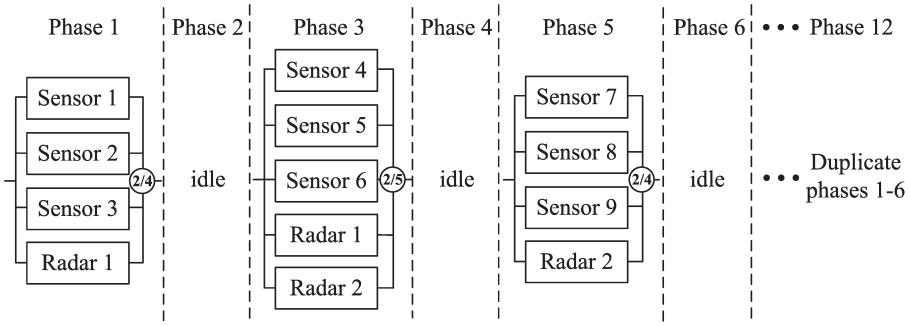


Fig. 9. Train-speed monitoring mission

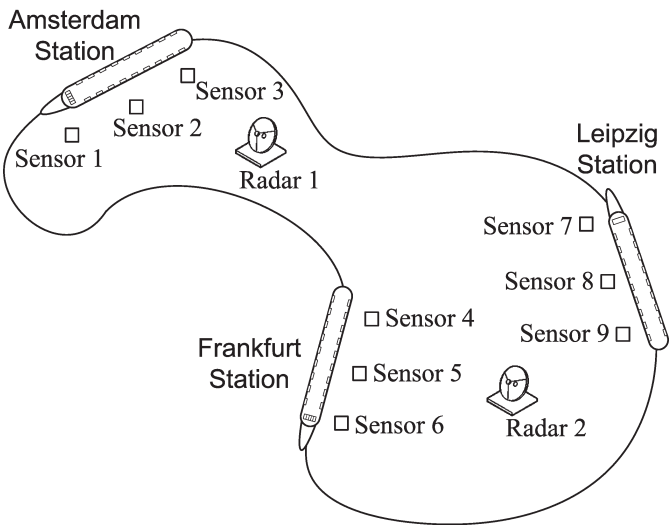


Fig. 10. PMS model for the speed-monitoring mission

Table 1. Parameters of the speed-monitoring PMS

Parameters	Value (Time unit: hour)
Phase duration T_i (i - phase index)	$T_i = \begin{cases} 0.2, & i = 1,3,5,...,17 \\ 3, & i = 2,4,6,...,18 \end{cases}$
Failure rates	$\lambda_K = \begin{cases} 2\cdot 10^{-3}, & K = \text{Sensors} \\ 10^{-3}, & K = \text{Radars} \end{cases}$ (λ_K remains fixed in all phases.)
Repair rates μ_K	$\mu_K = 0.01$ (for all components) (μ_K remains fixed in all phases.)

	CBV	*PV*
Phase 1	$CBV_1^{(1)} = (\mathbf{U}_1^{(S1)}, \mathbf{U}_1^{(S2)}, \mathbf{E}_1^{(S3)}, \mathbf{E}_1^{(R1)})$ $CBV_1^{(2)} = (\mathbf{U}_1^{(S1)}, \mathbf{D}_1^{(S2)}, \mathbf{U}_1^{(S3)}, \mathbf{E}_1^{(R1)})$ $CBV_1^{(3)} = (\mathbf{U}_1^{(S1)}, \mathbf{D}_1^{(S2)}, \mathbf{D}_1^{(S3)}, \mathbf{U}_1^{(R1)})$ $CBV_1^{(4)} = (\mathbf{D}_1^{(S1)}, \mathbf{U}_1^{(S2)}, \mathbf{U}_1^{(S3)}, \mathbf{E}_1^{(R1)})$ $CBV_1^{(5)} = (\mathbf{D}_1^{(S1)}, \mathbf{U}_1^{(S2)}, \mathbf{D}_1^{(S3)}, \mathbf{U}_1^{(R1)})$ $CBV_1^{(6)} = (\mathbf{D}_1^{(S1)}, \mathbf{D}_1^{(S2)}, \mathbf{U}_1^{(S3)}, \mathbf{U}_1^{(R1)})$	$CBV_1^{(1)} \sim CBV_1^{(6)}$
Phase 3	$CBV_3^{(1)} = (\mathbf{U}_3^{(S4)}, \mathbf{U}_3^{(S5)}, \mathbf{E}_3^{(S6)}, \mathbf{E}_3^{(R1)}, \mathbf{E}_3^{(R2)})$ $CBV_3^{(2)} = (\mathbf{U}_3^{(S4)}, \mathbf{D}_3^{(S5)}, \mathbf{U}_3^{(S6)}, \mathbf{E}_3^{(R1)}, \mathbf{E}_3^{(R2)})$ $CBV_3^{(3)} = (\mathbf{U}_3^{(S4)}, \mathbf{D}_3^{(S5)}, \mathbf{D}_3^{(S6)}, \mathbf{U}_3^{(R1)}, \mathbf{E}_3^{(R2)})$ $CBV_3^{(4)} = (\mathbf{U}_3^{(S4)}, \mathbf{D}_3^{(S5)}, \mathbf{D}_3^{(S6)}, \mathbf{D}_3^{(R1)}, \mathbf{U}_3^{(R2)})$ $CBV_3^{(5)} = (\mathbf{D}_3^{(S4)}, \mathbf{U}_3^{(S5)}, \mathbf{U}_3^{(S6)}, \mathbf{E}_3^{(R1)}, \mathbf{E}_3^{(R2)})$ $CBV_3^{(6)} = (\mathbf{D}_3^{(S4)}, \mathbf{U}_3^{(S5)}, \mathbf{D}_3^{(S6)}, \mathbf{U}_3^{(R1)}, \mathbf{E}_3^{(R2)})$ $CBV_3^{(7)} = (\mathbf{D}_3^{(S4)}, \mathbf{U}_3^{(S5)}, \mathbf{D}_3^{(S6)}, \mathbf{D}_3^{(R1)}, \mathbf{U}_3^{(R2)})$ $CBV_3^{(8)} = (\mathbf{D}_3^{(S4)}, \mathbf{D}_3^{(S5)}, \mathbf{U}_3^{(S6)}, \mathbf{U}_3^{(R1)}, \mathbf{E}_3^{(R2)})$ $CBV_3^{(9)} = (\mathbf{D}_3^{(S4)}, \mathbf{D}_3^{(S5)}, \mathbf{U}_3^{(S6)}, \mathbf{D}_3^{(R1)}, \mathbf{U}_3^{(R2)})$ $CBV_3^{(10)} = (\mathbf{D}_3^{(S4)}, \mathbf{D}_3^{(S5)}, \mathbf{D}_3^{(S6)}, \mathbf{U}_3^{(R1)}, \mathbf{U}_3^{(R2)})$	$PV_3^{(1)} \sim PV_3^{(60)}$
Phase 5	$CBV_5^{(1)} = (\mathbf{U}_5^{(S7)}, \mathbf{U}_5^{(S8)}, \mathbf{E}_5^{(S9)}, \mathbf{E}_5^{(R2)})$ $CBV_5^{(2)} = (\mathbf{U}_5^{(S7)}, \mathbf{D}_5^{(S8)}, \mathbf{U}_5^{(S9)}, \mathbf{E}_5^{(R2)})$ $CBV_5^{(3)} = (\mathbf{U}_5^{(S7)}, \mathbf{D}_5^{(S8)}, \mathbf{D}_5^{(S9)}, \mathbf{U}_5^{(R2)})$ $CBV_5^{(4)} = (\mathbf{D}_5^{(S7)}, \mathbf{U}_5^{(S8)}, \mathbf{U}_5^{(S9)}, \mathbf{E}_5^{(R2)})$ $CBV_5^{(5)} = (\mathbf{D}_5^{(S7)}, \mathbf{U}_5^{(S8)}, \mathbf{D}_5^{(S9)}, \mathbf{U}_5^{(R2)})$ $CBV_5^{(6)} = (\mathbf{D}_5^{(S7)}, \mathbf{D}_5^{(S8)}, \mathbf{U}_5^{(S9)}, \mathbf{U}_5^{(R2)})$	$PV_5^{(1)} \sim PV_5^{(360)}$
Phase 2, 4, 6	$CBV_i^{(1)} = (\mathbf{E}_i^{(S1)}, ..., \mathbf{E}_i^{(R2)})$ $(i = 2,4,6)$	

Fig. 11. CBVs of phase 1-6

Table 2. Reliability and computation time for the speed-monitoring PMS

Time (Hour)	Traditional BDD & Markov		Proposed method ($MPE=10^{-6}$)		Petri-net simulation
	PMS reliability	Computation time (s)	PMS reliability	Computation time (s)	PMS reliability
9.6 (phase-6 end)	0.999995	0.07	0.999995	0.04	0.999995
19.2 (phase-12 end)	0.999919	23.9	0.999918	0.93	0.999920

In order to verify the correctness of our algorithm, we compare the non-simulation approaches with the Petri-net simulation. From Table 2, we can see that three kinds of methods generate the similar reliabilities for this PMS. The non-simulation algorithms are programmed by the MATLAB language, and implemented with a 1.8 GHz processor. The average computation time in Table 2 is recorded without multi-threading. And we can see that our approach is much less time consuming than the BDD & Markov method. In the Petri-net simulation, the model contains one Petri net for the phase-index increment (with low priority), one Petri net for the detection of system failure, and 11 Petri nets for the independent equipments. We carry out the Petri-net simulation using the GRIF software [8] with 10^7 iterations. In the next instance, we show that the truncation step is indispensable in avoiding the excessive memory cost of the traditional BDD & Markov method.

3.2. Regular test mission in the oil and gas system

In the oil and gas industry, the regular test is one essential part of safety procedures. Consider a pressure protection system which is used to stop flow in case of overpressure. Equipments in the system may suffer from higher pressure during the test phases than the normal phases. Hence, the pressure protection system can be model as the PMS whose structure remains fixed during the entire mission, as shown in Fig. 12.

It can be seen from Fig. 12 that the phase 1 and the phase 2 are duplicated as the test is periodically carried out. In this PMS, components can be modeled as repairable if they can be quickly renewed after failure. Generally, the duration of the test phase is much shorter than that of the normal phases. Here, we suppose the life and the repair time are exponentially distributed. Relevant parameters of the PMS are presented in Table 3.

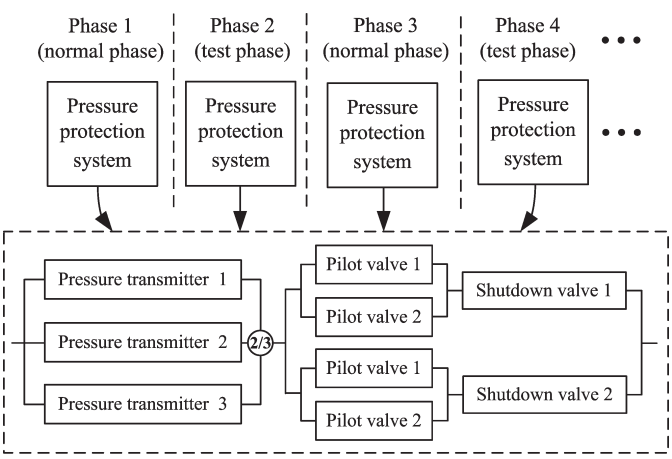


Fig. 12. PMS model for the pressure protection system

In the traditional BDD & Markov method, the time cost and the space cost increases exponentially with the number of phases in the PMS. The experimental results in Table 4 show that the computing capacity of the BDD & Markov algorithm is 5 phases. Further computation will result in the problem of the insufficient computer memory. However, when we use the proposed flexible truncation (with MPE=10-5), the computing capacity of our approach increases to 8 phases. The computing capacity will increase if *MPE* gets bigger. In our experiment, the algorithms are programmed without multi-threading, and run on a 1.8 GHz processor to record the computation time.

Table 3. Parameters of the pressure-protection PMS

Parameters	Value (Time unit: hour)
Phase duration T_i	$T_i = \begin{cases} 168, \text{ normal phase.} \\ 2, \quad \text{test phase.} \end{cases}$
Failure rates λ_K	$\lambda_K = \begin{cases} 10^{-4}, \quad \text{normal phase.} \\ 5 \cdot 10^{-4}, \text{ test phase.} \end{cases}$ (for all components)
Repair rates μ_K	$\mu_K = 0.05$ (for all components)

Table 4. Reliability and computation time for the pressure-protection PMS

Time (Hour)	Traditional BDD & Markov		Proposed method ($MPE=10^{-5}$)		Petri-net simulation
	PMS reliability	Computation time (s)	PMS reliability	Computation time (s)	PMS reliability
170 (end of phase 2)	0.998871	0.02	0.998871	0.02	0.998867
340 (end of phase 4)	0.997389	7.23	0.997388	1.22	0.997468
510 (end of phase 5)	0.995930	146	0.995927	3.73	0.996095
680 (end of phase 8)	Insufficient memory		0.994425	185	0.994594

The proposed method is validated by the Petri-net simulation using the GRIF software [8] with 10^7 iterations. In the Petri-net model, there is one Petri-net determining the phase transition, one Petri net determining the PMS failure, and 9 Petri nets determining the states of every component. By comparing the BDD & Markov method with our approach, we can see that the truncation step is indispensable to obtain the PMS reliability. In many real-world applications, the PMS model often contains a large number of repetitive phases, and the components in the PMS are usually considered as repairable. The proposed method is efficient to analyze this kind of PMS.

4. Conclusion

In the modeling of practical PMS, the system normally consists of several subsystems, and the subsystems may comprise many equipments. To keep the PMS model concise, reliability engineers usually

neglect the details of equipment configurations, and consider each subsystem as an entity whose behavior is exponentially distributed. Hence, the number of components in the PMS may not be very large if the engineers want a small-scale model. However, most PMS models cannot avoid a large number of phases because the real-world mission usually contains many repetitive tasks. This paper proposes an efficient approach for the reliability assessment of these PMS.

When the PMS model contains a multitude of phases, existing approaches may suffer from the BDD-explosion problem. Our approach uses truncations to analyze the PMS with many phases, and uses BDD to analyze the PMS with many components. In our approach, the decreasing truncation limit can keep the truncation error within the user-defined maximum permissible error. The case study shows that the truncation is a necessary step to solve the explosion problem as the number of phases increases. A possible direction of future works is to explore the application of our method in the field of the aerospace, assembling, and nuclear power industry, and to explore better truncation strategies using classic methods.

Acknowledgement

This research is partly carried out with the Reliability, Availability, Maintainability and Safety (RAMS) group at the Norwegian University of Science and Technology (NTNU), and it is funded by the National Natural Science Foundation of China with the grant no. 71071159.

References

1. Alam M, Alsaggaf UM. Quantitative reliability evaluation of repairable phased-mission systems using Markov approach. IEEE Transactions on Reliability 1986; 35(5): 498-503, http://dx.doi.org/10.1109/TR.1986.4335529.
2. Amari SV, Xing LD, Shrestha A et al. Performability analysis of multistate computing systems using multivalued decision diagrams. IEEE Transactions on Computers, 2010; 59(10): 1419-1433, http://dx.doi.org/10.1109/TC.2009.184.
3. Burdick G, Fussell J, Rasmuson D et al. Phased mission analysis: A review of new developments and an application. IEEE Transactions on Reliability 1977; 26(1): 43-49, http://dx.doi.org/10.1109/TR.1977.5215072.
4. Cekyay B, Ozekici S. Performance measures for systems with Markovian missions and aging. IEEE Transactions on Reliability 2012; 61(3): 769-778, http://dx.doi.org/10.1109/TR.2012.2207529.
5. Čepin M. Analysis of truncation limit in probabilistic safety assessment. Reliability Engineering and System Safety 2005; 87(3): 395-403, http://dx.doi.org/10.1016/j.ress.2004.06.009.
6. Chew SP, Dunnett SJ, Andrews JD. Phased mission modelling of systems with maintenance-free operating periods using simulated Petri nets. Reliability Engineering and System Safety 2008; 93(7): 980-994, http://dx.doi.org/10.1016/j.ress.2007.06.001.
7. Epstein S, Rauzy A. Can we trust PRA? Reliability Engineering and System Safety 2005; 88(3): 195-205, http://dx.doi.org/10.1016/j.ress.2004.07.013.
8. Grif-workshop software, SATODEV LLC. 2014. http://grif-workshop.com/
9. Horn RA, Johnson C. The Hadamard Product - Chapter 5 - Topics in matrix analysis. London: Cambridge University Press, 1990
10. Huang HZ, Zhang H, Li Y. A new ordering method of basic events in fault tree analysis. Quality and Reliability Engineering International 2012; 28(3): 297-305, http://dx.doi.org/10.1002/qre.1245.
11. Jung WS, Han SH, Yang JE. Fast BDD truncation method for efficient top event probability calculation. Nuclear Engineering and Technology 2008; 40: 571-580, http://dx.doi.org/10.5516/NET.2008.40.7.571.
12. Jung WS, Yang J-E, Ha J. Development of measures to estimate truncation error in fault tree analysis. Reliability Engineering and System Safety 2005; 90(1): 30-6, http://dx.doi.org/10.1016/j.ress.2004.09.007.
13. Lu J-M, Wu X-Y. Reliability evaluation of generalized phased-mission systems with repairable components. Reliability Engineering and System Safety 2014; 121: 136-145, http://dx.doi.org/10.1016/j.ress.2013.08.005.
14. Lu J-M, Wu X-Y, Liu Yiliu, Mary Ann Lundteigen. Reliability analysis of large phased-mission systems with repairable components based on success-state sampling. Reliability Engineering and System Safety 2015; 142: 123-133, http://dx.doi.org/10.1016/j.ress.2015.05.010.
15. Mo YC. New insights into the BDD-based reliability analysis of phased-mission systems. IEEE Transactions on Reliability 2009; 58(4): 667-678, http://dx.doi.org/10.1109/TR.2009.2026804.
16. Mo YC. Variable ordering to improve BDD analysis of phased-mission systems with multimode failures. IEEE Transactions on Reliability, 2009; 58(1): 53-57, http://dx.doi.org/10.1109/TR.2008.2011673.
17. Mo YC, Han J, Zhang Z et al. Approximate reliability evaluation of large-scale distributed systems. Journal of Information Science and Engineering 2014; 30(1): 25-41.
18. Mo YC, Xing L, Amari SV. A multiple-valued decision diagram based method for efficient reliability analysis of non-repairable phased-mission systems. IEEE Transactions on Reliability 2014; 63(1): 320-330, http://dx.doi.org/10.1109/TR.2014.2299497.
19. Mo YC, Zhong F, Zhao X. New results to BDD truncation method for efficient top event probability calculation. Nuclear Engineering and technology 2012; 44: 755-766, http://dx.doi.org/10.5516/NET.03.2011.058.
20. Mura I, Bondavalli A. Markov regenerative stochastic Petri nets to model and evaluate phased mission systems dependability. IEEE Transactions on Computers 2001; 50(12): 1337-1351, http://dx.doi.org/10.1109/TC.2001.970572.
21. Ou Y, Dugan JB. Modular solution of dynamic multi-phase systems. IEEE Transactions on Reliability 2004; 53(4): 499-508, http://dx.doi.org/10.1109/TR.2004.837305.
22. Peng R, Zhai QQ, Xing LD et al. Reliability of demand-based phased-mission systems subject to fault level coverage. Reliability Engineering and System Safety 2014; 121: 18-25, http://dx.doi.org/10.1016/j.ress.2013.07.013.
23. Rauzy A. Binary decision diagrams for reliability studies. Handbook of performability engineering 381-396. Berlin: Springer Press, 2008, http://dx.doi.org/10.1007/978-1-84800-131-2_25.
24. Shrestha A, Xing LD, Dai YS. Reliability analysis of multistate phased-mission systems with unordered and ordered states. IEEE Transactions on Systems Man and Cybernetics Part a - Systems and Humans 2011; 41(4): 625-636, http://dx.doi.org/10.1109/TSMCA.2010.2089513.
25. Wang D, Trivedi KS. Reliability analysis of phased-mission system with independent component repairs. IEEE Transactions on Reliability 2007; 56(3): 540-551, http://dx.doi.org/10.1109/TR.2007.903268.

26. Wu Xin-Yang, Wu X-Y. Extended object-oriented Petri net model for mission reliability simulation of repairable pms with common cause failures. Reliability Engineering and System Safety 2015; 136: 109-119, http://dx.doi.org/10.1016/j.ress.2014.11.012.
27. Xing L. Reliability evaluation of phased-mission systems with imperfect fault coverage and common-cause failures. IEEE Transactions on Reliability 2007; 56(1): 58-68, http://dx.doi.org/10.1109/TR.2006.890900.
28. Xing L, Dugan JB, Morrissette BA. Efficient reliability analysis of systems with functional dependence loops. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2009; 43(3): 65-69.
29. Xing L, Levitin G. Bdd-based reliability evaluation of phased-mission systems with internal/external common-cause failures. Reliability Engineering and System Safety 2013; 112: 145-153, http://dx.doi.org/10.1016/j.ress.2012.12.003.
30. Zang X, Sun H, Trivedi KS. A BDD-based algorithm for reliability analysis of phased-mission systems. IEEE Transactions on Reliability 1999; 48(1): 50-60, http://dx.doi.org/10.1109/24.765927.
31. Zhang T, Bai GH, Guo B. Success probability model of phased mission systems with limited spares. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2012; 14(1): 24-32.

Ji-Min LU
College of Information System and Management
National University of Defense Technology
De Ya Road 109, Changsha, Hunan 410073, China

Mary Ann LUNDTEIGEN
Yiliu LIU
Department of Production and Quality Engineering
Norwegian University of Science and Technology, Valgrinda,
N-7491 Trondheim , Norway

Xiao-Yue WU
College of Information System and Management
National University of Defense Technology
De Ya Road 109, Changsha, Hunan 410073, China

E-mails: jeemanlv@gmail.com, mary.a.lundteigen@ntnu.no,
yiliu.liu@ntnu.no, xiaoyuewucn@gmail.com

Article citation info:
TOPIĆ D, ŠLJIVAC D, STOJKOV M. Reliability model of different wind power plant configuration using sequential monte carlo simulation. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 237–244, http://dx.doi.org/10.17531/ein.2016.2.11.

Danijel TOPIĆ
Damir ŠLJIVAC
Marinko STOJKOV

RELIABILITY MODEL OF DIFFERENT WIND POWER PLANT CONFIGURATION USING SEQUENTIAL MONTE CARLO SIMULATION

MODEL NIEZAWODNOŚCI RÓŻNYCH KONFIGURACJI ZESTAWU ELEKTROWNI WIATROWEJ OPARTY NA SEKWENCYJNEJ SYMULACJI MONTE CARLO

Paper presents an enhanced model for calculation of reliability indices for different wind power plants configuration concepts used over past two decades. The autoregressive – moving average (ARMA) model is used combined with the sequential Monte Carlo simulation in order to predict expected energy not served (EENS) more accurately during the failure. Statistical database of LWK (Land Wirtschafts Kammer) is used for determining different wind power plant configuration types component reliability (performance) used for calculating influence of individual wind power plant configuration concepts on expected energy not served. Furthermore, a comparison of the distribution of EENS of different wind power plants configuration concepts have been presented, as well as the influence of the predominantly mechanical and electrical components failures on both EENS and failure rates.

***Keywords**: wind turbine; components; reliability; failure rate; Monte Carlo, ARMA.*

W pracy przedstawiono udoskonalony model służący do obliczania wskaźników niezawodności dla różnych koncepcji konfiguracji zestawów elektrowni wiatrowych jakie stosowano w ostatnich dwóch dziesięcioleciach. Wykorzystano autoregresyjny model średniej ruchomej (ARMA), który w połączeniu z symulacją sekwencyjną Monte Carlo pozwala z większą dokładnością przewidzieć oczekiwaną wartość energii niedostarczonej (EENS) podczas awarii. Baza statystyczna LWK (Land Wirtschafts Kammer) posłużyła autorom do określania niezawodności (wydajności) części składowych elektrowni wiatrowych przy różnych typach konfiguracji zestawu. Otrzymane wartości wykorzystano do obliczenia wpływu poszczególnych koncepcji konfiguracji zestawu elektrowni wiatrowej na oczekiwaną wartość energii niedostarczonej. Ponadto, przedstawiono porównanie rozkładu EENS dla różnych koncepcji konfiguracji zestawu elektrowni wiatrowej jak również omówiono wpływ uszkodzeń części mechanicznych i elektrycznych elektrowni na EENS oraz awaryjność.

***Słowa kluczowe**: turbina wiatrowa; części składowe, niezawodność; awaryjność; Monte Carlo, ARMA.*

1. Introduction

According to the International Energy Agency (www.iea.org) in 2009 total electricity generation share of the nuclear power plants was 13.4 %, natural gas 21.4 %, oil 5.1 % and coal power plants 40.6 %, while share of renewables (without large hydropower plants – 16.2 %) was 3.3 %. Those non-renewable power generation units, though generally highly dispatchable and reliable, have in turn huge environmental influence, especially on the global warming, due to the great share of fossil fuels usage [17].

In recent years, share of renewables in total electricity generation is increasing: excluding big hydropower plants, the greatest share in electricity generation from renewables is coming from the wind power plants. Total installed capacity in renewables in 2013 was 560 GW respectively, out of which 56.7 % or 318 GW from wind power plants [28]. In regard to a conventional power plant (thermal, hydro and nuclear) which implies adequacy of primary energy and constant power generation, wind power plants are depending on wind speed that is highly variable.

Due to the great increase of the capacity of wind power plants in recent years, their reliability is more and more important. That resulted in many wind turbine reliability models and methods being developed. However, due to the recent availability of long-term wind power plant statistics and fast development of wind turbine technology and size, it is important to continue developing new and more accurate wind turbine reliability models. This paper is an effort in that direction presenting advanced reliability modelling for different wind turbine types, taking into account different wind power plant configurations developed over the past years and component performance statistics.

The former developed reliability models of wind power plants in [5, 7, 8, 22, 23, 28] have so far related to the wind power plant as a whole. These reliability models have yielded a detailed overview of the impact of wind power plants on the reliability and availability, but they were not taking into account the impact of individual components of the each single wind power plant on the calculation of reliability indices of wind power plant. In [3] an analytical model is presented that describes in detail the reliability of wind power plant by taking into account diversity of the wind power plant configuration from the generator aspect. Also, according to [2] and described in chapter 2, there are four dominant configuration groups of wind power plants differed by type of generator, by network connection, by power control and by speed rate.

2. Wind turbine's construction characteristics

Wind turbine's main components are namely: rotor (hub and blades), brake, gearbox, generator, electronic control system, yaw system, nacelle, drive train, anemometer and tower. According to [2],

there are four typically configuration concepts of a wind turbine: Type A, Type B, Type C and Type D.

Type A configuration presents wind turbine with constant speed, stall power regulation and squirrel cage induction generator. Type B configuration corresponds to the limited variable speed wind turbine with variable generator rotor resistance, known as OptiSlip. It uses a wound rotor induction generator (WRIG) and has been used by the Danish manufacturer Vestas since the mid-1990s. The generator is directly connected to the grid [2]. Type C, which is known as the doubly fed induction generator (DFIG) concept, corresponds to the limited variable speed wind turbine with a wound rotor induction generator (WRIG) and partial scale frequency converter (rated at approximately 30 % of nominal generator power) on the rotor circuit [2]. Type D configuration corresponds to the full variable-speed wind turbine, with the generator connected to the grid through a full-scale frequency converter. Some full variable-speed wind turbine systems have no gearbox [2].

3. Statistical data of a wind turbine components

For a wind turbine reliability assessment, it is very important to have statistical data on performances and failures of wind turbine components. Detailed data on performance of wind turbines were presented by Windstats [35]. Windstats is a commercial newsletter for the wind industry and records details of performance of wind turbines in many countries [33]. Tavner, Xiang and Spinato have performed detailed analysis of Windstats reliability data for German (WSD) and Danish (WSDK) wind turbines for 10-year period from October 1994 to September 2004 presented in [33]. Except the Windstats data, Spinato, Tavner, van Bussel and Koutoulakos in [31] have also analysed reliability data for wind turbines obtained from a survey performed by the Land Wirtschafts Kammer (termed in this paper LWK) in Schleswig Holstein, Germany. LWK data are based on 11-year long period. Also, data on the performance of wind turbines extracted from WMEP (Wissenschaftliches Mess und Evaluierungsprogramm) are available. WMEP data are based on 17-year period of research and include more than 1500 wind turbines. Those data were analysed and presented in [13, 16] and [32]. Other performance statistics of reliability of wind for Sweden, Germany and Finland were presented in [29] and [30]. In this paper will be used data from LWK database. In research presented in this paper, reliability analysis for type A configuration of wind turbine will be performed using LWK database for wind turbine Nordex N52/N54, for type B configuration of wind turbine will be performed using LWK database for wind turbine Vestas V39 and for type D configuration of a wind turbine will be performed using LWK database for wind turbine Enercon E66. For type C configuration of wind turbine there are no available data and this type of configuration will not be taken in to consideration. Table 1 presents statistical data on the failure rates and downtimes of particular components of a wind turbine according to LWK database [31].

Table 1. Reliability data for different wind turbine types according to LWK database [31]

Wind turbine component	Failure rate [1/year]			Downtime [h/failure]
	Nordex N52/N54 (Type A)	Vestas V39 (Type B)	Enercon E66 (Type D)	
Electrical system	0.28	0.34	0.50	255
Electronic system	0.15	0.27	0.31	60
Generator	0.11	0.09	0.13	160
Hydraulic system	0.19	0.26	-	70
Yaw system	0.12	0.10	0.17	60
Mechanical brake	0.08	0.01	0.02	45
Sensors	0.03	0.08	0.26	42
Anemometer	0.13	0.06	0.07	4
Pitch control	-	0.10	0.47	70
Rotor and blades	0.46	0.17	0.14	125
Gearbox	0.51	0.18	-	335
Air brake	0.06	-	-	110
Shaft/bearings	0.07	-	0.05	130
Other	0.43	0.20	0.24	65

4. Model description

4.1. Wind speed model

When it comes to a reliability assessment of a wind turbine, it is very important to have data about wind speed because output power of wind turbine is directly dependent on wind speed as previously described. For the reliability assessment of a wind turbine, data about wind speed can be used directly or, as it will be described below, can be modelled. In the previous research wind speed with the time-series (autoregressive, moving average or autoregressive moving average model), Weibull, Rayleigh or normal distribution or Markov chain was modelled.

Billinton, Chen and Ghajar in [6] have modelled the wind speed with time series (autoregressive moving average - ARMA) and in described in detail. Also, many other authors have used ARMA model for wind speed modelling as in [1, 5 ,7 ,8, 21, 22, 23, 27, 28]. Deshmukh and Ramakumar in [11] have used a Weibull distribution for wind speed modelling. Also, Weibull distribution for wind speed modelling in [3, 10, 12] and [20] was used. Giorseto and Utsurogi in [14], Wang, Dai, Hui and Thomas in [34] and Attvwa and El-Saadany in [4] have used Rayleigh distribution for wind speed modelling. Normal distribution for wind speed modelling in [15, 18, 19, 25] and Markov chain in [9, 24, 26] was used.

In this paper the wind speed model described in [6] is used. In this paper wind speed with the time series will be modelled. General form of the time series (ARMA model) is given with the following expression:

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + ...\Phi_n y_{t-n} + \alpha_t - \Theta_1 \alpha_{t-1} - \Theta_2 \alpha_{t-2} - ... - \Theta_m \alpha_{t-m} \tag{1}$$

Where: $\Phi_i = 1,2,...,n$ autoregressive parameters, $\Theta_i = 1,2,...,m$ moving average parameters, $\{\alpha_t\}$ white noise with normally distributed with mean zero and variance σ_a^2. Wind speed model according to [6] with the following equation can be expressed:

$$V_t = \mu + y_t \tag{2}$$

Where is: V_t – simulated wind speed in the hour t, μ – mean value of the wind speed of the all observed measured data, y_t – time series of the wind speed described with expression (1).

4.2. Wind power plant model

In this chapter, enhanced reliability model of wind turbine is shown and de-

scribed. New introduced and developed reliability model of wind turbine has the aim to give a detailed calculation of reliability indices of wind turbine. Output power of wind turbine with the following equation is presented:

$$P = \begin{cases} 0 & 0 \le v_t \le v_{ci} \\ (A + Bv_t + Cv_t^2)P_r & v_{ci} \le v_t \le v_r \\ P_r & v_r \le v_t \le v_{co} \\ 0 & v_t \ge v_{co} \end{cases} \tag{3}$$

Where is: P_r – rated power of wind turbine, v_{ci} – cut in wind speed, v_r – rated wind speed, v_{co} – cut out wind speed v_t – wind speed. Coefficient's A, B and C can be calculated by:

$$\begin{aligned} A &= \frac{1}{(v_{ci} - v_r)^2}\left[v_{ci}(v_{ci} + v_r) - 4(v_{ci} \cdot v_r)\left[\frac{v_{ci} + v_r}{2v_r}\right]^3\right] \\ B &= \frac{1}{(v_{ci} - v_r)^2}\left[4(v_{ci} + v_r)\left[\frac{v_{ci} + v_r}{2v_r}\right]^3 - (3v_{ci} + v_r)\right] \\ C &= \frac{1}{(v_{ci} - v_r)^2}\left[2 - 4\left[\frac{v_{ci} + v_r}{2v_r}\right]^3\right] \end{aligned} \tag{4}$$

Wind turbine model with series reliability model is presented. Failure of any component of wind turbine causes outage of wind turbine. Failure rate of whole wind turbine λ_{WT} can be calculated by:

$$\lambda_{WT} = \sum_{i=1}^{n} \lambda_i \tag{5}$$

Where is: λ_i – failure rate of i-th component of wind turbine. Since wind turbine is with series reliability model presented, then average downtime of whole wind turbine r_{WT} based on failure rates and downtimes of idividual components can be calculated by:

$$r_{WT} = \frac{\sum_{i=1}^{n} \lambda_i \cdot r_i}{\sum_{i=1}^{n} \lambda_i} = \frac{1}{\mu_{WT}} \tag{6}$$

Where is: r_i – downtime of i-th component of wind turbine, μ_{WT} – repair rate of whole wind turbine.

Once all the required input data have been entered and wind speeds and the corresponding output power of wind power plant have been listed, calculation of the reliability index of wind power plant have been based on the sequential Monte Carlo simulation (SMCs). As previously stated, for sequential Monte Carlo simulation exponential distribution is assumed. An example of sequential Monte Carlo simulation is presented in Figure 1. Duration of the period of correct work or *time to failure TTF* and duration of failures time or *time to repair TTR*, wind speed and power output of the wind power plant are monitored and presented in Figure 1. For sampling *TTF* and *TTR* the random number generated U_1 and U_2 must be transformed into time according to equations (7) and (8).

$$TTF = -\frac{1}{\lambda}\ln U_1 \tag{7}$$

$$TTR = -\frac{1}{\mu}\ln U_2 \tag{8}$$

At the moment of failure and after, wind speed is continually measured. Electric energy not supplied during downtime depends on wind speed during downtime. Two similar failures in duration, but significantly different by expected energy not served (*EENS*) are presented in Figure 1.

Sequential Monte Carlo simulation ends when convergence conditions have been reached. If default number of simulations is executed without the desired conditions of convergence achieved, it is necessary to increase the number of simulations N.

The convergence criterion for stopping the simulation is applied on the slowest variable convergence. In this case, the expected energy not served (*EENS*) is the slowest variable convergence, and then sufficient accuracy of simulation is achieved. Coefficient of accuracy α, for described case when the converges variable is expected energy not served (*EENS*), can be written by the following expression:

$$\alpha = \frac{\sigma(EENS)}{\overline{EENS}} \tag{9}$$

Where $\sigma(EENS)$ represents the standard deviation of expected energy not served and $\overline{EENS}$ represents average expected energy not served of all previous simulations. In this paper, as a condition of sufficient accuracy the value is set as: $\alpha = 0.05$. Block diagram of the enhanced reliability model of wind power plant is shown in Figure 2.

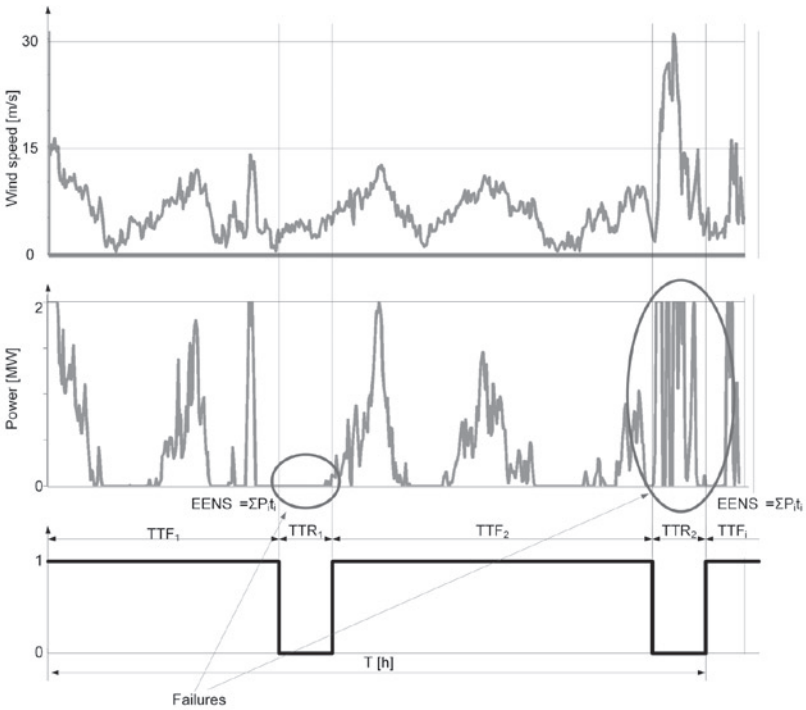


Fig. 1. An example of sequential Monte Carlo simulation

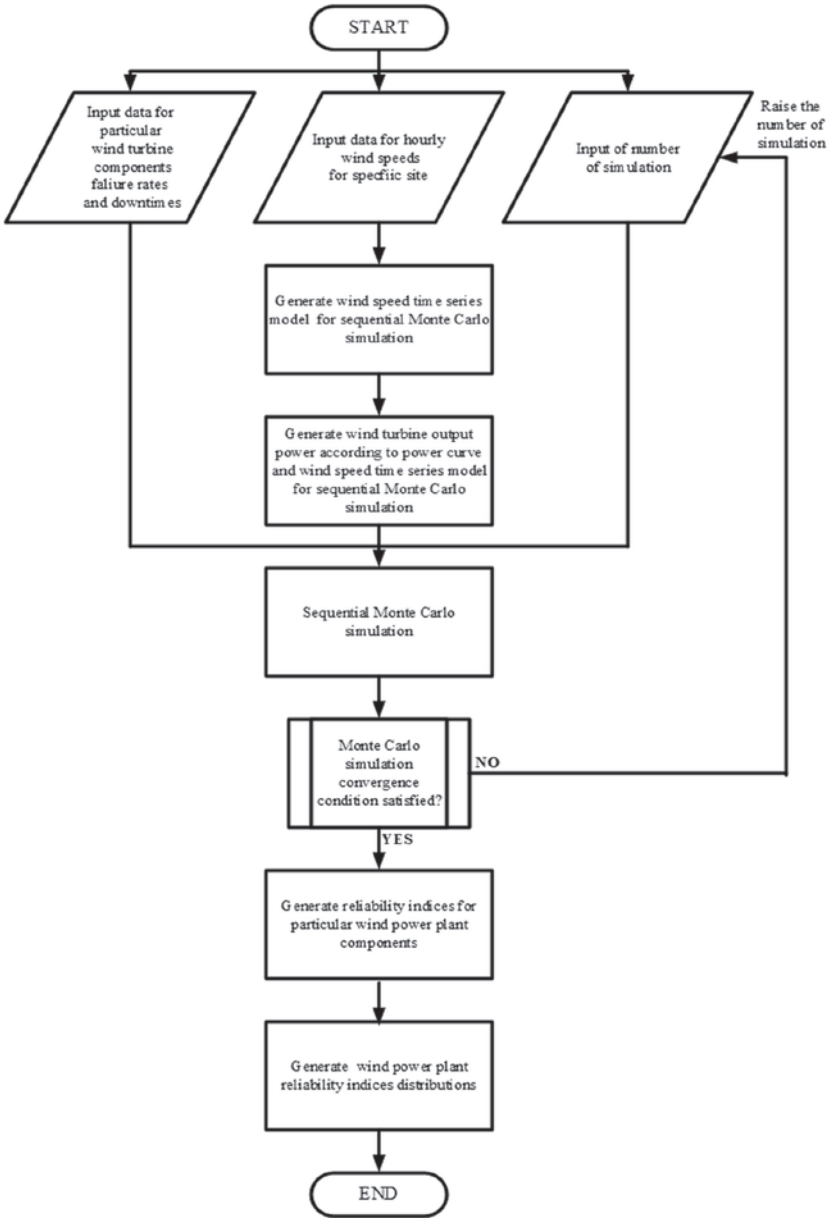


Fig. 2. Block diagram of wind turbine reliability model

5. Results

In this example, enhanced reliability model of the wind power plant is applied for the calculation of reliability indices of several different wind power plant configurations, based on concepts of type A, B and D. Data on the failure rates and downtimes of wind power plant configuration types A, B and D have already been presented in Table 1. The wind power plant which will be used as a case study of enhanced reliability model to get reliability indices has the following data: $P_r = 2$ MW, $v_{ci} = 5$ m/s, $v_r = 12$ m/s, $v_{co} = 25$ m/s.

To model wind speed "System identification Tool" after ARMA model ARMA (5,0) is obtained, further on as AR (5) abbreviated. The part of wind speed sequence is presented in Figure 3. Actual measured wind speeds are marked with a black line, while simulated values of wind speed using the AR (5) model are illustrated with blue lines. The resulting autoregressive parameters of AR (5) model are: $\Phi_1 = 0.9635$, $\Phi_2 = -0.06591$, $\Phi_3 = 0.03536$, $\Phi_4 = 0.0042$ and $\Phi_5 = 0.05521$. Having obtained the parameters of ARMA model, they are then used for enhanced reliability model of wind power plant. Wind speed model based on described procedure is then used in the Monte Carlo simulation.

The impact of individual components downtime of the wind power plant configuration types A, B and D on the *EENS* is shown in Figure 4.

From Figure 4 it is easy to see that gearbox has the largest contribution to the expected energy not served with the highest failure rate for type A. After the gearbox, electrical system, rotor and rotor blades failures have the greatest contributions to the *EENS* for type A. Downtimes of these three components (gearbox, electrical systems and rotor) caused about 48 % of all wind power plant failures, but they are responsible for about 75 % of the *EENS*. The remaining 60 % downtime of wind power plant is responsible for only about 25 % of the *EENS*.

From the same figure 4, it is obvious that electric system failures, gearbox failure, rotor and blades failures and hydraulic system failures have the largest contribution to the *EENS* for type B. Electric system has also the highest failure rate. Electronic control system has the second highest failure rate, but also has significantly smaller contribution to the total *EENS*. Failures of electric system and gearbox failures caused about 28 % of all downtime of the wind power plant. On the other hand, failure of above mentioned components are responsible for almost 60 % of the *EENS*. Remaining components failures make up about 72 % of downtimes, but they are responsible for only about 40 % of power plant's *EENS*, configuration type B.

It is very clear from Figure 4 that the electric system is the most unreliable component which caused the largest part of expected energy not served of wind power plant, configuration type D. The *EENS*, due to failures of electric system is 40.39 MWh/year. The rotor blades (pitch control) failures followed according to the partial amount of *EENS* is 10.93 MWh/year.

A comparison of the distribution of *EENS* of the wind power plants configuration types A, B and D is shown in Figure 5.

The comparison of absolute contributions of predominantly mechanical and electrical components fail-

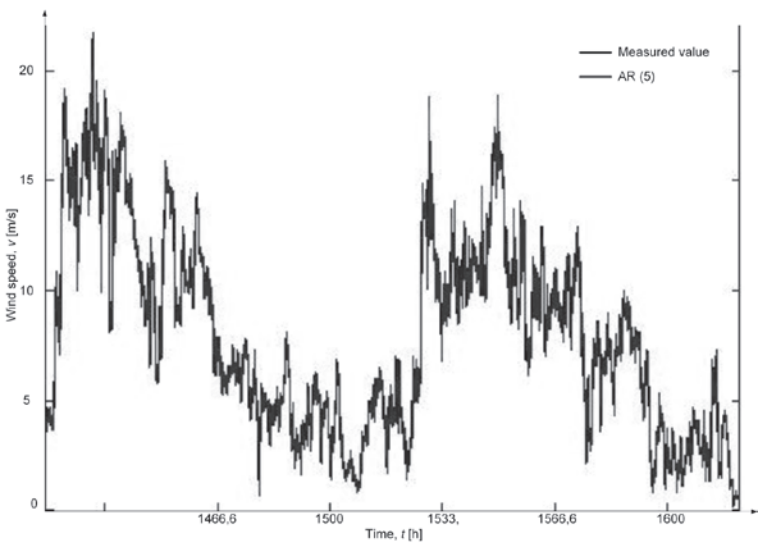


Fig. 3. An example of the simulated sequence of wind speed

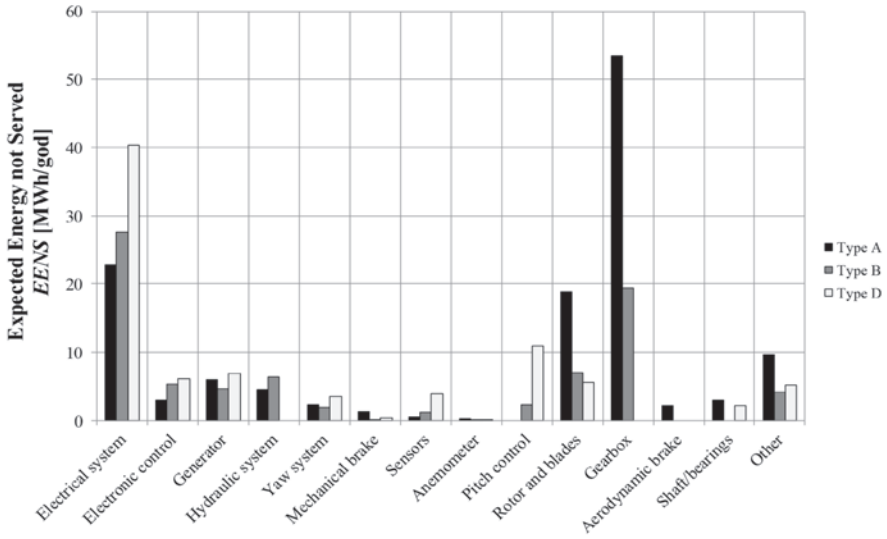


Fig. 4 Comparison of the EENS caused by failures of individual components of wind power plant for configuration concepts A, B and D

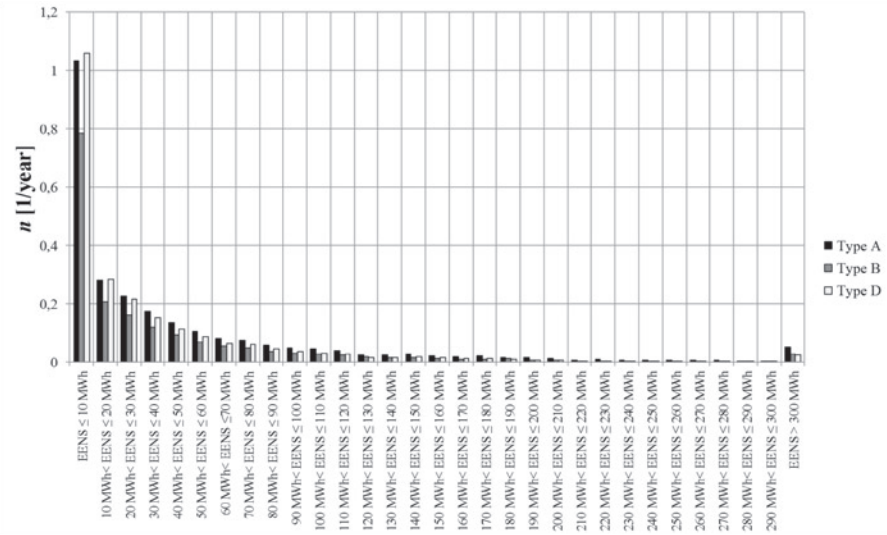


Fig. 5. A comparison of the distribution of EENS of the wind power plant configuration types A, B and D

ures of wind power plant, configuration types A, B and D, to *EENS* is presented in Figure 6.

The most significant contribution to the *EENS* of wind power plant, configuration type A is caused by mechanical components, as presented in Figure 6. For wind power plant of configuration type B, mechanical components failures and electric system components failures have fairly uniform contribution to *EENS*, while for wind power plant configuration type D, the most significant contribution to the *EENS* is caused mostly by failures of the electrical components. If absolute contributions of all components in all three analysed wind power plant configurations are compared (Figure 6), the largest contributions to the *EENS* are caused by: mechanical components of the wind power plant configuration type A, electrical components wind power plant configuration D, mechanical and electrical components of the wind power plant configuration type B, electrical components of the wind power plant configuration type A and mechanical components of wind power plant configuration type D, respectively. A comparison of the relative contribution to expected energy not served caused by mechanical and electrical components of the wind power plants of all analysed configuration types (A, B and D) is presented in Figure 7. Shares of mechanical and electric components failure rates in the overall intensity of downtime for wind power configuration types A, B and D are illustrated in Figure 8.

6. Conclusion

The paper presents the availability and expected energy not served calculation of different type of wind turbines based on the statistical data on the performance of wind turbine components. Recently installed wind turbines in Germany (after 2008) have, in general, three blades, pitch power regulation, variable speed and synchronous or double-fed induction generator. Those characteristics correspond to wind turbine configuration concepts C and D. It can therefore be expected that newly installed wind turbine in near future will have, in general, configuration concept C or D, while on the other hand, many wind turbines currently in operation have configurations concepts A or B. Therefore results of the performed availability analysis of all configuration concepts presented in this paper can be used for reliability modelling of both operation of the existing and planning and design of new wind turbines which is the focus of the future research.

Failures of predominantly mechanical components in wind power plant configuration of type A are responsible for about 67 % of the expected energy not served, in wind power plant configuration type B mechanical components are responsible for 48 %, and at least in wind power configuration type D for about 18 %. On the other hand, failures of mainly electrical components in wind power configuration of type A are responsi-

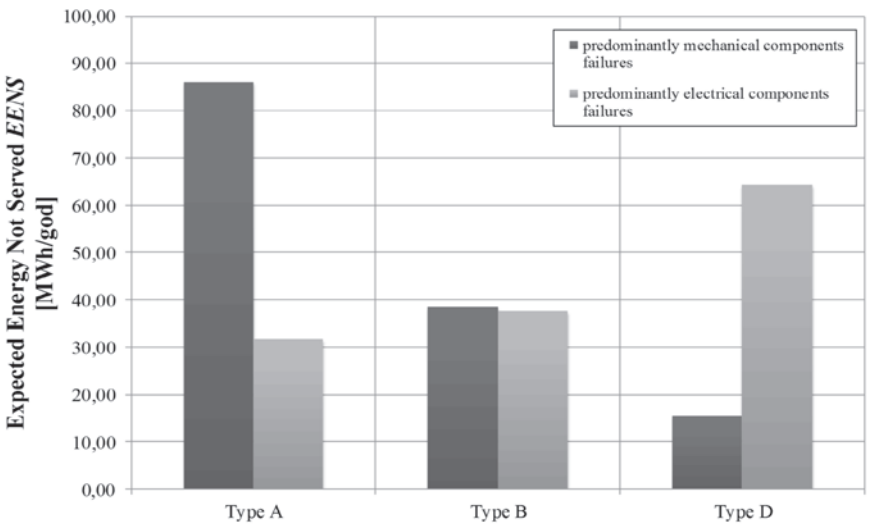


Fig. 6. The comparison of absolute contributions of predominantly mechanical and electrical components failures of wind power plant, configuration types A, B and D, to EENS

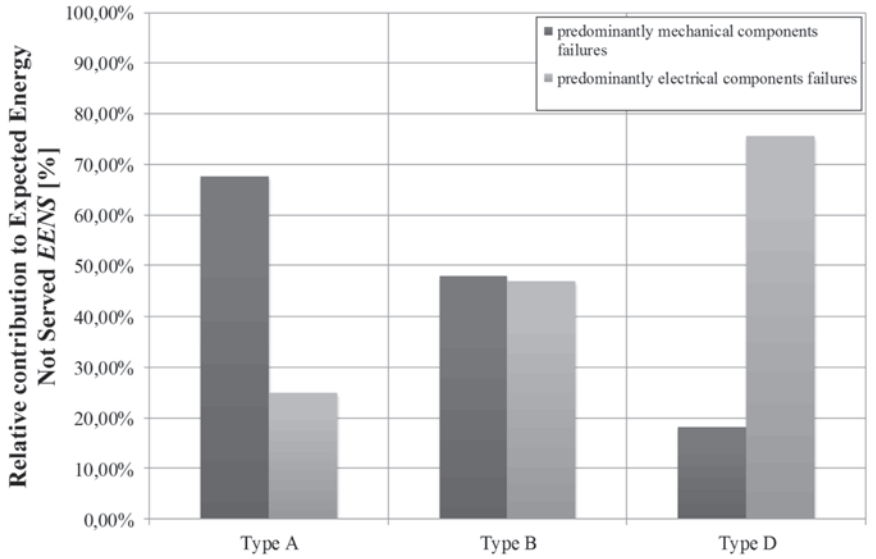


Fig. 7. Comparison of relative contribution to EENS, caused by mechanical and electrical components of the wind power plants of all analysed configuration types (A, B and D)

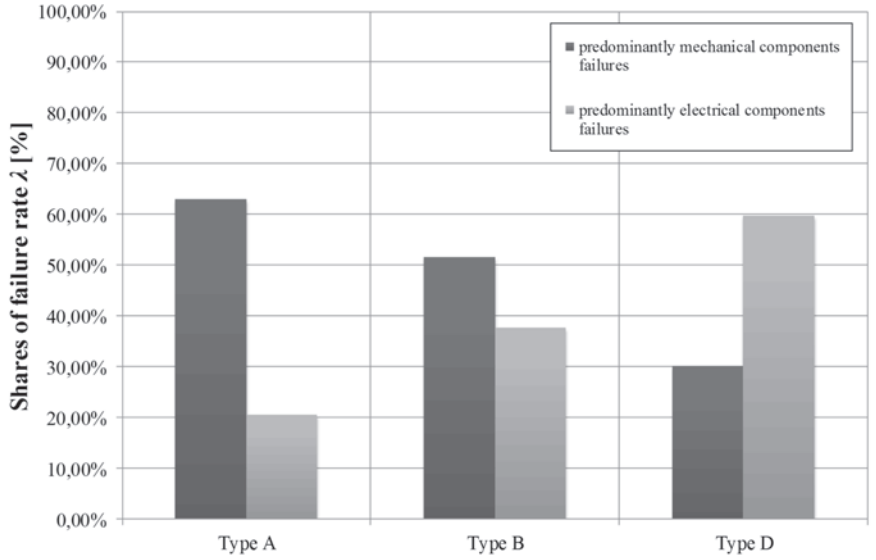


Fig. 8. Shares of failure rates of mechanical and electric components in the overall failure rates for wind power configuration types A, B and D

ble for about 25 % of the expected energy not served, in wind power plant configuration of type B electrical components are responsible for about 47 %, and in wind power plant configuration type D, they are responsible for as much as 75 %.

The main reason why failures of mechanical components in wind power plant configuration type D have a contribution of only about 18 % to the expected energy not served is that they have not transmission gearbox, which is previously described as a component with major contribution to the expected energy not served in wind power plant configuration of type A and also as a component with the second largest contribution to the *EENS* in wind power plant configuration type B (after the electrical system). Additional reason for mechanical components have such neglected cause of total failures in wind power plant configuration D is that the wind power plant configuration type D, as already stated, has no hydraulic but rather electric actuators (for example in rotor blades control position system). Finally, the wind power plant configuration type D has no hydraulic system and therefore there are no contributions of the component failures on *EENS*.

On the other hand, the reason why failures of mainly electrical components of the wind power plant configuration type D are responsible for about 75 % of the *EENS* is that the hydraulic actuators have been replaced by electric. In this case, wind power plant configuration type D have synchronous generators that have lower rotating speed, which means increasing number of pole pairs and increasing number of electric windings – according to [31], it might be one reason for the higher intensity downtime of generators and electrical systems. Another reason why the wind power plants configuration type D have significant contribution of failures of electrical components in overall expected *EENS*, is that they have converters with full frequency regulation, unlike other types of wind power converters.

Wind power plants configuration type B have a fairly uniform contribution to expected energy not served due to failures of mechanical components, as well as to failures of electric components. Unlike wind power plant configuration type D, the wind power plant configuration type B have a gearbox, as well as the hydraulic system, so these are reasons why failures of mechanical components have a greater contribution to the expected energy not served. On the other side, the wind power plants, configuration type B have no electric actuators in a system such as rotation of wind power plant or in the system for the rotation of the rotor blades. Wind power plant configuration type B has induction generator with slip rings, which is more robust than generators in wind power plant configuration type D, without converter and frequency controlling unit, and these are the reasons why it has smaller contribution to expected energy not served, due to failures of electrical components in relation to the wind power plants configuration type D.

Wind power plants configuration type A has the highest *EENS* caused by failures of mainly mechanical components. One reason is that wind power plants configuration Type A have gearbox similar as wind power plants configuration type B whose failures have quite significant contribution to *EENS*. Furthermore, it can be noted that the significant contribution to *EENS* have failures of rotor and rotor blades from all mechanical components. Wind power plants configuration type A opposed to the wind power plants configuration type B and D, have the power regulation by intentionally loosing of wind speed achieved by aerodynamic design of the rotor blades (stall control). In this case, there is no possibility for the regulation of the angle of the rotor blades, causing increase of mechanical forces and stresses of the rotor blades. This may be the reason for a greater contribution of mechanical components failure to the *EENS*. On the other hand, the reason why the wind power plant configuration type A has the smallest contribution of electric components failures to *EENS* (both in absolute and relative amount) may be that this type of configuration has a quite robust squirrel-cage induction generator and also there are no power electronic components (frequency converter).

References

1. Abul'Wafa A R. Reliability/cost evaluation of a wind power delivery system. Electric Power Systems Research 2011; 81: 873-879, http://dx.doi.org/10.1016/j.epsr.2010.11.028.
2. Ackermann, T. Wind Power in Power Systems. Wiley, 2005, http://dx.doi.org/10.1002/0470012684.
3. Arabian-Hoseynabadi H, Oraee H, Tavner P J. Wind turbine productivity considering electrical subassembly reliability. Renewable Energy 2010; 35: 190-197, http://dx.doi.org/10.1016/j.renene.2009.04.014.
4. Attwa Y M, El-Saadany E F. Wind Based Distributed Generation; Uncertainties and planning Obstacles. Proceedings of IEEE Power Engineering Society General Meeting, Tampa, USA, June 2007: 1–5, http://dx.doi.org/10.1109/pes.2007.385800.
5. Billinton R, Bagen. A Sequential Simulation Method for the Generating Capacity Adequacy Evaluation of Small Stand - Alone WECS. IEEE Proceedings of Canadian Conference on Electrical and Computer Engineering 2002; 1: 72-77.
6. Billinton R, Chen H, Ghajar R. Time-series models for reliability evaluation of power systems including wind energy. Microelectronics Reliability 1996; 36: 1253-1261, http://dx.doi.org/10.1016/0026-2714(95)00154-9.
7. Billinton R, Gao Y. Multistate Wind Energy Conversion System Models for Adequacy Assessment of Generating Systems Incorporating Wind Energy. IEEE Transactions on Energy Conversion 2008; 23: 163-170, http://dx.doi.org/10.1109/TEC.2006.882415.
8. Billinton R, Hua C, Ghajar R. A sequential Simulation Technique for Adequacy Evaluation of Generating Systems Including Wind Energy. IEEE Transactions on Energy Conversion 1996; 11: 728-734, http://dx.doi.org/10.1109/60.556371.
9. Castro Sayas F, Allan N R. Generation availability assessment of wind farms. IET Generation, Transmission and Distribution, 1996; 143; 507-518, http://dx.doi.org/10.1049/ip-gtd:19960488.
10. D'Annunzio C, Santoso S. Wind power generation reliability analysis and modelling. Proceedings of IEEE Engineering Society General Meeting, San Francisco, USA, 2005; 35-39.
11. Deshmukh R G, Ramakumar R. Reliability Analysis of combined wind-electric and conventional generation systems. Solar Energy 1982; 28; 345-352, http://dx.doi.org/10.1016/0038-092X(82)90309-7.
12. Ehsani A, Fotuhi M, Abbaspour A, Ranjbar A M. An Analytical Method for the Reliability Evaluation of Wind Energy Systems. Proceedings of IEEE Region 10 TENCON, Melbourne, Australia, November 2005; 1-7, http://dx.doi.org/10.1109/tencon.2005.300832.
13. Faulstich S, Lyding P, Tavner P J. Effects of Wind Speed on Wind turbine Availability. Proceedings of EWEA 2011, Brussels, Belgium, March 2011.
14. Giorsetto P, Utsurogi K F. Development of a new procedure for reliability modeling of wind turbine generators. IEEE Transactions on Power Apparatus and Systems 1983: 134-143, http://dx.doi.org/10.1109/TPAS.1983.318006.
15. Haghifam M R, Omidvar M. Wind Farm Modeling in Reliability Assessment of Power System. Proceedings of International Conference on Probabilistic Methods Applied to Power Systems, Stockholm, Sweden, June 2006: 1-5, http://dx.doi.org/10.1109/pmaps.2006.360414.
16. Hahn B, Durstewitz M, Rohrig K. Reliability of wind turbines - Experiences of 15 years with 1500 WTs. (Institut für Solare Energieversorgungstechnik (ISET), Verein an der Universität Kassel, Kassel), Germany, 2006
17. International Energy Agency. www.iea.org, accessed November 2012
18. Jaeseok C, Jeongje P, Kyeonghee C, Taegon O, Shahidehpour M. Probabilistic Reliability Evaluation of Composite Power Systems Including Wind Turbine Generators. Proceedings of International Conference on Probabilistic Methods Applied to Power Systems, Singapore, Singapore, June 2010: 802-807.
19. Kaigui X, Billinton R. Energy and reliability benefits of wind energy conversion systems. Renewable Energy 2011; 36: 1983–1988, http://dx.doi.org/10.1016/j.renene.2010.12.011.
20. Karaki S H, Chedid R B, Ramadan R. Probabilistic Performance Assessment of Wind Energy Conversion Systems. IEEE Transactions on Energy Conversion 1999; 14: 217-224, http://dx.doi.org/10.1109/60.766986.
21. Karki R, Billinton R. Cost-Effective Wind Energy Utilization of Reliable Power Supply. IEEE Transactions on Energy Conversion 2004; 19: 435-440, http://dx.doi.org/10.1109/TEC.2003.822293.
22. Karki R, Hu P, Billinton R. Reliability Evaluation of a Wind Power Delivery System Using an Approximate Wind Model. Proceedings of the 41st International Universities Power Engineering Conference, Newcastle, UK, September 2006: 113-117, http://dx.doi.org/10.1109/UPEC.2006.367726.
23. Karki R, Hu P. Wind Power Simulation Model for Reliability Evaluation. Proceedings of Canadian Conference on Electrical and Computer Engineering, Saskatoon, Canada, May 2005: 541-544, http://dx.doi.org/10.1109/ccece.2005.1556988.
24. Leite A P, Borges C L T, Falcao D M. Probabilistic Wind Farms Generation Model for Reliability Studies Applied to Brazilian Sites. IEEE Transactions on Power Systems 2006; 21: 1493-1501, http://dx.doi.org/10.1109/TPWRS.2006.881160.
25. Liang W, Jeongje P, Jaeseok C, El-Keib A A, Shahidehpour M, Billinton R. Probabilistic Reliability Evaluation of Power Systems Including Wind turbine Generators Using a Simplified Multi State Model: A case Study. Proceedings of IEEE Power & Energy Society General Meeting, Calgary, Canada, July 2009: 1-6.
26. Manco T, Testa A. A Markovian Approach to Model Power Avaliability of a Wind Turbine. Proceedings of IEEE Power Tech, Lausanne, Switzerland, July 2007: 1256-1261.
27. Peng W., Billinton R. Reliability Benefit Analysis of Adding WTG to a Distribution System. IEEE Transactions on Energy Conversion 2001; 16: 134-139, http://dx.doi.org/10.1109/60.921464.
28. Renewables 2013 – Global Status Report. http://www.ren21.net, accessed January 2014
29. Ribrant J, Bertling L. Survey of Failures in Wind Power Systems With Focus on Swedish Wind Power Plants During 1997 – 2005. IEEE Transactions on Energy Conversion 2007; 22: 167-173, http://dx.doi.org/10.1109/TEC.2006.889614.
30. Ribrant, J. Reliability performance and maintenance - A survey of failures in wind power systems, Stockholm: Master thesis, KTH School of Electrical Engineering, 2006.
31. Spinato F, Tavner P J, Van Bussel G J W, Koutoulakos E. Reliability of wind turbine subassemblies. IET Renewable Power Generation 2008; 3: 387-401, http://dx.doi.org/10.1049/iet-rpg.2008.0060.

32. Tavner P J, Greenwood D M, Whittle M W G, Gindele R, Faulstich S, Hahn B. Study of weather and location effects on wind turbine failure rates. Wind Energy 2012; 16: 175-187, http://dx.doi.org/10.1002/we.538.
33. Tavner P J, Xiang J, Spinato F. Reliability Analysis for Wind Turbines. Wind Energy 2007; 10: 1-18, http://dx.doi.org/10.1002/we.204.
34. Wang X, Dai Hui Z, Thomas R J. Reliability Modeling of Large Wind Farms and Associated Electric Utility Interface Systems. IEEE Transactions on Power Apparatus and Systems 1984: 569-575, http://dx.doi.org/10.1109/TPAS.1984.318746.
35. Windstats. www.windstats.com, accessed November 2010.

Danijel TOPIĆ
Faculty of Electrical Engineering
Josip Juraj Strossmayer University of Osijek
Ulica kneza Trpimira 2b, 31000 Osijek, Croatia

Damir ŠLJIVAC
Faculty of Electrical Engineering
Josip Juraj Strossmayer University of Osijek
Ulica kneza Trpimira 2b, 31000 Osijek, Croatia

Marinko STOJKOV
Mechanical Engineering Faculty
Josip Juraj Strossmayer University of Osijek
Trg Ivane Brlić Mažuranić 2, 35000 Slavonski Brod, Croatia

E-mails: danijel.topic@etfos.hr, damir.sljivac@etfos.hr, mstojkov@sfsb.hr

Article citation info:
QIAN X, WU Y. An electricity price-dependent control-limit policy for condition-based maintenance optimization for power generating unit. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 245–253, http://dx.doi.org/10.17531/ein.2016.2.12.

Xinbo QIAN
Yonggang WU

AN ELECTRICITY PRICE-DEPENDENT CONTROL-LIMIT POLICY FOR CONDITION-BASED MAINTENANCE OPTIMIZATION FOR POWER GENERATING UNIT

ZASTOSOWANIE STRATEGII UZALEŻNIAJĄCEJ TERMIN PRZEGLĄDU OD CENY PRĄDU ELEKTRYCZNEGO DO OPTYMALIZACJI UTRZYMANIA RUCHU AGREGATU PRĄDOTWÓRCZEGO Z UWZGLĘDNIENIEM JEGO STANU TECHNICZNEGO

For the control-limit policy of condition-based maintenance (CBM), it usually focuses on the internal condition of the equipment while neglecting the un-constant external conditions. However, the electricity price-dependent downtime cost have influence on the cost-effectiveness of control-limit policy for a generating unit in a power system. To make a linkage between CBM and the non-constant cost model, an electricity price-dependent control-limit policy (EPCLP) is proposed to accommodate the time-dependent downtime costs. For the proposed EPCLP, preventive maintenance control-limits is much flexible to be adjusted to different electricity price levels, and the maintenance cost reduction can be achieved among the planning horizon as a result. The optimal control-limits and maintenance costs for different downtime-cost ratios, reliabilities, covariate processes and electricity price scenarios are analysed to compare the performances between the proposed policy and the constant control-limit policy. Through the sensitivity analysis, the application scope of the proposed policy is evaluated.

Keywords: *Power generating unit, Multi-component system, Condition-based maintenance, Control-limit policy, Electricity price.*

Stosując strategie utrzymania ruchu uwzględniające bieżący stan techniczny obiektu (condition based maintenance, CBM) oparte na pojęciu progu konserwacji koniecznej (control limit), najczęściej przywiązuje się wagę do stanu samego sprzętu, ignorując przy tym niestałe warunki zewnętrzne. Należy jednak pamiętać, że w przypadku agregatów prądotwórczych wchodzących w skład układów elektroenergetycznych, koszty przestoju zależne od ceny energii elektrycznej mają wpływ na opłacalność stosowania strategii progu konserwacji koniecznej. Aby powiązać CBM z modelem kosztów niestałych, zaproponowano strategię progu konserwacji koniecznej, w której wysokość progu uzależniona jest od ceny prądu elektrycznego (electricity price-dependent control-limit policy, EPCLP). Przyjęcie takiej strategii pozwala uwzględnić koszty przestojów zależne od czasu. W EPCLP, progi czasowe konserwacji zapobiegawczej są bardzo elastyczne, co pozwala na ich regulację zgodnie z aktualną ceną energii elektrycznej. Strategia umożliwia redukcję kosztów w danym horyzoncie planowania. W celu porównania proponowanej strategii ze strategią stałego progu konserwacji koniecznej, w pracy przeanalizowano optymalne progi czasowe konserwacji koniecznej oraz koszty utrzymania ruchu dla różnych stosunków przestoju do kosztu, różnych wartości niezawodności, różnych procesów kowariantnych oraz różnych scenariuszy zmian cen energii elektrycznej. Zakres zastosowania proponowanej strategii oceniano za pomocą analizy czułości.

Słowa kluczowe: *agregat prądotwórczy, system wieloelementowy, utrzymanie zależne od bieżącego stanu technicznego, strategia progu konserwacji koniecznej, cena energii elektrycznej.*

1. Introduction

Preventive maintenance (PM) is important for the operation of a generating unit to avoid deteriorations and catastrophic failures. Generally, there are two types of preventive maintenance strategies, i.e. scheduled maintenance and condition-based maintenance (CBM) [1, 14, 17]. For scheduled maintenance (also called planned maintenance), the PM is carried out in accordance with established time intervals. With rapid development of instrumentation and measurement technology, more efficient maintenance types are needed, such as condition-based maintenance (CBM). For CBM, the maintenance action taken at each inspection is determined once the state of the system is higher than the specified control-limit [3, 9, 16]. In the class of control-limit policy, maintenance decision is made by comparison of degradation level to the critical thresholds, e.g. in [3]. Furthermore, functional failure and potential failure was defined by the failure threshold in [16]. The problem of what PM control-limit should be and how it should be obtained has been an increasingly attractive research topic, e.g. in [4, 5, 25].

The PM thresholds can be categorized into two main types corresponding to various system performance measures, such as a condition monitoring index and an integrated reliability index [14]. For the threshold based on a condition monitoring index [7, 24], the condition monitoring index can be obtained from monitoring items, such as wear, temperature, pressure, etc. For example in [24] the PM threshold was based on the wear measurement, while a kind of control-limit based on laser's operating current was studied in [7]. Another type of threshold is set on the integrated reliability index, and it is derived from both event data and condition monitoring data [9, 14]. When the reliability of the assets is influenced and/or indicated by different risk factors, which are so-called covariates, the condition monitoring data can be extended to both environment covariates and condition monitoring covariates [9]. Most of the covariate models are developed based on the Proportional Hazards Model (PHM) [6] and more extended models can be referred to [9, 14]. For example, a kind of

PM threshold of control-limit policy (CLP) was proposed based on the PHM in [2]. A kind of preventive replacement threshold was proposed based on the deterioration level in presence of environmental covariates in [26].

Alternatively, dynamic control-limit policy, as a new kind of control-limit, has been applied to construct dynamic thresholds. It has been shown to have a cost-saving and a better generalization capability than conventional constant control-limit policy. The studies about dynamic control-limit policy can be classified mainly into three categories, such as inspection rate-dependent threshold [5], age-dependent threshold [4, 8, 15] and degradation-dependent threshold [26]. For an inspection rate-dependent threshold [5], the higher the inspection rate, the higher the degradation-type PM threshold. It makes sense that more frequent inspections lead to more timely warning if system is near to failure. For an age-dependent threshold, PM threshold decreases in age and should be triggered by smaller signal values when it deteriorates [4]. The PM threshold can also be non-decreasing in age and it can be more tolerant of larger signal values for older systems, if there is increasing accuracy in predicting the future signal value [8]. For a degradation-dependent threshold, Zhao et al. [26] proposed an adaptive maintenance decision to take into account the state of covariates to dynamically adapt the PM threshold. However, much further improvements of dynamic control-limit policy are needed with respect to both internal condition (e.g. degradations) and un-constant external condition (e.g. non-constant cost), especially for the application of a generating unit.

In a power system, to raise the market competitiveness for a generating unit, not only the reliability should be improved, but the economical performance should be improved as well. Although most of the CBM decisions are made based on degradation condition, it is worth noting that the time-dependent downtime cost can occur for each maintenance and replacement action. During the maintenance duration, the downtime cost is an important part of the total maintenance cost. The downtime cost is fluctuant according to the time-dependent electricity price, since electricity price is the main influence factor for time-varying downtime cost. As a result, the problem needs to be addressed that how to further reduce the maintenance cost in terms of time-varying downtime cost for a generating unit.

Although control-limit policy is widely applied to condition-based maintenance, existing control-limit policies do not examine the non-constant cost in CBM optimization. In terms of the CBM optimization based on cost criteria for a generating unit, the time-dependent electricity price can not be ignored since it can cause the fluctuation of downtime cost. However, the research about this problem is limited [2], so extended CLP considering non-constant cost is in needed of research. Inspired by the control-limit policy for CBM, we decide to make a linkage between non-constant cost and the control-limit policy for CBM. As a result, we extend the constant control-limit policy to an electricity price-dependent control-limit policy to deal with the electricity price-dependent downtime cost in CBM. In this paper, we consider the influence of both the degradation and the electricity price on the control-limit policy. Compared to CLP, the proposed EPCLP can take advantage of time-dependent electricity price in order to perform PMs economically to achieve the minimal maintenance cost, by assigning different thresholds to different electricity price levels. The proposed EPCLP can be much more flexible for a generating unit, in terms of both reliability and cost-effectiveness.

The rest of this paper is structured as follows: section 2 describes the proposed electricity price-dependent control-limit policy; in section 3, an extensive computational analysis is conducted for evaluation of the proposed EPCLP in a comparison to the constant CLP; concluding remarks are given in section 4.

2. Electricity price-dependent control-limit policy

After symbols and abbreviations are listed in section 2.1, section 2 has three primary stages. In section 2.2, the conventional control-limit policy is described before the proposed policy. Next, in section 2.3, the structure of the proposed electricity price-dependent control-limit policy is illustrated, and the innovations and characteristics are proposed. And then, in section 2.4, the modelling and calculation procedures are described for CBM optimization in terms of the proposed policy.

2.1. Symbols and abbreviations

Model parameter

h_i	hazard rate of component i
β_i , η_i	Weibull shape parameters and scale parameters for component i
a_{it}	age of component i at time t
z_{it}	covariate value of component i at time t
γ_i	corresponding coefficient of the covariate for component i
ep_t	electricity price at time t
K_i	difference between the cost of per CM and the cost of per PM for component i
C_{dL}, C_{dH}, C_{dM}	downtime cost during above-average, average and below-average electricity price periods respectively
F_{it}	probability of sudden failure at time t for component i
Δt	inspection interval
$c_{\mathrm{c}i}$, $c_{\mathrm{p}i}$, $c_{\mathrm{o}i}$	cost for each CM, PM and OM, respectively
$c_{\mathrm{d}t}$	downtime cost at time period t
x	simulation scenario
W_x	cost rate for simulation scenario x
Ns	number of degradation simulations
$E(C)$	expected cost rate
λ	downtime cost ratio, i.e., DCR
T_{cL}, T_{cM}, T_{cH}	number of continuous periods for below-average price, average price and above-average price, respectively
T_{sL}, T_{sM}, T_{sH}	sum of the periods for below-average price, average price and above-average price, respectively

Model variable

$d_1(ep_t)$	level-1 thresholds dependent on the electricity price at time t
$d_{1\mathrm{H}}$, $d_{1\mathrm{M}}$, $d_{1\mathrm{L}}$	level-1 thresholds for above-average, average and below-average electricity price, respectively
d_2	level-2 threshold

Abbreviations

PM	Preventive Maintenance
CBM	Condition-based Maintenance

CM	Corrective Maintenance
OM	Opportunistic Maintenance
PHM	Proportional Hazards Model
SARIMA	Seasonal Autoregressive Integrated Moving Average
CLP	Control-limit policy
EPCLP	Electricity price-dependent control-limit policy

2.2. Control-limit policy

A valuable statistical procedure for estimating the risk of equipment failure is the proportional hazard model (PHM) if it is subjected to condition monitoring [6]. A form of PHM combines a baseline hazard function h_0 along with a factor that takes into account covariates to improve the prediction of failure. In this paper, a Weibull PHM [14] is applied and it is calculated by Eq. (1). It is a joint model of PHM and Markov property for covariate evolution.

$$h_{it} = h_0 \exp(\gamma_i z_{it}) = \beta_i (a_{it}/\eta_i)^{\beta_i - 1} \exp(\gamma_i z_{it}) / \eta_i \tag{1}$$

where β_i, η_i are Weibull shape parameters and scale parameters, respectively, for component i. a_{it} is the age of component i at time t, z_{it} is the covariate value of component i at time t, and γ_i is the corresponding coefficient of the covariate.

A possible realization of the time evaluation of a component is illustrated in Fig. 1. For each component of a power generating unit, both sudden failure and degradation failure can occur [9]. Therefore, at each inspection time, there can be four possible events for each component, including CM, PM, OM or no maintenance actions. If the hazard h_{it} exceeds the predetermined control-limit levels, PM will be performed. If the multi-component system continues on operation without any maintenance action, covariate state of each component i will follow the Markov process. Based on the state transition probability matrix we can obtain the sample covariates z_{k+1} at next inspection based on the current value of covariates z_k. Covariates of each component can be selected by the EXAKT software [2]. For each component, the covariate state can be modelled by the Markov process [13, 18, 19]. The accumulated values of ppm metals behaved as a homogenous Markov process in [13], and a continuous time homogenous Markov process was modelled for the degradation state in [19]. The control-limit replacement policies for deteriorating systems were established by PHM which was dependent on its age and the condition monitoring state [18].

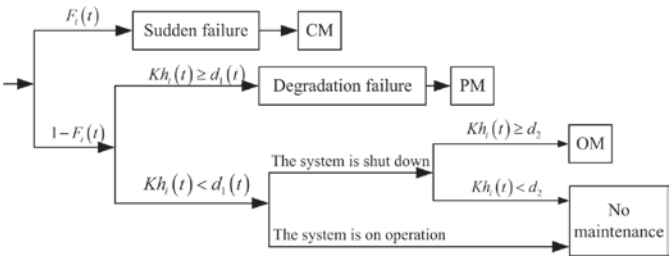


Fig. 1. Tree of possible events for component i at time

2.3. Structure of an electricity price-dependent control-limit policy

To deal with the time-dependent downtime cost due to the fluctuating electricity price, we extend the PHM based CBM policy for multi-component system to an electricity price-dependent control-limit policy (EPCLP) from a constant control-limit policy (CLP). Instead of the constant control-limit, the threshold for the EPCLP is adjusted to different electricity price levels (e.g. above-average, average and below-average). The level-1 thresholds d_1 are d_{1H}, d_{1M} and d_{1L} for above-average, average and below-average electricity price, respectively, such that:

$$\mathbf{d}_1(ep_t) = \begin{cases} d_{1H} & \text{if } ep_t \in \left(ep^{m} + \Delta ep, ep^{U}\right] \\ d_{1M} & \text{if } ep_t \in \left[ep^{m} - \Delta ep, ep^{m} + \Delta ep\right], \\ d_{1L} & \text{if } ep_t \in \left[ep^{L}, ep^{m} - \Delta ep\right) \end{cases} \tag{2}$$

where ep^U, ep^L are the upper and lower bound of electricity price, respectively. Meanwhile, $ep^m+\Delta ep$, $ep^m-\Delta ep$ are the upper and lower bound of average level for electricity price, respectively.

The proposed EPCLP for CBM for a hydro generating unit as a multi-component system is proposed as follows:

1) For component i, perform CM if a sudden failure occurs with the probability F_{it}.
2) For component i, perform PM if $K_i h_{it} \geq d_1(ep_t)$;
3) If the system is shut down for the maintenance (e.g. PM or CM), perform OM on the component l if $K_l h_{lt} > d_2$.

A schematic representation of the EPCLP is presented in Fig. 2 to illustrate the cost-effectiveness of the proposed policy, compared to the constant CLP. For simplicity, the example describes a single-component system instead of a multi-component system. As the hazard rate of the component increases from the initial time or after each maintenance, it can meet $d_L < d$ (respectively, $d_H > d$) during the low (respectively, high) price periods. Here, d_L, d_H, d_M are threshold for EPCLP, while d is the constant threshold. As a result, PM can be performed during low price periods with much higher probability by EPCLP compared to the constant CLP. So the downtime cost is in total ($C_{dH}+C_{dH}+C_{dM}$) for the constant CLP, while the downtime cost is in total ($C_{dL}+C_{dH}+C_{dM}$) for EPCLP. The proposed policy can achieve a cost-reduction of (C_{dH}-C_{dL}) compared to the constant CLP in this example, since EPCLP is more flexible than CLP. From the illustration, EPCLP can be cost-efficient by balancing the timing of the outages, as well as by balancing the PMs and CMs.

However, for the trade-off among the costs of PM, CM and outages to achieve a minimal cost for the EPCLP, PM cannot necessarily be performed during below-average price periods with high probability for all cases. The problem is complicated since optimal thresholds are

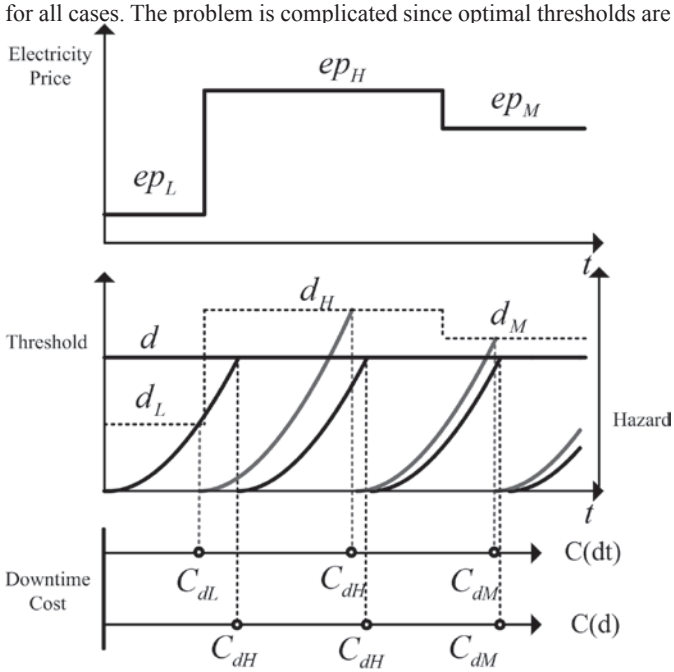


Fig. 2 Schematic representation of the epclp compared to clp

sensitive to both the reliabilities of the multi-component system and the relationships between the various costs, as well as the degradation process and electricity price scenarios.

Qualitatively, for cases with high reliability levels and less number of above-average price periods, it can be cost-efficient to perform PM during above-average price periods, since the PMs and outages can be reduced effectively. While for cases with low reliabilities and a large number of above-average price periods, it would be cost-efficient to perform PM during high price periods.

When the downtime cost is higher, it could be cost-effective to perform fewer PMs, more CMs and fewer outages. So the threshold can be lower for the below-average electricity price level with fewer continuous periods. Conversely, if the downtime cost is lower, it could be cost-efficient to perform more PMs with fewer CMs. So the threshold of the electricity price level which has more continuous periods may be much lower. The advantage of the proposed policy in terms of the cost-efficiency can be verified by the results of the simulations in section 3.3.

However, for cases with much higher reliabilities, or with much lower downtime costs, the advantage of the proposed policy is not so significant. For cases with much higher reliabilities, few PMs are needed, so the cost-saving of EPCLP can be insignificant, the same goes for the cases with much lower downtime costs. Sensitivity analysis will be proposed in Section 3.3.

2.4. Modelling of an electricity price-dependent control-limit policy for CBM

Since the failure cost is generally much higher than that of performing PM, sufficient PMs can reduce failures but cost will be increased as well. So a trade-off exists between the scheduling of PMs and CMs. The objective of the problem is to obtain the optimal PM/OM thresholds to minimize the average cost of maintenance and loss due to downtime during the planning horizon.

Based on these explanations, the EPCLP can be stated as follows: given the fluctuating electricity price, with having the stochastic degradation process and the relations specifying the permissible of maintenance actions, the problem is to the assign the dynamic thresholds to PMs and OMs, and the maintenance cost rate is minimized. The assumptions used in this problem are as follows:

1) The component degradation process can be described by the proportional hazard model, with covariates observable at each inspection.
2) The components of the system are independent in degradations and failure processes.
3) The components are economically dependent. The loss of productivity during downtime is incurred if the generating unit is shut down for CM or PM. The loss of productivity is incurred only once for each outage.
4) We focus on the maintenance optimization in this study, so the inspection interval is not a design variable in the optimization problem. The inspections are assumed to be scheduled according to fixed intervals, so we can assume they are performed at zero cost.
5) Each component may suffer random failure only once during each inspection interval. A failure may happen during the inspection interval or just at the inspection time. For simplification, we assume failures just occur at the inspection time, and so do the maintenance actions.
6) During downtime the components do not deteriorate.
7) Both CM and PM can fully recover the component to an as-good-as-new condition.

The mathematical formulation for the problem is given as follows:

$$EC = \min_{\mathbf{d}_1(ep_t),d_2} \sum_{x=1}^{Ns} p_x W_x\left(\mathbf{d}_1(ep_t),d_2\right) \tag{3}$$

$$W_x\left(\mathbf{d}_1(ep_t),d_2\right) = \frac{1}{NT\cdot\Delta t}\sum_{t=1}\left(\sum_{i=1}^{NC}\left(I_{cit}^x c_{ci} + I_{pit}^x c_{pi} + I_{oit}^x c_{oi}\right) + I_{dt}^x c_{dt}\right) \quad \forall x \tag{4}$$

$$\text{s.t.} \quad I_{cit}^x = \begin{cases} 1 & \text{when } u_{it}^x < F_{it} \\ 0 & \text{otherwise} \end{cases} \quad \forall i,t,x \tag{5}$$

$$I_{pit}^x = \begin{cases} 1 & \text{when } I_{cit}^x = 0 \;\&\; Kh_{it} \ge \mathbf{d}_1(ep_t) \\ 0 & \text{otherwise} \end{cases} \quad \forall i,t,x \tag{6}$$

$$I_{oit}^x = \begin{cases} 1 & \text{when } \prod_i^{NC}\left(1-I_{cit}^x\right)\left(1-I_{pit}^x\right) = 0 \;\&\; \mathbf{d}_1(ep_t) > Kh_{it} \ge d_2 \\ 0 & \text{otherwise} \end{cases} \quad \forall i,t,x \tag{7}$$

$$\min \mathrm{d}_1(ep_t) > d_2 \tag{8}$$

$$F_{it} = 1 - \exp\left\{-h_{it}\Delta t\right\} \tag{9}$$

$$u_{it}^x \sim U(0,1) \tag{10}$$

Eq. (3) is the objective function which indicates the purpose of identifying electricity price-dependent threshold sets that achieves the minimum expected maintenance cost over all possible degradation scenarios. Eqs. (4) state the maintenance cost rate for each degradation scenario x over the whole planning horizon NT, including maintenance costs and downtime costs. Constraints (5) state a sudden failure will occur with the probability F_{it}, and it can be determined by the stochastic variable u_{it}. Constraints (5) ~ (7) determine whether CM, PM and OM will occur respectively for component i during period t for degradation scenario x. Constraints (8) guarantee the first-level thresholds be greater than the second-level threshold. Constraints (9) determine the sudden failure probabilities [23]. Constraints (10) impose the normal distribution on the stochastic variable u_{it}.

Furthermore, to determine the value of PM threshold, the optimization of thresholds can be achieved by certain criteria such as cost, reliability and availability criteria. For example, an optimal PM threshold was derived by minimizing the average maintenance cost in [2, 18]. A double-level PM threshold for a multi-component system was optimized on cost criteria as well in [23].

The proposed maintenance policy is evaluated by the average cost in the planning horizon, taking into account the cost of each type of maintenance action and downtime costs. In order to have sensible results, the maintenance cost evaluation is implemented using Monte Carlo simulation [22]. The simulation generates a large number of degradation scenarios. Each component undergoes stochastic transitions between the possible covariate states, evolving through conditions of availability and unavailability due to PM or sudden failure. For each simulation scenario x, the cost rate, including total maintenance cost and loss of productivity, will be calculated, denoted by W_x if the thresholds, $\boldsymbol{d}_1(ep_t)$ and d_2, are given. The probability for each degradation scenario is p_x. After N_s simulations, the expected cost rate EC under given thresholds, $\boldsymbol{d}_1$ and d_2, can be estimated. The simulation procedures will continue until it satisfy the convergence criterion [20], and the number of valid simulations can be determined. The overall Monte Carlo simulation for EPCLP for a power generating unit is outlined in Fig. 3.

```
Algorithm: Monte-Carlo simulation for EPCLP for a power generation unit
Inputs: degradation process; price scenarios; PHM parameter.
Outputs: optimal threshold set for EPCLP.
Begin:
    while (( d1(epl),d2 )∈ valid sets) do
        n←1; t←1; i←1;                    //initialization
        while ( |ECn- ECn-1|> tolerance limit ) do
            while ( t< NT ) do
                while (i ≤ NC) do
                  generate u ~ U(0,1)          // generation of sudden failure
                  if u ≤ F( i, t)
                      then fcm (i, t)=0 and Ccm← Ccm + ccm(i) otherwise fcm (i, t)=1
                end while
                i←1;
                while (i ≤ NC) do              // judgement of PM
                  if fcm (i, t)=0 and Kh(i, t)≥d1(ep(t))
                      then fpm (i, t)=0 and Cpm← Cpm + cpm(i) otherwise fpm (i, t)=1
                end while
                i←1;
                while (i ≤ NC) do              // judgement of OM
                    if prod( fcm ) * prod( fpm )=0 and fpm (i, t)=1 and Kh(i, t)≥ d2
                        then fom (i, t)=0 and Com← Com + com(i) otherwise fom (i, t)=1
                 end while
                 if prod( fcm ) * prod( fpm )=0 then Cd← Cd +cd(ep(t))
            end while
            recording the average cost for each simulation ECn
            n=n+1;
        end while
    end while
    report the optimal threshold set
end
```

Fig. 3. Monte carlo simulation method for EPCLP for a power generating unit

4. Computational experiments

In summary, the proposed electricity price-dependent control-limit policy (EPCLP) is an enhancement of constant control-limit policy (CLP). The maintenance actions using EPCLP and CLP and the comparison results of these two kinds of policy are discussed in Section 3.2. In order to achieve better insight into the performance of the EPCLP for CBM in Section 3.3, EPCLP is compared with the CLP in terms of the following four factors, downtime cost ratios (Section 3.3-1), reliabilities of the multi-component system (Section 3.3-2), covariate processes (Section 3.3-3), electricity price scenarios (Section 3.3-4).

3.1. Dataset

We present a simulation study to illustrate the effectiveness of the proposed policy for a generating unit as a multi-component system. The capacity of the generating unit is set to be 25MW. PHM parameters, including a shape parameter, a scale parameter and a covariate parameter for each component, are listed in Table 1. The parameter γ is the covariate parameter which indicates the degree of influence a covariate has on the hazard. The costs of various maintenance actions for each component are listed in Table 1, including costs for CM, PM and OM, and the unit of the each maintenance cost is \$1000. The possible values for z_{it} are Z_l, which represent selected covariate bands. In this example, the values are set to be Z_0=0, Z_1=35, Z_2=60 and Z_3=85 which represent the condition monitoring states. The state transition matrix of each component, listed in Table 2(a-c), is derived from a month interval, and the covariates of each component are assumed to be independent.

Table 1. Parameters of proportional hazards model and CM, PM and OM cost for critical components

component	Shape Parameter β	Scale Parameter η	Covariate Parameter γ	CM cost (\$1000)	PM cost (\$1000)	OM cost (\$1000)
hydro turbine	3	1000	0.060	213	24	12
generator	2	1500	0.044	150	20	10
transformer	3	800	0.053	210	24	12

Table 2(a) Transition probability matrix of condition indictor of hydro turbine

Bands	[0,35)	[35,60)	[60,85)	[85,100]
[0,35)	0.72350	0.25340	0.02258	0.00052
[35,60)	0.03301	0.85120	0.11490	0.00089
[60,85)	0.01800	0.19220	0.78710	0.00270
[85,100]	0	0	0	1

Table 2(b) Transition probability matrix of condition indictor of generator

Bands	[0,35)	[35,60)	[60,85)	[85,100]
[0,35)	0.79850	0.18180	0.01921	0.00049
[35,60)	0.02815	0.83270	0.13840	0.00075
[60,85)	0.02110	0.12250	0.74320	0.11320
[85,100]	0	0	0	1

Table 2(c) Transition probability matrix of condition indictor of transformer

Bands	[0,35)	[35,60)	[60,85)	[85,100]
[0,35)	0.73590	0.23310	0.03038	0.00062
[35,60)	0.00926	0.82920	0.16070	0.00084
[60,85)	0.00794	0.09850	0.81030	0.08326
[85,100]	0	0	0	1

The electricity price history profiles are from the Monthly Locational Marginal Pricing of the PJM Power Market [21]. For the identification of a statistical model, we use the weekly electricity price profile of twelve years from 1999 to 2012. A SARIMA model is applied to model the time series for the electricity price $ep_t, t\in\mathbb{Z}$ [11]. SPSS is used to identify a SARIMA(2,1,2)×(1,1,0)$_{52}$ model for the electricity price ep_t during period t. Introducing $Y_t := ep_t - ep_{t-1} - (ep_{t-52} - ep_{t-53})$, it reads:

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_1^* \left(y_{t-52} - \phi_1 y_{t-53} - \phi_2 y_{t-54}\right) = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \theta_0$$

The estimated model coefficients are:

$$\left(\phi_1, \phi_2, \phi_1^*\right) = \left(-0.604, 0.346, -0.47\right), \left(\theta_1, \theta_2, \theta_0\right) = \left(-0.175, 0.8, -0.37\right)$$

Furthermore, ε_t, $t\in\mathbb{Z}$ are independent, normally distributed random variables with mean 0 and variance 11.26.

The stochastic electricity price process is approximated by a scenario tree [10]. According to the SARIMA equation, a large number of simulated price scenarios are generated using an i.i.d. realization of t. Then the empirical means of the forecasting electricity price can be derived from simulation price scenarios. In this study, the planning horizon is set to be three years. The monthly electricity price of the second and third years are assumed to be identical with the price process of first planning year, since historical electricity price records are not enough for predicting electricity price process accurately over the next two years.

The ranges of the average price, above-average price and below-average price are defined to be $[ep^m - \Delta ep, ep^m + \Delta ep]$, $[ep^m + \Delta ep, ep^U]$, and $[ep^L,$

$ep^m - \Delta ep$], respectively. ep^m represents the mean value of the expected foresting electricity price and is \$52/MWh in this case. The parameter Δep is set to be \$5/MWh to divide the forecasting electricity price into three different levels.

The thresholds $[d_1, d_2]=[d_{1L}, d_{1M}, d_{1H}, d_2]$ are defined within the range (\$1000/day) while the discretization accuracy ($\log_{10}$)=0.5 for each component. For simplification the unit of the threshold (\$1000/day) will not be mentioned, and the values of the thresholds are all in logarithm scale (base 10) for the rest of the paper.

3.2. Comparison of EPCLP and CLP for the Mean Forecasting Electricity Price

To confirm the performance of the proposed policy, we take the mean forecasting electricity price scenario for an example. The results (Table 3) show that the electricity price-dependent threshold $d_1=[d_{1L}, d_{1M}, d_{1H}]$ is [-1, -1, 0]. The thresholds of below-average and average price periods are much lower than that of above-average price periods. The minimum of the price-dependent thresholds, min(d_1)=-1, is much lower than the constant threshold d_1=-0.5. Meanwhile, the maximum of the price-dependent thresholds max(d_1)=0 is greater than the constant threshold. Compared to the constant CLP, more PMs can be performed by the proposed policy to reduce failures. Specifically, the average number of PMs is increased to 14.0 from 8.0 by the proposed policy, the average number of CMs is reduced to 1.3 from 2.2, and the average number of outages is increased to 11.7 from 9.8. As a result, the average cost is reduced by (960-890)/960=7%. The standard error of the estimated cost is around \$5/day and the standard error of the average number of the event (outage, CM, PM or OM) is 0.1.

The optimal thresholds of the proposed policy make sense in this case. Generally, the total cost for each PM (including PM cost and downtime cost) is much lower than the total cost for each failure (including CM cost and downtime cost). Therefore, it is cost-effective to reduce failures by performing more PMs compared to the CLP.

Furthermore, the comparison of the event probability distributions between the EPCLP and CLP is shown in Fig. 4. The vertical axis indicates the probability of the event (e.g., outage, CM, PM or OM) occurred at time t. The event probability can be approximately estimated by the percentage of times the event occurs at time t over all the total simulations. For example, if the failure event at time t occurs 100 times over a total of 3000 simulations, the probability of the failure is estimated to be 100/3000=3.3%. Compared to the constant CLP, PM can be performed during below-average and average price periods with much higher probabilities by the proposed policy as shown in Fig. 4. Meanwhile, the probabilities of failure (or CM) by the EPCLP are generally lower than that of the constant CLP during the planning horizon. So this case shows that the number of failures can be reduced significantly by the proposed policy.

Table 3. Results for the case with downtime cost ratio 0.12 for EPCLP compared to constant CLP

policy	d_1	d_2	Outages	CM	PM	OM	cost (\$/day)	cost-saving
EPCLP	[-1,-1,0]	-1.1	11.7	1.3	14.0	3.2	890	7%
CLP	-0.5	-1	9.8	2.2	8.0	5.6	960	

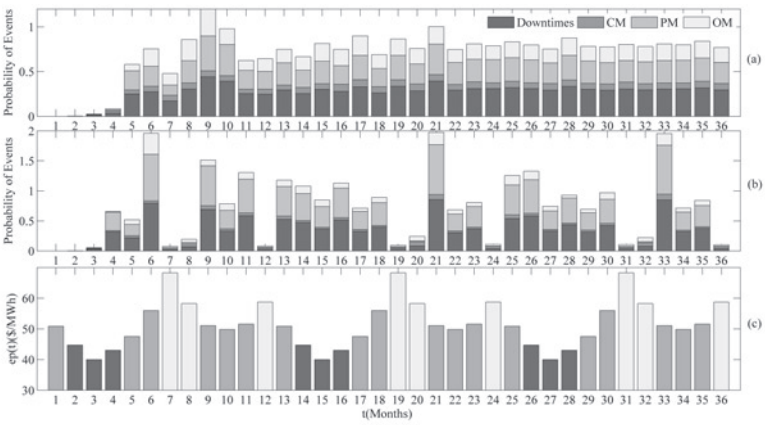


Fig. 4. Comparison of the event probability distributions between the and CLP a) and EPCLP b) and CIS the forecasting electricity price

3.3. Study Results and Discussions for Sensitivity Analysis for EPCLP

1) Influence of downtime cost ratios (DCR)

To analyze the effect of different DCR to optimal EPCLP, DCR is defined by $\lambda=c_d/(c_d+c_p+c_c)$. Here, c_p is the average of c_{pi}, c_c is the average of c_{ci} and c_d is the average of c_{dt}. DCR increases with higher downtime cost while c_p and c_c are fixed. For the mean price scenario, the proposed policy can offer a 3% to 7% cost-saving over the CLP (Table 4) within the valid range of DCR $\lambda \leq 0.35$. As λ increases from 0.12 to 0.35, the number of CMs increases with the number of PMs. It can be cost-effective to reduce the outages by increasing CM for higher DCR. Therefore, CM is increasing, and number of PMs and outages are decreasing while λ is increasing. For higher DCR (e.g. λ>0.35), EPCLP is no longer cost-effective since less PM is needed.

The valid DCR can be categorized into low and high downtime cost in terms of the different optimal thresholds. Thresholds $\boldsymbol{d}_1$ are [−0.5, −1, 0] and [−1.5, 0, 0] for low DCR (e.g. λ=0.12 or λ=0.21) and high DCR (e.g. λ=0.35), respectively. The reason for the difference of the optimal thresholds is that performing more PMs are cost-effective for cases with lower DCR while fewer PMs are needed for cases with higher DCRs.

2) Influence of reliability levels

The threshold of the proposed policy can be influenced by the reliability of the multi-component system, as well as the DCR. In the sensitivity analysis of reliability levels to the PM threshold, five cases with different values of the scale parameter η in the PHM model are tested. For case R1, the scaling parameter η are [1000 1500 800] and for case R2, the scaling parameter η are set to be [1000 1500 800]×2, and so on. The higher the scale parameter η, the higher the reliability is.

For case R1 with λ=0.12, the optimal thresholds are [−1, −1, 0] (Table 5). Whereas, for cases with the higher reliability levels, e.g. R2, R3, R4, the optimal thresholds are [−1.5, 0,

Table 4. Optimal results of the EPCLP for mean price scenario with various downtime cost ratios

λ	$\boldsymbol{d}_1$	d_2	Outages	CM	PM	OM	Cost of EPCLP	cost of CLP	cost -saving
0.12	[−0.5,−1,0]	−1.1	11.7	1.3	13.9	3.2	900	960	7%
0.21	[−0.5,−1,0]	−1.1	10.9	1.5	12.8	3.0	1200	1240	3%
0.35	[−1.5, 0,0]	−2	8.4	2.7	9.9	11.5	1730	1810	4%
<0.5	−0.5	−1	9.5	2.2	8.5	7.1			

Table 5. Optimal results for the case with mean price scenario and λ=0.12 via the EPCLP

case	d_1	d_2	Outages	CM	PM	OM	cost ($/day)
R1	[−1,−1,0]	−1.1	11.7	1.3	13.9	3.2	890
R2	[−1.5,0,0]	−2	4.2	1.3	5.3	5.1	510
R3	[−1.5,0,0]	−2	3.1	0.7	4.1	3.1	320
R4	[−1.5,0,0]	−2	2.7	0.4	3.3	2.5	230

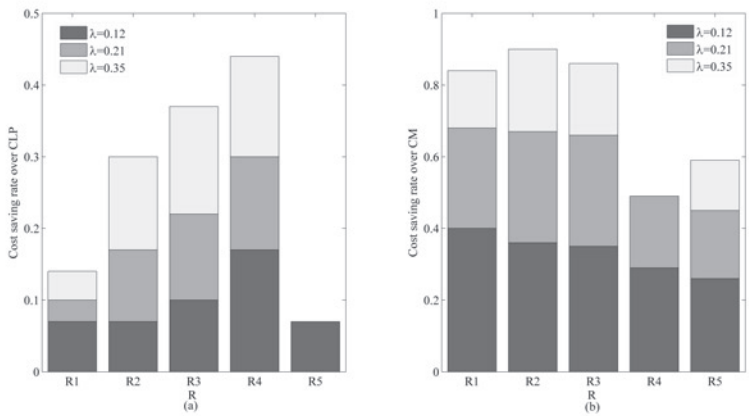


Fig. 5. Cost saving rate of EPCLP over CLP a) and CM b) for different reliabilities and DCRS

0]. Since fewer PMs are needed for latter cases, the threshold of the price level which has the minimal continuous price periods should be the minimum. Nevertheless, for the mean price scenario with low DCR (e.g. λ=0.12), more PMs are needed for cases with lower reliability levels. So for low reliability cases the optimal thresholds are [−1, −1, 0] while the optimal thresholds are [−1.5, 0, 0] for high reliability cases. Since fewer PMs are needed as the reliability increases, the threshold is minimal during the price periods for which the number of continuous time periods is minimal. It is economical to perform fewer PMs for cases with higher reliability levels. And it is cost-efficient to set the threshold of the low price periods to be minimal since the continuous time periods ratio for different electricity price levels is T_{cL}:T_{cM}:T_{cH} =3:9:6.

The cost-saving of the proposed policy over the constant CLP (or over the CM policy) for various downtime cost ratios and reliability levels are shown in Fig. 5(a) and Fig. 5(b), respectively. The results show that the proposed policy can provide a cost saving of about 3% to 17% over CLP for the cases with valid reliabilities (e.g. R1, R2 and R3) and DCRs (e.g. λ≤0.35). Whereas for cases with higher reliabilities (e.g. R4 and R5), the valid range of the DCR becomes much smaller, e.g. such as λ≤0.21 and λ≤0.12 for case R4 and case R5, respectively. For cases with higher reliabilities (e.g. R5) and higher DCR (e.g. λ≥0.21), the proposed policy can be no longer cost-effective than CLP since few PMs are needed.

3) Influence of covariate processes

To analyze the influence of the covariate on the performance of the proposed policy, the parameterized transition matrix is:

$$M_\alpha = \begin{bmatrix} 1-\alpha & \alpha/3 & \alpha/3 & \alpha/3 \\ \alpha/3 & 1-\alpha & \alpha/3 & \alpha/3 \\ \alpha/3 & \alpha/3 & 1-\alpha & \alpha/3 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Compared to the CLP, the cost-saving rate of EPCLP generally decreases as the parameter α increases (Fig. 6). For example, in the case when α=0.01 and λ=0.12, the cost-saving rate is 15%, and it is reduced to 6% as a is increased to 0.2.

Since more PMs are needed as a increases, the optimal price-dependent thresholds can be affected as a result. For example, when DCR is λ=0.12, the optimal price-dependent level-1 thresholds for 0.05≤α<0.2 and 0.2≤α<0.9 are [−1.5, −0.5, 0] and [−0.5, −1.5, 0], respectively. The changes of the optimal thresholds show that more PMs are needed for rapid covariate degradation.

The threshold of the average price periods is set to be lowest since the number of continuous middle price periods (T_{cM}) is the highest. From the relationships between the optimal thresholds and the price scenarios (more details are discussed in Section 3.3-4), the ratio for continuous periods of low, average, and high price T_{cL}:T_{cM}:T_{cH} is 3:9:6. The changes in the optimal thresholds show that more PMs are needed

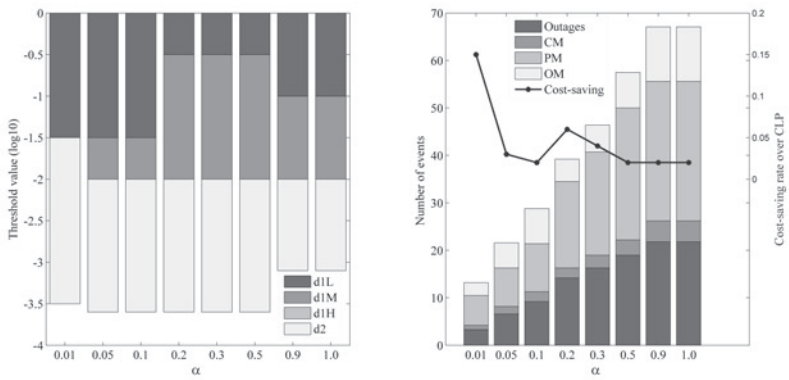


Fig. 6. Comparison of EPCLP and CLP in terms of degradation parameter alpha

for rapid covariate degradation. So the lowest threshold of the price-dependent thresholds is set during the average price periods since the number of continuous periods for the average price is the largest.

4) Influence of electricity price scenarios

For the electricity price-dependent control-limit policy, it is instructive to change the property of the electricity price scenario to analyze the influence of different price scenarios to the performance of the two policies.

From the results of the optimal thresholds for different price scenarios, there may exist specific relationships between the optimal thresholds and some specific properties of the electricity price process. To verify this conjecture, 13 typical price scenarios are selected from the total 59049 price scenarios (e.g. price scenario 00001, 01249, 01250, 04150, 06292, 31321, 31869, 48037, 58002, 58888, 58917, 58937 and 59030 are the label for the forecasting electricity scenarios). The price scenarios can be generated and reduced by the scenario generation method [12].

Fig. 7. Schematic representation of continuous periods C and cumulative periods S

To summarize the relationships between the optimal thresholds and the electricity price scenarios from the simulation results, some properties of the electricity price scenarios can be detected. Firstly, T_{cL}, T_{cM}, T_{cH} are defined to be the numbers of continuous periods for above-average price, average price and below-average price, respectively. Secondly, T_{sL}, T_{sM}, T_{sH} denote the sum of the periods for a low price level, an average price level and a high price level, respectively. As shown in Fig. 7, the ratio among the numbers of continuous periods for below-average price, average price and above-average price is T_{cL}: T_{cM}: T_{cH}=3:2:1. Moreover, the ratio among the sum of the periods for a low price level, an average price level and a high price level is T_{sL}: T_{sM}: T_{sH}=4:5:1.

The relationships between the optimal thresholds and price scenarios (Table 6) are summed up from the optimal thresholds of various price scenarios. The relationship between the optimal thresholds and price scenarios can be summarized in terms of different DCRs.

(1) Various price scenarios with low DCRs and low reliabilities;

For the cases with lower DCRs (e.g. λ<0.2) and lower reliabilities (e.g. case R1), more PMs are needed to be performed. The relationship between the optimal thresholds and the price scenarios can be divided into three categories.

For the category (a) (e.g. scenario 00001, 58888, 58917, 58937, 59030), the minimal threshold can be assigned to the price levels which achieve the highest number of continuous price periods max[T_{cL}, T_{cM}, T_{cH}]. For the price scenario 00001, the largest number of continuous price periods is the average price, i.e. max[T_{cL}, T_{cM}, T_{cH}]=max[6, 9, 3]= T_{cM}, the minimal threshold can be assigned to the average price levels, i.e. min[d_{1L}, d_{1M}, d_{1H}]= d_{1M}.

For the category (b) (e.g. scenario 04150, 06292, 31321, 48037), for the price level which achieves the highest number of continuous price periods, and the sum of the corresponding price periods is relatively much higher than other price levels, the minimal thresholds can be assigned to the other price levels. For the price scenario 04150, the largest number of continuous price periods is the average price, i.e. max[T_{cL}, T_{cM}, T_{cH}]=max[3, 7, 3]= T_{cM}. And the sum of the periods for different price levels are [T_{sL}, T_{sM}, T_{sH}]=[9, 18, 9]. The sum of the periods for average price level (T_{sM}=18) is much larger than that of low and high price levels. If a minimal threshold is assigned to the average price level, much more PMs can be performed since T_{sM} is much higher than T_{sL} and T_{sH}. As a result, too much PMs will be performed and the maintenance cost can be increased. So it will be cost effective to perform PM during the low and high price periods instead of during the average price periods for price scenario 04150.

For the category (c) (e.g. scenario 01249, 01250, 31869, 58002), the largest continuous price periods max[T_{cL}, T_{cM}, T_{cH}] are much higher than other price levels, the minimal thresholds can be assigned to the price levels with second largest continuous price periods. For the price scenario 01250, the largest number of continuous price periods is the average price, i.e. max[T_{cL}, T_{cM}, T_{cH}]=max[6, 12, 6]=12=T_{cM}, then min[d_{1L}, d_{1M}, d_{1H}] can be assigned to the low price level.

(2) Various price scenarios with higher DCRs (e.g. $\lambda \geq 0.2$) or higher reliabilities.

For the cases with higher downtime cost ratios (e.g. $\lambda \geq 0.2$) or higher reliabilities of the multi-component system (e.g. case R2), fewer PMs are needed. So the minimal threshold can be assigned to the price levels which have the lowest continuous price periods (e.g. scenario 01249, 1250, 04150, 06293, 31321, 31869, 48037, 58002, 58917 and 59030) or the second lowest continuous price periods (e.g. scenario 00001, 58888 and 58937).

4. Conclusions

This paper has proposed an electricity price-dependent control-limit policy (EPCLP) for the power generating unit taking into account the electricity price-dependent downtime cost for CBM. Since the proposed EPCLP can take full advantage of the changes of the fluctuating electricity prices, it can make further maintenance cost reduction in comparison of the constant control-limit policy. From the extensive computational analysis, it can be concluded that the EPCLP holds a significant advantage of cost-saving if the downtime cost rates, reliabilities of the multi-component system and the covariate are among the valid ranges. Future extensions of this work will focus on the investigating optimal CBM policy for complex multi-component system that build on our control-limit policy. In addition, although we assume the forecasting electricity is deterministic, it is also possible to consider the case in which the forecasting electricity is of uncertainty.

Table 6. Specific relationships between the price scenarios and optimal thresholds via the price-dependent control-limit policy for case R1

category	price scenario	T_{cL}	T_{cM}	T_{cH}	T_{sL}	T_{sM}	T_{sH}	min[d_{1L}, d_{1M}, d_{1H}]	
								λ<0.2	$\lambda \geq 0.2$
a	00001	6	9	3	24	9	3	d_{1M}	d_{1M}
	58888	6	9	3	18	9	9	d_{1M}	d_{1M}
	58937	3	9	6	9	9	18	d_{1M}	d_{1M}
	58917	6	9	6	12	12	12	d_{1M}	d_{1L}
	59030	3	9	6	9	9	18	d_{1M}	d_{1L}
b	04150	3	7	3	9	18	9	d_{1H}, d_{1L}	d_{1L}
	06292	3	7	3	9	15	12	d_{1H}, d_{1L}	d_{1L}
	31321	3	7	3	9	18	9	d_{1H}, d_{1L}	d_{1L}
	48037	3	7	3	9	18	9	d_{1H}, d_{1L}	d_{1L}
c	01249	6	10	3	15	18	3	d_{1L}	d_{1L}
	01250	6	12	6	15	15	6	d_{1L}	d_{1L}
	31869	6	12	6	12	12	12	d_{1L}	d_{1L}
	58002	6	12	6	12	12	12	d_{1L}	d_{1L}

Acknowledgement

This work is supported by National Natural Science Foundation of China (Grant No. 71501148, 51379081). The research work is also supported by the Key Laboratory of Metallurgical Equipment and Control of Education Ministry of Wuhan University of Science and Technology (Grant No. 2015B05).

References

1. Ahmad R, Kamaruddin S. An overview of time-based and condition-based maintenance in industrial application. Computers & Industrial Engineering 2012; 63(1): 135-149, http://dx.doi.org/10.1016/j.cie.2012.02.002.
2. Banjevic D, Jardine AKS, Makis V, Ennis M. A control-limit policy and software for condition-based maintenance optimization. INFOR 2001; 39(1): 32-50.
3. Castanier B, Grall A, Berenguer C. A condition-based maintenance policy with non-periodic inspections for a two-unit series system. Reliability Engineering & System Safety 2005; 87(1): 109-120, http://dx.doi.org/10.1016/j.ress.2004.04.013.
4. Chen A. and Wu G. Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging Markovian deterioration. International Journal of Production Research 2007; 45(15): 3351-3379, http://dx.doi.org/10.1080/00207540600677617.
5. Chen D, Trivedi K S. Optimization for condition-based maintenance with semi-Markov decision process. Reliability Engineering & System Safety 2005; 90(1): 25-29, http://dx.doi.org/10.1016/j.ress.2004.11.001.
6. Cox D. Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological) 1972; 34(2): 187-220.
7. Diyin T, Jinsong Y. Optimal Replacement Policy for a Periodically Inspected System Subject to the Competing Soft and Sudden Failures. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2015; 17(2): 228-235, http://dx.doi.org/10.17531/ein.2015.2.9.
8. Elwany A H, Gebraeel N Z, Maillart L M. Structured Replacement Policies for Components with Complex Degradation Processes and Dedicated Sensors. Operations Research 2011; 59(3): 684-695, http://dx.doi.org/10.1287/opre.1110.0912.
9. Gorjian N, Ma L, Mittinty M, Yarlagadda P, Sun Y. A review on degradation models in reliability analysis, in Proceedings of the 4th World Congress on Engineering Asset Management 2009. Athens.
10. Gröwe-Kuska N, Kiwiel K, Nowak M, Nowak M P, Römisch W, Wegner I. Power management under uncertainty by Lagrangian relaxation. in Proceedings of the 6th International Conference Probabilistic Methods Applied to Power Systems PMAPS 2000. INESC Porto.
11. Hamilton J D. Time series analysis. New York: Cambridge Univ Press, 1994.
12. Heitsch H, Römisch W. Scenario reduction algorithms in stochastic programming. Computational optimization and applications 2003; 24(2): 187-206, http://dx.doi.org/10.1023/A:1021805924152.
13. Jardine A K S, Makis V, Banjevic D, Braticevic D, Ennis M. A decision optimization model for condition-based maintenance. Journal of Quality in Maintenance Engineering 1998; 4(2): 115-121, http://dx.doi.org/10.1108/13552519810213644
14. Jardine A K S, Lin D, and Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing 2006; 20(7): 1483-1510, http://dx.doi.org/10.1016/j.ymssp.2005.09.012.
15. Jiang R. A multivariate CBM model with a random and time-dependent failure threshold. Reliability Engineering & System Safety 2013; 119: 178-185, http://dx.doi.org/10.1016/j.ress.2013.05.023.
16. Jiang R. Optimization of alarm threshold and sequential inspection scheme. Reliability Engineering & System Safety 2010; 95(3): 208-215, http://dx.doi.org/10.1016/j.ress.2009.09.012.
17. Liu Q, Dong M, Lv W, Geng X, Li Y. A novel method using adaptive hidden semi-Markov model for multi-sensor monitoring equipment health prognosis. Mechanical Systems and Signal Processing 2015; 64-65: 217-232, http://dx.doi.org/10.1016/j.ymssp.2015.03.029.
18. Makis V, Jardine A K S. Optimal replacement in the proportional hazards model. INFOR 1992; 30(1): 172-183.
19. Makis V, Jiang X. Optimal Replacement Under Partial Observations. Mathematics of Operations Research 2003; 28(2): 382-394, http://dx.doi.org/10.1287/moor.28.2.382.14484.
20. Marseguerra M, Zio E, Podofillini L. Condition-based maintenance optimization by means of genetic algorithms and Monte Carlo simulation. Reliability Engineering & System Safety 2002; 77(2): 151-165, http://dx.doi.org/10.1016/S0951-8320(02)00043-1.
21. PJM. Hourly Real-Time & Day-Ahead LMP. [cited 2014; Available from: http://www.pjm.com/markets-and-operations/energy/real-time/onthlylmp.aspx.
22. Qian X. Control-limit Policy of Condition-based Maintenance Optimization for Multi-component System by Means of Monte Carlo Simulation. IAENG International Journal of Computer Science 2014; 41(4): 269-273.
23. Tian Z, Liao H. Condition based maintenance optimization for multi-component systems using proportional hazards model. Reliability Engineering & System Safety 2011; 96(5): 581-589, http://dx.doi.org/10.1016/j.ress.2010.12.023.
24. Wang W. A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance. International Journal of Production Research 2000; 38(6): 1425-1436, http://dx.doi.org/10.1080/002075400188933.
25. Yang K, Yang G. Degradation Reliability Assessment Using Severe Critical Values. International Journal of Reliability, Quality and Safety Engineering 1998; 5(1): 85-95, http://dx.doi.org/10.1142/S0218539398000091.
26. Zhao X, Fouladirad M, Bérenguer C, Bordes L. Condition-based inspection/replacement policies for non-monotone deteriorating systems with environmental covariates. Reliability Engineering & System Safety 2010; 95(8): 921-934, http://dx.doi.org/10.1016/j.ress.2010.04.005.

Xinbo QIAN
The Key Laboratory of Metallurgical Equipment and Control of Education Ministry
Wuhan University of Science and Technology
No. 947 Heping Avenue, Wuhan 430081, China

Yonggang WU
School of Hydropower and Information Engineering
Huazhong University of Science and Technology
No. 1037 Luoyu road, Wuhan 430074, China

E-mails: xinboqian@wust.edu.cn, ygangwu@163.com

Article citation info:
TCHÓRZEWSKA-CIEŚLAK B, PIETRUCHA-URBANIK K, URBANIK M. Analysis of the gas network failure and failure prediction using the Monte Carlo simulation method. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 254–259, http://dx.doi.org/10.17531/ein.2016.2.13.

Barbara TCHÓRZEWSKA-CIEŚLAK
Katarzyna PIETRUCHA-URBANIK
Marek URBANIK

ANALYSIS OF THE GAS NETWORK FAILURE AND FAILURE PREDICTION USING THE MONTE CARLO SIMULATION METHOD

ANALIZA AWARYJNOŚCI SIECI GAZOWYCH ORAZ PROGNOZOWANIE AWARII Z ZASTOSOWANIEM SYMULACYJNEJ METODY MONTE CARLO

The scope of the article includes the analysis of the gas network failure based on a material obtained from field tests covering the years 2004-2014, conducted on the gas network of 120 thousand city, allowing to specify the failure rate of the gas network with division into material, pressure and pipelines diameter and indicate the main causes of failure on gas networks. On the base of the results of this analysis the Monte Carlo method to predict failures in gas pipe network has been presented.

***Keywords**: failure of gas network, Monte Carlo method, analysis of the failure structure, failure prediction.*

Artykuł swoim zakresem obejmuje analizę awaryjności sieci gazowej na podstawie uzyskanego materiału z badań eksploatacyjnych obejmujących lata 2004-2014 prowadzonych na terenie Zakładu Gazowniczego w 120 tys. mieście, co pozwoliło na podanie intensywności uszkodzeń sieci gazowych z podziałem na materiał, ciśnienie i średnice rurociągów oraz podanie głównych przyczyn powstawania awarii na sieciach gazowych. Na podstawie wyników analizy zaprezentowano zastosowanie metody Monte Carlo do prognozowania awarii sieci gazowych.

***Słowa kluczowe**: awaryjność sieci gazowej, metoda Monte Carlo, analiza struktury awaryjności, prognozowanie awaryjności.*

1. Introduction

In the report of the Committee of Union Gas in the World Congress in Nice in 1972, the issue connected with the failure intensity of the gas network was presented. Operation of Gas Supply System is inseparably linked with the occurrence of the fire and explosion risk. Failures, explosions and fires long-term statistics conducted by the gas services and organizational units of the State Fire Service indicate that despite the continuous increase of the safety level of gas supply system and number of actions taken in this regard, still many failures are recorded, which often generate serious threat for the external and the inner environment of a man, as well as create risk for human health and life. The consequences of failures, during which the release and dispersion of natural gas occurs, depend on its type, nature and causes, as well as from the efficiency of its removal.

Failure frequency evaluation in the operation of gas supply subsystem should be one of the priorities of the gasworks, which should attach greater importance to assess the proper functioning of the newly designed, as well as the implemented systems [31]. With the increasing pressure from the environmentalists and stricter standards for acceptable environmental pollution and structures vulnerability in terms of its protection, it seems to be a necessity [6, 24, 28]. The issue connected with failure gas pipeline analysis in perspective of improper design, construction and maintenance is presented in work [11], the increased failure affected by the lack or improper conduct of repairs and modernization, also incorrect or lack of risk management program. When planning renovations or modernization of gas networks the gasworks employees should make the classification of pipelines, that means determine which sections of the network require immediate repair, and which can be repaired later [7, 15, 26]. For this reason, making the appropriate maintenance (repair, renovation, replacement) only after the damage of the element may be irrational, control human intervention detects and removes faults being a potential source of failure. Widely used solution become the preventive renewals, aimed at reducing the loss of utility of a given element in different environment [4, 8, 21]. Very important from operator perspective is the failure reason [18, 20, 29], as well as modernizing actions that should be taken to avoid such undesirable situations. The strategy of the preventive renewals is to establish such timing of the renewals which will enable to achieve the maximum profitability of the project, through the use of the periodic strategies involving the renewals after a certain period of element operation [5, 11]. Such classification can be made, for example, on the basis of failure prediction for certain sections of the gas network, using the Monte Carlo method [30, 34]. The Monte Carlo methods include all proceedings aimed at finding approximate solutions of some problems (mathematical, technical or operational) [27]. The Monte Carlo method involves estimating the probability of occurrence of certain events based on previous studies [1, 12]. The assessment of the polyethylene pipes properties of gas networks in terms of operation safety, as well as activities that influence the increase of operation reliability and safety improvement was proposed in [3]. Interesting approach for assessment of gas pipes defects was presented by [23], in which corrosion and gouges defects were included. In work [17] significant issues related to ensuring the safety of pipelines, through research methods and the improvement of the technical condition were presented. Such composition was also prescribed in [10, 14, 16, 19] as to identify the most common failure causes in

addition to the existing design and construction data, as well as visual physical inspection. Nevertheless random character of failure occurrence makes analysis and assessment in this area very complex and based mainly on the analysis of operational data and implementation of methods and analysis of the failure mechanisms under real conditions as shown in works [9, 22, 25, 32, 33].

According to US Department of Transportation the trend in pipeline safety has demonstrated a stable decline in incidents concerning deaths and injuries, in the last twenty years and decrease from about sixty in nineties to forty. But on the other hand still more than fifty percentage of gas network was constructed in fifties or sixties, what can cause much serious failures or even gas explosion. Also many programs were implemented to improve this situation, there is still necessity for continuous improvement of gas network condition in order to reduce the failure rate. As to perform this actions, the analysis of typical gas network was proposed in this work, which aim will be to eliminate failure or serious pipeline incidents.

Also the aim of the work is the possibility of forecasting (prediction) failure as to minimize their possible impact, what is very important for safety reasons for users in subsystem of natural gas supply (SNGS), for this purpose, the Monte Carlo simulation method was used.

2. Analysis of failure in subsystem of natural gas supply based on operational data

2.1. Preliminary assessment of pipelines' technical state

The SNGS is powered by a ring high-pressure network through the 43 first degree reduction and metering stations, which supply medium pressure rings of various districts of the city and the surrounding regions. The age and material structure of the network are presented in Fig. 1 and 2.

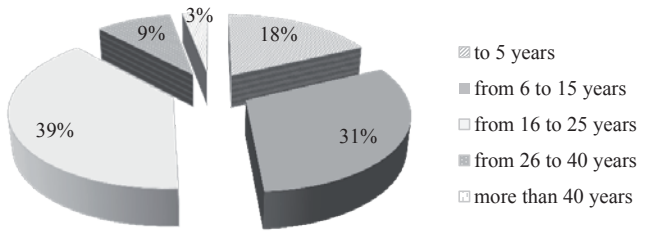


Fig. 1. The age structure of the gas network - a state for 2014, in %

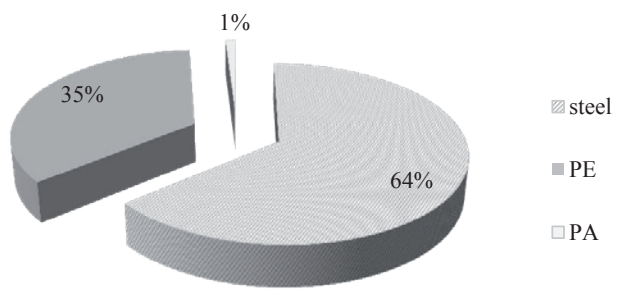


Fig. 2. The material structure of the gas network - a state for 2014, in %

Data on failures on the main taps and reduction points as well as distribution networks in the years 2004-2014 are presented in Fig. 3.

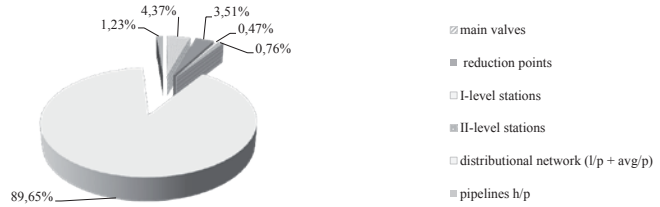


Fig. 3. Failures occurring in the gas network in the years 2004-2014, in %

Figure 4 shows graphically the results of the analysis of causes of failure in the l/p (low pressure) steel and plastic networks: PE - polyethylene and PA - polyamide and m/p (medium pressure) steel and plastic networks. Figure 5 summarizes the number of failures in the l/p steel and plastic networks and in m/p steel and plastic networks, in the assumed diameters ranges.

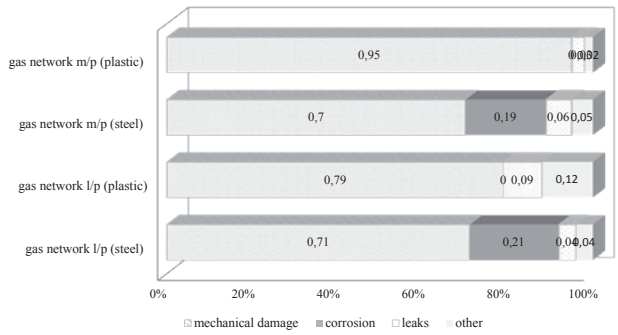


Fig. 4. Failures of gas networks l/p and m/p in the years 2004-2014 with division into the causes of their occurrence, in %

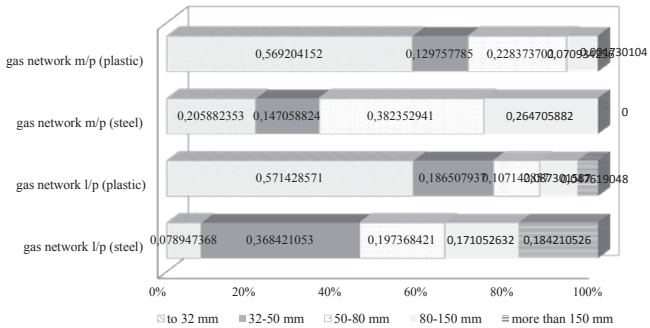


Fig. 5. Failures in the assumed range of pipelines diameters in the years 2004-2014, in %

2.2. Water network failure

The unit failure rate λ_i for l/p and m/p gas pipelines, with division into material, was calculated according to the formula (1):

$$\lambda_i = k_i\,(t,t+\Delta t)/(\,l_i \cdot \Delta t)$$

where:

λ_i	–	unit failure rate for i^{th} type of network or i^{th} type of fittings, [number of failures / km · year];
$k_i\,(t,t+\Delta t)$	–	the total number of failures in the time interval Δt in a given type of network,
l_i	–	length of a given type network, in a given period of time, in which failures occurred, [km];
i	–	type of network;

Δt – the time interval, [one year].

2.3. Discussion of results

The detailed analysis showed that the average values of the failure rate index of gas pipelines are:

- failure rate of l/p gas pipelines $\lambda_{avgusteel}$ = 0.006987 [number of failures/km · year],
- failure rate of l/p gas pipelines $\lambda_{avguplastic}$ = 0.062352 [number of failures/km · year],
- failure rate of m/p gas pipelines $\lambda_{avgusteel}$ = 0.005432 [number of failures/km · year],
- failure rate of m/p gas pipelines $\lambda_{avguplastic}$ = 0.027429 [number of failures/km · year].

In detail the overall situation of failure rate can be distinguished as follows on the Figure 6.

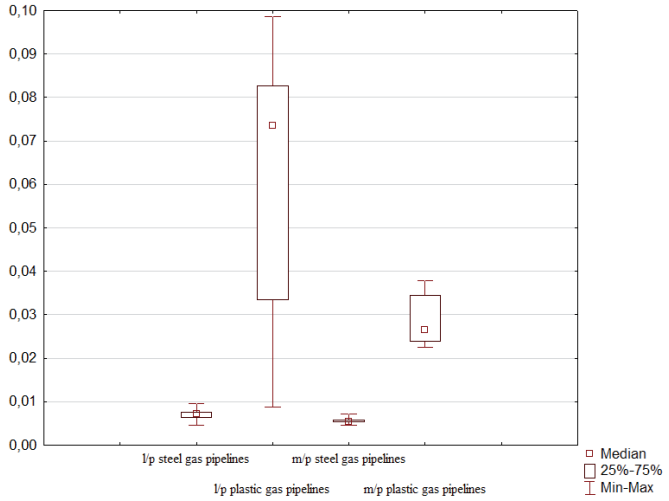


Fig. 6. Failure rate of a medium pressure gas pipeline, low pressure depending on the material type and on the network altogether

Most emergency pipes occurred to be plastic sections, they constitute nearly 65% of all, due to the fact that the sections made of this material constitute the most networks, what is confirmed by failure rate index. The mean operating time between failures T_{Pavg} in the considered years was 1.9 d. This analysis shows that the main cause of failures in gas distribution networks is the corrosion of steel pipes and mechanical damages of plastic pipes. The total average failure rate of gas pipelines was λ_{avg} = 0.02555 [number of failures/ km · year]. The analysis showed seasonality of the failures in the gas distribution networks, in spring and summer the number of failure increases and during autumn and winter it decreases, so it is important to increase the frequency of gas pipelines inspections and the use of monitoring during the periods of increased failure rate. The performed analyses show, among others, that the replacement of steel pipes by plastic pipes made in recent years, significantly reduced the number of network failures due to corrosion.

In the damage structure on all kinds of pipes, mechanical failure dominate and represent 78.7% of the total sum of failure. The effect of such situation could be negligent backfill placement of pipes, which during operation are susceptible to settlements and consequently breaking and too shallow laying pipes in the ground, so that they are exposed to high loads. The largest share of failures concerned network, which is older than 16 but having not more than 25 years. The cause of such situation is that many older pipes had already been replaced. The smallest number of failures occurred on the elements younger than 5 years.

3. The use of the Monte Carlo simulation method for predicting failures in SNGS

A concept of the adopted Monte Carlo simulation method can be presented in the simplest way by means of the following procedures [2]:

1. Calculate the probability $P(k,t)$ of the adopted subsystem reliability measure.

In the analysed example it is the probability of k failures in SNGS, calculated on the basis of data from exploitation. We used [13]:

$$P(k, t) = ((n \cdot \lambda \cdot t)\, k / k!) \exp(-n\lambda t) \tag{2}$$

2. Establish equal intervals of random numbers with lengths corresponding to the calculated probabilities.
3. Generate a sequence of independent random numbers occurring with equal probability ($N = (25 \div 100) \cdot 10^3$), for each random variable included in the reliability analysis of SNGS. In the shown example it is a number of failures "*k*", which appears in SNGS within a specified period of time.
4. After making a sufficiently large number of operations (draws), calculate the number of results found in the emergency areas for every failure "*k*".
5. Calculate the U_k index, which determines the probability of the occurrence of *k* failures within a specified time interval.

$$U_k = N_1/N \tag{3}$$

where N_1 is a number of hit random numbers in equal probability intervals, corresponding to a certain number of failures at time *t* and *N* is a number of all executed draws.

6. The measure of reliability in the analysed period is the index *K* calculated as:

$$K = 1 - U_k \tag{4}$$

Generally, the Monte Carlo method can be used for every element of SNGS, if only the values of failure probability are known. Computer programs gives the possibility to use the simple Monte Carlo simulation method to assess the reliability of SNGS. Figure 7 shows the program algorithm performed for simulation calculations of the number of failures on distribution networks in SNGS.

Once you start the program you must provide the following inputs:

- t - time,
- L - the length of the network,
- k_{max} - the maximum number of failures at time *t*,
- λ - the average unit failure rate.

Based on these data, the program calculates the probability $P(k,t)$ and determines the intervals of random numbers. Then the random numbers are generated, at every number the program checks whether a given number falls within the numbers for the ith "*k*", at the same time counting how many of the generated numbers are within the proper range. After the appropriate number of draws is made the program calculates the index U_k and the procedure is repeated from the beginning for the next *k*.

The result of the program is a series of grouped indexes U_k for the corresponding *k*. A simulation of 1000000 draws was performed for each *k* failure for l/p steel and PE networks and m/p steel and PE networks. The results are shown in graphical form in Figures 8, 9, 10, 11.

Implementation of the Monte Carlo method shows the prognosis of technical condition of gas pipe. From the analysis of presented figure 8-11 it seems that the intensified time for inspection and then rehabilitation of network made from PE in case of average pressure is the range from 101 to 154 years of exploitation, which greatly ex-

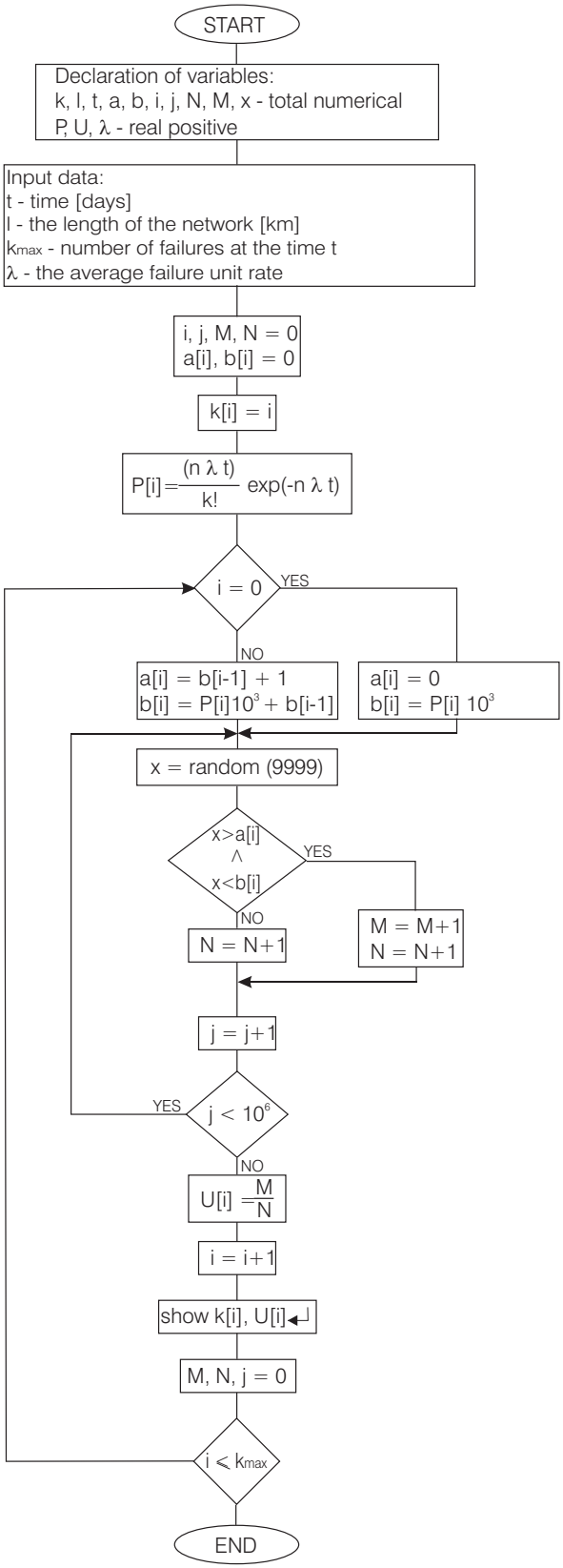


Fig. 7. The algorithm of the program to assess the reliability of the SNGS by the Monte Carlo method, where M is the number of random numbers getting hit with compartments of equal probability, N is number of all performed lotteries, U is the probability, a predetermined range of time when k failure occurs, P is the probability of k failure occurrence at the t time [27]

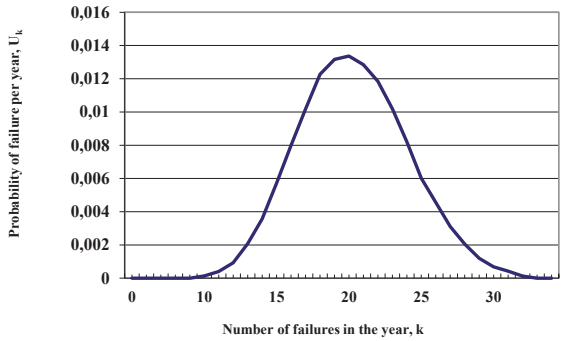


Fig. 8. Simulation using the Monte Carlo probability U(k,t) of the forecasted k failures in the network of steel l/p, Λ_{avg} = 16,4 failure/year

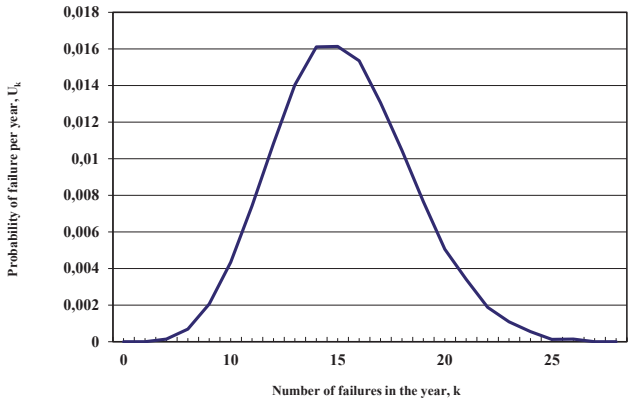


Fig. 9. Simulation using the Monte Carlo probability U(k,t) of the forecasted k failures in the network of PE l/p, Λ_{avg} = 13,8 failure/year

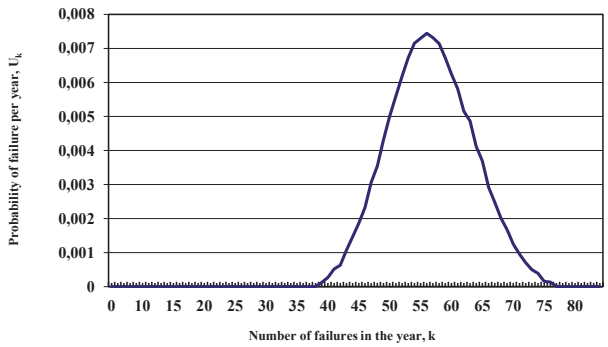


Fig. 10. Simulation using the Monte Carlo probability U(k,t) of the forecasted k failures in the network of steel avg/p, Λ_{avg} = 53,6 failure/year

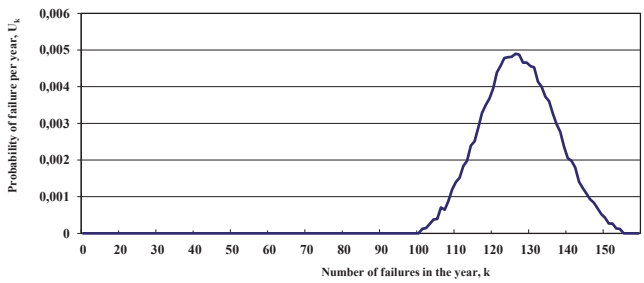


Fig. 11. Simulation using the Monte Carlo probability U(k,t) of the forecasted k failures in the network of PE avg/p, Λ_{avg} = 137,2 failure/year

ceeds the lifetime of the gas network and constitute good prognosis for investments of network rehabilitation. Also on the example of steel pipes of average pressure the time of intensified observation should be delayed longer in time from 37 to 74 years. The situation is different in case of network of low pressure, it seems that the probability of failure considerably increases, for example U_k, which equal 1,20E-02 and is reached after eleven years of exploitation for PE pipe. For the same probability of failure, but for the steel pipe (l/p) it is attained after eighteen years of gas network functioning. Such prognosis should point a direction for conducting preventive modernization of gas pipes.

4. Conclusion and perspectives

The simulation Monte Carlo methods can be used to predict failures occurring in the gas networks, which allows to classify properly the elements of the subsystem requiring modernisation or general overhaul. It could be very helpful in performing and planning of operation strategy prediction.

As to perform the prioritisation pipes for rehabilitation, the failure and prognosis analysis of the gas network should be conducted. It constitute the crucial element of the management of urban gas network, mainly in the strategic modernization plans, as well as it supports the rehabilitation techniques.

Further research should address the introduction of methods for analyzing failure during which more information data from different gas network will be gathered and constitute guidelines describing the possibility of failure occurrence on the gas pipes. It also seems necessary to indicate to discuss the criterion concerning the effectiveness of the gas supply system functioning.

References

1. Adhikari R, Agrawal R K. Performance evaluation of weights selection schemes for linear combination of multiple forecasts, Artificial Intelligence Review 2014; 4(42): 529-548, http://dx.doi.org/10.1007/s10462-012-9361-z.
2. Borgoń J, Jaźwiński J, Klimaszewski S, Żmudziński Z, Żurek J. Simulation methods for testing the safety of flights, WNT, Warsaw 1998.
3. Brown, N., Crate, J.M.: Analysis of a failure in a polyethylene gas pipe caused by squeeze off resulting in an explosion, Journal of Failure Analysis and Prevention, 12 (1)2012: 30-36, http://dx.doi.org/10.1007/s11668-011-9527-z.
4. Cang S, Yu H. A combination selection algorithm on forecasting. European Journal of Operational Research 2014; 1(234): 127-139, http://dx.doi.org/10.1016/j.ejor.2013.08.045.
5. Chandrashekara AS, Ananthapadmanabha T, Kulkarni AD. A neuro-expert system for planning and load forecasting of distribution systems, International Journal of Electrical Power & Energy Systems 1999; 5(21): 309-314, http://dx.doi.org/10.1016/S0142-0615(98)00057-X.
6. Che JX. Optimal sub-models selection algorithm for combination forecasting model. Neurocomputing 2015; 1(151): 364-375, http://dx.doi.org/10.1016/j.neucom.2014.09.028.
7. Costantini M, Pappalardo C. A hierarchical procedure for the combination of forecasts, International Journal of Forecasting 2010; 4(26): 725-743, http://dx.doi.org/10.1016/j.ijforecast.2009.09.006.
8. Gonzalez PA, Zamarreno JA. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. Energy and Buildings 2005; 37: 595-601, http://dx.doi.org/10.1016/j.enbuild.2004.09.006.
9. Hao, Y.-M., Zhang, C.-S., Shao, H., Wang, M.-T. Bayes network quantitative risk analysis for failure of natural gas pipelines, Journal of Northeastern University, Volume 32, Issue SUPPL. 2011; 2: 321-325.
10. Huang K, Yang H, Shi J, Long T, Suxin: Analysis on natural gas pipeline network rehabilitation technology. Natural Gas Industry 2006; 26(4): A19-A20, 119-121.
11. Kent Muhlbauer W. Pipeline Risk Management, Gulf Publishing Company, London 1992.
12. Kizilaslan R, Karlik B. Combination of neural networks forecasters for monthly natural gas consumption prediction. Neural Network World 2009; 19(2): 191-199.
13. Knapik K, Wieczysty A. Theoretical aspects of the application of the Poisson distribution in determining the reliability of the water distribution subsystem. Mat. Conf. Water supply, quality and water protection, Cracow 2000.
14. Kowalski D, Miszta-Kruk K. Failure of water supply networks in selected Polish towns based on the field reliability tests. Engineering Failure Analysis 2013; 35: 736-742, http://dx.doi.org/10.1016/j.engfailanal.2013.07.017.
15. Lux T, Morales-Arias Leonardo. Relative forecasting performance of volatility models: Monte Carlo evidence. Quantitative Finance 2013; 9(13): 1375-1394, http://dx.doi.org/10.1080/14697688.2013.795675.
16. Majid ZA, Mohsin R, Yaacob Z, Hassan Z. Failure analysis of natural gas pipes. Engineering Failure Analysis 2010; 17(4): 818-837, http://dx.doi.org/10.1016/j.engfailanal.2009.10.016.
17. Majid ZA, Mohsin R, Yusof MZ. Experimental and computational failure analysis of natural gas pipe. Engineering Failure Analysis 2012; 19 (1): 32-42, http://dx.doi.org/10.1016/j.engfailanal.2011.09.004.
18. Matvienko AF, Filippov YuI, Sagaradze VV, Pecherkina NL, Baldin AV, Grigor'ev PA. Stress-corrosion cracking of steels for gas-main pipelines: III. Failure of pipes in the heat-affected zone. Physics of Metals and Metallography 2000; 90(3): 309-317.
19. Mohsin R, Majid ZA, Yusof MZ. Multiple failures of API 5L X42 natural gas pipe: Experimental and computational analysis, Engineering Failure Analysis 2013; 34: 10-23, http://dx.doi.org/10.1016/j.engfailanal.2013.07.007.
20. Mohsin R, Majid ZA. Erosive failure of natural gas pipes. Journal of Pipeline Systems Engineering and Practice 2014; 5(4), http://dx.doi.org/10.1061/(ASCE)PS.1949-1204.0000170.
21. Ondrejka Harbulakova V, Estokova A, Stevulova N, Luptakova A. Different aggressive media influence related to selected characteristics of concrete composites investigation. International Journal of Energy and Environmental Engineering 2014; 5 (2-3): 1-6, http://dx.doi.org/10.1007/s40095-014-0082-8.
22. Pietrucha-Urbanik K, Tchórzewska-Cieślak B. Water Supply System operation regarding consumer safety using Kohonen neural network; in: Safety, Reliability and Risk Analysis: Beyond the Horizon - Steenbergen et al. (Eds), Taylor & Francis Group, London 2014: 1115-120.

23. Pluvinage G, Capelle J, Schmitt C, Mouwakeh M. Domain failure assessment diagrams for defect assessment of gas pipes. 19th European Conference on Fracture: Fracture Mechanics for Durability, Reliability and Safety, ECF 2012.
24. Potocnik P, Govekar E, Grabec I. Short-term natural gas consumption forecasting. DeFelice F (ed.), Proceedings of the 16th Iasted International Conference On Applied Simulation And Modelling, Palma de Mallorca, Spain Aug 29-31, 2007, IASTED International Conference on Modelling and Simulation 2007: 353-357.
25. Szybka J, Broniec Z, Pilch R. Forecasting the failure of a thermal pipeline on the basis of risk assessment and exploitation analysis. Eksploatacja i Niezawodnosc -Maintenance and Reliability 2011; 4: 5-10.
26. Tchórzewska-Cieślak B, Rak J. Analysis of gas networks failure in chosen cities. XII Conf. of Heat Engineers, Solina 2000.
27. Tchórzewska-Cieślak B. Reliability of selected elements of the natural gas supply subsystem, PhD Dissertation, Cracow 2002.
28. Valis D, Vintr Z, Malach J. Selected aspects of physical structures vulnerability - state-of-the-art. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2012; 3: 189-194.
29. Wilmott MJ, Diakow DA. Factors influencing stress corrosion cracking of gas transmission pipelines: detailed studies following a pipeline failure. Part 2: pipe metallurgy and mechanical testing. Proceedings of the International Pipeline Conference, IPC, 1/2016: 573-585.
30. Winkler RL, Makridakis S. The combination of forecasts. Journal of the Royal Statistical Society Series A-Statistics in Society 1983; 146, 150-157, http://dx.doi.org/10.2307/2982011.
31. Witek M. Risk of gas transmission through network operations. Gas, Water and Sanitary Engineering 2001; 1: 19-23
32. Xiangpeng Luo, Shunli Lu, Jianfeng Shi, Xiang Li, Jinyang Zheng Numerical simulation of strength failure of buried polyethylene pipe under foundation settlement Engineering Failure Analysis 2015; 48: 144-152, http://dx.doi.org/10.1016/j.engfailanal.2014.11.014.
33. Yin, Y.-L., Lin, G.-L.: Risk analysis of the city gas pipeline network based on the fault tree, IEEM 2009 - IEEE International Conference on Industrial Engineering and Engineering Management 2009: 2477-2481.
34. Zhou Y, Ma L, Mathew J, Sun Y, Wolff R. Asset life prediction using multiple degradation indicators and failure events: a continuous state space model approach. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2009; 4: 72-81.

Barbara TCHÓRZEWSKA-CIEŚLAK
Katarzyna PIETRUCHA-URBANIK
Marek URBANIK

Department of Water Supply and Sewerage Systems
Rzeszow University of Technology
Al. Powstancow Warszawy, 35-959 Rzeszow, Poland

E-mails: cbarbar@prz.edu.pl, kpiet@prz.edu.pl, marekurbanik@vp.pl

Article citation info:
LIU H, JIANG W, HULIO Z, WANG Q. Estimation of system reliability by using the PLS-regression based corrected response surface method. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 260–270, http://dx.doi.org/10.17531/ein.2016.2.14.

Huahan LIU
Wei JIANG
Zahid Hussain HULIO
Qiuzhi WANG

ESTIMATION OF SYSTEM RELIABILITY BY USING THE PLS-REGRESSION BASED CORRECTED RESPONSE SURFACE METHOD

OCENA NIEZAWODNOŚCI SYSTEMU Z WYKORZYSTANIEM POPRAWIONEJ METODY POWIERZCHNI ODPOWIEDZI OPARTEJ NA REGRESJI CZĄSTKOWYCH NAJMNIEJSZYCH KWADRATÓW

A new computational method, referred as PLS-regression (PLSR) based corrected response surface method, has been developed for predicting the reliability of structural and mechanical systems subjecting to random loads, material properties, and geometry. The method involves a Corrected-Response Surface Model (C-RSM) based on the Partial Least Squares Regression Method (PLSRM) combined with some correction factors, and Monte Carlo Simulation (MCS), which is named as the Corrected-Partial Least Squares Regression-Response Surface Method (C-PLSRRSM). In order to develop an accurate surrogate model for the region determining the reliability of the system, a proper coefficient is presented to determine the sampling region of the input random variables. Due to a small number of original function evaluations, the proposed method is effective, particularly when a response evaluation entails costly finite-element, mesh-free, or other numerical analysis. Three numerical examples involving reliability problems of two structural systems and a mechanical system illustrate the method developed. Results indicate that the proposed method provides accurate and computationally efficient estimates of reliability. The proposed correction method, the PLSR based corrected response surface (C-PLSR-RS), can be the accurate surrogate model for calculating system reliabilities, especially for the implicit performance functions.

Keywords: *Reliability; Mechanical System; Partial Least Squares Regression; Response Surface Method; Correction Method.*

Nowa metoda obliczeniowa o nazwie "poprawiona metoda powierzchni odpowiedzi oparta na regresji PLS" (C-PLSRRSM) została opracowana dla potrzeb przewidywania niezawodności systemów konstrukcyjnych i mechanicznych poddanych obciążeniom losowym oraz charakteryzujących się losową geometrią oraz losowymi właściwościami materiałowymi. W metodzie uwzględniono pewne czynniki korekcyjne oraz symulację Monte Carlo. W celu opracowania odpowiedniego modelu zastępczego dla regionu stanowiącego o niezawodności systemu, przedstawiono współczynnik, który pozwala określić obszar pobierania próbek wejściowych zmiennych losowych. Ze względu na niewielką liczbę ocen funkcji początkowych, proponowana metoda jest skuteczna zwłaszcza wtedy, gdy ocena odpowiedzi wymaga kosztownej analizy numerycznej metodą elementów skończonych czy metodą automatycznie generowanej siatki (free mesh). Opracowaną metodę zilustrowano za pomocą trzech przykładów numerycznych dotyczących niezawodności dwóch systemów konstrukcyjnych oraz jednego układu mechanicznego. Wyniki wskazują, że proponowana metoda zapewnia dokładne i wydajne obliczeniowo oszacowanie niezawodności. Proponowana metoda C-PLSR-RS może stanowić trafny model zastępczy do obliczania niezawodności systemu, zwłaszcza w przypadku uwikłanych funkcji stanu granicznego.

Słowa kluczowe: *niezawodność; układ mechaniczny; regresja cząstkowych najmniejszych kwadratów; metoda powierzchni odpowiedzi; metoda korekcji.*

1. Introduction

Suppose a system has m limit-state functions associated with its constituting components. If the relationship between the system reliability and component reliability is known, it is possible to compute the system reliability pRs through the component reliability pR_j. For a series system, the system works well only if all components operate well, then system reliability is the probability of intersection of the component reliability events [21,29], as shown in Eq. (1). For a parallel system, the system is reliable if any of the components works well, then system reliability pRs is, therefore, computed by the probability of the union of the component reliability events [29], as shown in Eq. (2).

$$pRs = \Pr\left\{\bigcap_{j=1}^{m} Y_j > 0\right\} \text{ (for a series system)}, \tag{1}$$

and

$$pRs = \Pr\left\{\bigcup_{j=1}^{m} Y_j > 0\right\} \text{ (for a parallel system)}. \tag{2}$$

Where Y_j is the j_{th} component limit-sate function included in the system performance function, which is expressed by:

$$Y_j = g_j(\mathbf{X}), (j = 1, \cdots, m), \tag{3}$$

and the size of the random variables $\mathbf{X} = [X_1, X_2, \cdots, X_n]^T$ is n. Then the component reliability event E_j is defined by the event $g_j(\mathbf{X}) > 0$, and the component reliability is then the probability:

$$pR_j = \Pr\{g_j(\mathbf{X}) > 0\}, (j = 1, \cdots, m). \quad (4)$$

The major task of system reliability analysis is to calculate pRs, given the joint distribution of $\mathbf{X}$ and the limit state functions $g_j(\mathbf{X}), (j = 1, \cdots, m)$. So far, one of the system reliability analysis method uses the component reliability without considering the dependency between the components failure. Approximate methods, such as the first- and second-order reliability methods (FORM/SORM) [4, 15, 28] and simulation methods [1, 23, 25, 26], are commonly employed to estimate the component reliability. Considering the failure dependency expressed by the linear correlation coefficient ρ_{ij}, which is the failure relationship between the i_{th} component and the j_{th} component, another system reliability calculation method is obtained. The linear correlation coefficient ρ_{ij} can be easily found with the linearized limit-state function $g_i(\mathbf{X})$ and $g_j(\mathbf{X})$. Details about this type of system reliability analysis can be referred in [6, 21, 33, 38].

Without considering the correlation between the components failure, or using the linear correlation coefficient to express the dependency relationship, the accuracy of the component (or system) reliability result will deteriorate with the increase of the nonlinearity in nonnormal-to-normal transformation. To solve these problems, several more accurate methods have been developed by investigators [3, 7, 9]. Through extending the saddlepoint approximation (SA) method [8, 16] used in component reliability analysis, Du [7] developed a SA based system reliability analysis method. However, the accuracy of the results is largely determined by the accuracy of linearization of limit-state functions in the vicinity of their associated Most Likelihood Points (MLPs) and the MLPs are acquired by the optimization iteration process which affects the efficiency of the reliability calculation. Efficient Global Reliability Analysis (EGRA) method [2] was extended to solve the system reliability problems by Barron [3]. It is based on the creation of Gaussian process surrogate models that are required to be locally accurate only in the regions which have the significant contributions to the system failure. However, a large number of iterations and a complex optimization process are needed to get the surrogate model, which will decrease the efficiency of the system reliability analysis. An active learning reliability method combining Kriging and MCS was presented by Echard [9]. Two kinds of active learning method, which are used to add the experiment points to mend the meta-model, are presented. However, every point of the sample population obtained from the Monte Carlo Sampling is needed to search once during each active learning process, and high computational cost occurs if the number of the sample point population is large. A fuzzy multi-objective genetic algorithm approach [24] was proposed to optimize the system reliability.

This paper presents a new computational method for predicting reliability of structural and mechanical systems subjecting to random loads, material properties, and geometry. The proposed method involves a small number of exact or numerical evaluations of the performance function, generation of approximate values of the performance function at arbitrarily large number of inputs using the C-RSM, and the reliability evaluation by using the MCS. Three numerical examples involving reliability problems of structural and mechanical system illustrate the effectiveness and accuracy of the proposed method. Whenever possible, to evaluate the accuracy and computational efficiency of the proposed method, comparisons have been made with direct MCS method which calculates the original performance functions to get the system reliability.

Section 2 provides a brief introduction to the partial least squares regression and response surface method. Section 3 describes the proposed corrected response surface method based on PLSR method, which involves a new correction method with a coefficient, and a new sample method with a proposed proper coefficient to bound the distribution region of the input random variables. Section 4 gives the simulation theory of the MCS method which is used to analysis the system reliability with complex component failure dependencies. Three numerical examples are illustrated in Section 5, and comparisons have been made with direct MCS method.

2. Partial least square regression method and response surface method

Partial least square regression (PLSR) has two algorithms, PLS1 (Sequential algorithm) for the univariate response variables and PLS2 for the multivariate response variables [20]. PLSR was used to simultaneously correlate the parameters and responses. PLSR is a method for relating two data matrices, $\mathbf{x}$ and $\mathbf{y}$ (in this paper, representing a pair of realization matrix of $\mathbf{X}$ and $\mathbf{Y}$ at the sampling data), by a linear multivariate model, but goes beyond traditional regression in that it models also the structure of $\mathbf{x}$ and $\mathbf{y}$. The core concept of the PLSR approach is to solve the multicollinearity in regression or calibration, and the further details of the PLSR can be found in Ref. [35]. Nowadays, the PLSR method is applied to analysis the component reliability [39, 40]. PLSR derives its usefulness from its ability to analyze data with many, noisy, collinear, and even incomplete variables in both $\mathbf{x}$ and $\mathbf{y}$. Unlike the traditional Multiple Linear Regression (MLR) method, PLSR actually uses the responses variable information during the decomposition process [13]; even the $\mathbf{x}$-variables data tend to be many and also strongly correlated, PLSR method also works well. Many studies have shown the potential of PLSR for estimating the parameters and demonstrated that PLSR was a better alternative to conventional stepwise regression [18, 30, 32]. PLSR is also known as the projection to the latent structures which are included in a relatively recent multivariate regression method that combines the aspect of the principal component regression(PCR) and multiple linear regression (MLR). PLSR is pertinent statistical choice when [a] there are many variables x that are correlated with many responses y and [b] there is missing data on experimental work [5]. In this paper, PLSR method will be used to produce the surrogate model of the original performance functions of a system. The meta-model, with simple and low nonlinear form, will be used to calculate the system reliability.

Response surface method is used to explore interaction among the parameters and predict properties on the experimental region [5]. RSM is also a effective tool in assessing the reliability of complex structures which requires a deal between reliability algorithms and mechanical methods used to model the mechanical behavior, and the interest of this method is that the user is allocated to choose and check the mechanical experiments [12, 31]**.** RSM was used to explore interactions among parameters and predict the failure regions. RSM methodology is a collection of mathematical and statistical techniques based on fitting of polynomial equation to the experimental data, and becomes a powerful tool for describing the studied system ,so prediction of its behavior can be made by the surface responses plots that represents the system under studied region [5, 27]. RSM was also used for analyzing the surface maps for different responses and detecting of interactions among variables and quadratic models presented on

the responses [14]. The procedure of using a least square regression analysis to obtain the parameters of a response surface around a design point has earlier been used by Faravelli [10]. RSM method is used for the several reasons but the important one is that the numerical derivation on the analytical response surface is available, which reduces the number of mechanical computations required and provides the information to decision maker to choose the judicious experiment chemometric tools like Design of Experiment (DOE), RSM, PCR, or PLSR [12]. These methodologies can be helpful when many variables and responses are presented in various processes and correlation.

3. Corrected response surface based on PLSR

3.1. A simple PLSR algorithm

PLSR has the ability to model one or several dependent variables, responses, $\mathbf{Y} = [Y_1, \ Y_2, \cdots, Y_m]^T$, by means of a set of predictor variables $\mathbf{X} = [X_1, X_2, \cdots, X_n]^T$. With multivariate PLSR and the observation data, linear combinations of the predictor variables are formed sequentially and related to $\mathbf{Y}$ by ordinary least squares regression. It is shown that these linear combination, here called latent variables (also called components and factors similar to the components using in the principal components regression), is viewed as weighted averages of predictors, where each predictor holds the residual information that is not contained in the earlier latent variables. And the quantity to be predicted is a weighted average of the residuals from separately regressing each Y_j response against earlier latent variables. A modeling problem including *m*-dependent variables, responses $Y_1, Y_2, \ldots, Y_m$, and *n*-independent variables, predictor variables (or input variables) $X_1, X_2, \cdots, X_n$ are considered here to explain the PLSR theory. Where $Y_1, Y_2, \ldots, Y_m$ are seen as *n* responses all affected by input variables (independent with each other) $X_1, X_2, \cdots, X_n$. By selecting *N* observations (sample points) composed of *N* input data vectors $\boldsymbol{x}_k = (x_{k1}, x_{k2}, \cdots, x_{kn})^T, k = 1, 2, \cdots, N$ and the corresponding *N* responses values vectors $\boldsymbol{y}_k = (y_{k1}, y_{k2}, \cdots, y_{km})^T, k = 1, 2, \cdots, N$ obtained from calculating the original performance functions, two matrices $\mathbf{x}$ and $\mathbf{y}$ of dimensions ($N * n$) and ($N * m$) are formed. Data of a PLSR method can be arranged in two tables, and usually have been centered and scaled before the analysis [11], which are expressed by Eq. (5) and Eq. (6), respectively:

$$\mathrm{E}_0 = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nn} \end{bmatrix}, \tag{5}$$

$$\mathrm{F}_0 = \begin{bmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{N1} & \cdots & y_{Nm} \end{bmatrix}, \tag{6}$$

Here, a simple PLSR algorithm used in this paper is described as below:

(1) Finding the eigenvector corresponding to the maximum eigenvalue of matrix $\boldsymbol{E}_0{}^T \boldsymbol{F}_0 \boldsymbol{F}_0{}^T \boldsymbol{E}_0$ as $\boldsymbol{w}_1$; calculating the latent variable as $t_1 = \mathrm{w}_1^T \mathbf{X}$; evaluating the score vector as $\widehat{\boldsymbol{t}_1} = \boldsymbol{E}_0 \boldsymbol{w}_1$ and the residual matrix as $\boldsymbol{E}_1 = \boldsymbol{E}_0 - \widehat{\boldsymbol{t}_1} \alpha_1^T$. Where $\alpha_1 = \mathrm{E}_0^T \widehat{\boldsymbol{t}_1} \Big/ \left\| \widehat{\boldsymbol{t}_1} \right\|^2$.

(2) Finding the eigenvector corresponding to the maximum eigenvalue of the matrix $\boldsymbol{E}_1{}^T \boldsymbol{F}_0 \boldsymbol{F}_0{}^T \boldsymbol{E}_1$ as $\boldsymbol{w}_2$; calculating the latent variable as $t_2 = \mathrm{w}_2^T \mathbf{X}$; evaluating the score vector as $\widehat{\boldsymbol{t}_2} = \boldsymbol{E}_1 \boldsymbol{w}_2$ and the residual matrix as $\boldsymbol{E}_2 = \boldsymbol{E}_1 - \widehat{\boldsymbol{t}_2} \alpha_2^T$. Where $\alpha_2 = \mathrm{E}_1^T \widehat{\boldsymbol{t}_2} \Big/ \left\| \widehat{\boldsymbol{t}_2} \right\|^2$.

(3) The same processes carried out repeatedly until the p_{th} step. Then $\boldsymbol{w}_p$ equals to the eigenvector corresponding to the maximum eigenvalue of the matrix $\boldsymbol{E}_{p-1}{}^T \boldsymbol{F}_0 \boldsymbol{F}_0{}^T \boldsymbol{E}_{p-1}$. The latent variable is given by $t_p = \mathrm{w}_p^T \mathbf{X}$ and the score vector is estimated by $\widehat{\boldsymbol{t}_p} = \boldsymbol{E}_{p-1} \boldsymbol{w}_p$.

(4) According to the *p* latent variables extracted by above steps, the regression model of $\boldsymbol{F}_0$ is represented as $\boldsymbol{F}_0 = \widehat{\boldsymbol{t}_1} \beta_1^T + \cdots + \widehat{\boldsymbol{t}_p} \beta_p^T + \mathrm{F}_p$. Substituting $t_h = w_{h1}^* X_1 + \cdots + w_{hn}^* X_n, (h = 1, 2, \cdots, p)$ to the regression equations vector $\mathbf{Y} = t_1 \beta_1 + \cdots + t_p \beta_p$, we can get *m* response surrogate functions: $Y_j = a_{j1} X_1 + \cdots + a_{jn} X_n, (j = 1, 2, \cdots, m)$. Where two requirements should be fulfilled: $\widehat{\boldsymbol{t}_h} = \boldsymbol{E}_0 \boldsymbol{w}_h^*$ and $\boldsymbol{w}_h^* = \prod_{i=1}^{h-1} \left(\boldsymbol{I} - \boldsymbol{w}_i \boldsymbol{\alpha}_i^T \right) \boldsymbol{w}_h$.

3.2. Cross-validation theory

A strict test of the predictive significance of each PLS latent variable is necessary, and then stopping when latent variables start to be non-significant. Cross-validation (CV) is a practical and reliable way to test this predictive significance [11, 35, 36]. Cross-validation method is used to determine whether the next latent variable is needed to be extracted. Assuming the current latent variable is $\boldsymbol{t}_h$. Then the theory of this method, including two predictive residual sums of squares (PRESS), is analyzed as follows:

(1) First type-PRESS

The *N* sample points are divided into two groups each time, including one with *N-1* sample points and the other with one sample point. *N* parallel regression model is developed from the reduced data with one row of the observation data deleted. After developing a model, differences between actual and predicted $\mathbf{Y}$ -values are calculated for the deleted data. The sum of squares of these differences is computed and collected from all parallel models to form the predictive residual sum of squares (PRESS), which estimates the predictive ability of the model. The PRESS of the j_{th} response is expressed as:

$$PRESS_j(h) = \sum_{i=1}^{N} \left(y_{ij} - \widehat{y_{(i)j}}(h) \right)^2, (j = 1, 2, \cdots, m) \tag{7}$$

The PRESS of $\mathbf{Y}=(Y_1,Y_2,\cdots,Y_m)^T$ can be defined as:

$$PRESS(h)=\sum_{j=1}^{m}PRESS_j(h), \tag{8}$$

(2) Second type-SS

All sample points are used to regress the response functions, and the difference between the actual and predicted $\mathbf{Y}$ -values for each point are calculated. The PRESS of Y_j corresponding to all the sample points is presented as:

$$SS_j(h)=\sum_{i=1}^{N}\left(y_{ij}-\widehat{y_{(i)j}}(h)\right)^2, \tag{9}$$

Then the PRESS of $\mathbf{Y}=(Y_1,Y_2,\cdots,Y_m)^T$ can be defined as:

$$SS(h)=\sum_{j=1}^{m}SS_j(h), \tag{10}$$

(3) Stopping condition

An error threshold of the stopping condition is defined by:

$$C_{thre}=PRESS(h)/SS(h-1). \tag{11}$$

Where $SS(h-1)$ denotes the residual sum of squares before the current latent variable.

The ratio is calculated after each latent variable, and a latent variable is judged significantly from the ratio which is smaller than around 0.9025 for at least one of the $\mathbf{Y}$ -variables. If C_{thre} is less than 0.9025, then the h latent variables are enough to provide an accurate regression model; otherwise, another latent variable is needed to be extracted in order to reach the accurate level. The process continues until a latent variable is not significant.

3.3. Design of experiment

Direct sampling methods (MCS for example) for reliability analysis by evaluating a large number of original response functions with high complexity and nonlinearity can be prohibitively expensive. Various importance sampling methods have been developed to reduce expense by focusing on samples in the important regions of the random variable space [19, 34, 41]. Another method of reducing cost is the use of surrogate models. Typically, a relatively small set of points are selected through DOE method and the true response is calculated at each sample point. These points are then used to construct an approximation with simple and low nonlinear form of the true response using some regression methods (PLSR is used in this paper).

In order to develop the response surface surrogate model, Latin hypercube sampling (LHS) [22] is used to generate a group of sample observations in this paper. By considering the sample number used in dimensional reduction method (DRM) [37] through selecting the same number of sample points along every axis and several experiments, the proper number of samples used to construct the surrogate model is $N=4\times\mathrm{n}$. Where n is the number of the random variables affecting all of the responses which determine the performance of the mechanical system. Assuming the mean value vector of the random variables is:

$$\boldsymbol{\mu}_{\mathbf{X}}=\left[\mu_{X_1},\mu_{X_2},\cdots,\mu_{X_\mathrm{n}}\right]^T=\left[\mu_1,\mu_2,\cdots,\mu_\mathrm{n}\right]^T, \tag{12}$$

and the deviation vector of the random variables is:

$$\boldsymbol{\sigma}_{\mathbf{X}}=\left[\sigma_{X_1},\sigma_{X2},\cdots,\sigma_{X_\mathrm{n}}\right]^T=\left[\sigma_1,\sigma_2,\cdots,\sigma_\mathrm{n}\right]^T, \tag{13}$$

Then the sampling space used in LHS method can be given by:

$$\mathbf{S}=\left[\boldsymbol{\mu}_{\mathbf{X}}-\mathrm{f}\boldsymbol{\sigma}_{\mathbf{X}},\boldsymbol{\mu}_{\mathbf{X}}+\mathrm{f}\boldsymbol{\sigma}_{\mathbf{X}}\right], \tag{14}$$

The proper one, f=4.5, is chosen by constructing many experiments with ranging from 3 to 6. With the selected coefficient to bounding the sampling space, a more accurate reliability probability value will be obtained.

3.4. Response surface based on PLSR

With the N sample points, the corresponding N response function values of each Y_j, which are obtained from the structural analysis method (FEM for example), are computed by:

$$\mathrm{y}_{jk}=g_{jk}(\boldsymbol{x}_k),(j=1,2,\cdots,m;k=1,2,\cdots,N), \tag{15}$$

Then the two data matrices, $\mathbf{x}$ and $\mathbf{y}$, are enough to develop the surrogate models.

RSM consisting of a group of mathematical and statistical techniques, is used in the development of an adequate functional relationship between a response of interest and a number of associated input variables. Without containing the cross-product powers of $X_1,X_2,\cdots,X_n$, the second-degree RSM model of the j_{th} component performance function is shown by:

$$Y_j^{RS}=a_{j0}+\sum_{i=1}^{n}a_{ji}X_i+\sum_{i=1}^{n}a_{jii}X_i^2,(j=1,2,\cdots,m), \tag{16}$$

Then nonlinearity of the original performance function can be explained. Substituting the X_i^2 with Z_i , the RSM model is then defined as:

$$Y_j^{RS}=a_{j0}+\sum_{i=1}^{n}a_{ji}X_i+\sum_{i=1}^{n}a_{jii}Z_i,(j=1,2,\cdots,m), \tag{17}$$

Therefore, combined with the two matrices consisting of the input variables sample data, $\left[\boldsymbol{I};\mathbf{x};\mathbf{x}^2\right]$ (in dimensions $N\times(2n+1)$) and the corresponding response data, $[\mathbf{y}]$ (in dimensions $N\times m$), the parameters of the model shown in Eq. (16-17) can be calculated accurately by the PLSR method. Where $\boldsymbol{I}$ is a $N\times1$ matrix filled with 1.

3.5. Corrected RSM based on PLSR

Here, as only second powers of the input variables are considered and the cross-products of powers of the input variables are neglected, errors may be produced by the surrogate model. Therefore, some methods are needed to improve the accuracy of the surrogate model. In the field of reliability analysis, almost all types of the random variables have the following characteristics as most of the data is distributed around the mean value and the more the data close to the mean value, so the larger probability data will be selected. The two characteristics show that the data around the mean value of the input variables has the larger impact on the reliability result of the component or system. Then if the surrogate model is accurate at the mean value of the input variables, the reliability result will be accurate. Based on this thought, if the surrogate model is improved at the mean value of the input random variables, the accuracy of the reliability result calculated by surrogate model will be improved. In the light of this, a new correction method is proposed. The performance functions values at the mean value of the input random variables are firstly calculated for both of the original response functions and the surrogate functions. The difference between two types of response values are then calculated and used as the correction coefficient to modify the meta-model. The procedures for this method are list as follows:

(1) Calculate "mean" response function values

The value of the component response function is evaluated as:

$$y_j^{\mathrm{A}\_mean} = g_j\left(\boldsymbol{\mu}_{\mathbf{X}}\right), \left(j = 1, \cdots, m\right), \tag{18}$$

and the corresponding value obtained by surrogate model is estimated as:

$$y_j^{R\_mean} = a_{j0} + \sum_{i=1}^{n} a_{ji}\mu_i + \sum_{i=1}^{n} a_{jii}\mu_i^2, (j = 1, 2, \cdots, m), \tag{19}$$

(2) Calculate the coefficients

The coefficients are then represented by:

$$cf_j = y_j^{A\_mean} - y_j^{R\_mean}, \left(j = 1, 2, \cdots, m\right), \tag{20}$$

Where cf_j represents the correction quantity of the surrogate model of the $\mathrm{j_{th}}$ component limit state function corresponding to the original exact one at the mean value of the random variables. $y_j^{A\_mean}$ represents the component response value calculated by the original limit state function, and $y_j^{R\_mean}$ represents the value calculated by the surrogate model.

(3) Correct RSM

With the coefficient cf_j, the surrogate model can be revised in the form of:

$$Y_j^{corr} = a_{j0} + \sum_{i=1}^{n} a_{ji}X_i + \sum_{i=1}^{n} a_{jii}X_i^2 + cf_j, (j = 1, 2, \cdots, m). \tag{21}$$

Based on the corrected method, an accurate surrogate model will be obtained and the system reliability can be calculated more accurately. Combined with the MCS, the model will be used to simulate the system reliability.

4. MCS used in the reliability analysis of a dependent system

Assuming $Y_j > 0$ represents that the component is working well. For a series system, reliability of the system indicates that all components of the system work well. When using the component performance functions to express the system reliability, the performance function of the system can be defined as:

$$G_{series} = \min\left[Y_1, Y_2, \cdots, Y_m\right], \tag{22}$$

and the system reliability is defined as:

$$pRs = \Pr\left\{G_{series} > 0\right\}, \tag{23}$$

For a parallel system, the system is reliable if any of the components works well. Then the performance function of the system can be computed by:

$$G_{parrallel} = \max\left[Y_1, Y_2, \cdots, Y_m\right], \tag{24}$$

and reliability of the parallel system is evaluated by:

$$pRs = \Pr\left\{G_{parrallel} > 0\right\}, \tag{25}$$

Where all of the component functions $Y_1, Y_2, \cdots, Y_m$ are affected by the same input variables $X_1, X_2, \cdots, X_n$, which indicates the dependency of the system failure.

Then Monte Carlo method estimating $P_{R,S}$ and $P_{R,p}$ of the reliabilities of the series system and parallel system, respectively, are expressed as:

$$P_{R,S} = \frac{1}{N_S}\sum_{k=1}^{N_S} \ell\left[G_{series}^k > 0\right], \tag{26}$$

$$P_{R,P} = \frac{1}{N_S}\sum_{k=1}^{N_S} \ell\left[G_{parrallel}^k > 0\right], \tag{27}$$

Where G^k is the k_{th} realization of G, N_S is the sample size, and $\ell[\cdot]$ is an indicator function such that G^k is in the reliable set (i.e. when $G^k > 0$) and zero otherwise.

Since the proposed method facilitates explicit lower-dimensional approximation of a general multivariate function, the embedded MCS can be conducted for any sample size. The accuracy and efficiency of the reliability calculations using the developed method will be discussed in section 5.

5. Numerical examples

Three methods, including the Partial Least Squares Regression-Response Surface Method (PLSRRSM), Corrected-Partial Least Squares Regression-Response Surface Method (C-PLSRRSM), and direct MCS (D-MCS) with 10^6 samples, are discussed to analyze the system reliability with dependency. Accuracy of the proposed method is verified by three numerical examples. The system reliability calculated with the original performance functions by MCS is used as the

benchmark data. When comparing computational efforts, the number of original performance function evaluations is chosen as the primary metric in this paper. For the direct MCS, the number of original function evaluation is same as the sample size. However, the MCS (although with the same sample size as the direct MCS) embedded in the proposed method is conducted by using their response surface approximations. The difference in CPU times in evaluating an original function and its response surface approximation is significant when a calculation of the original function involves in expensive finite-element or mesh-free analysis.

5.1. Example 1—A ten-bar truss structural system

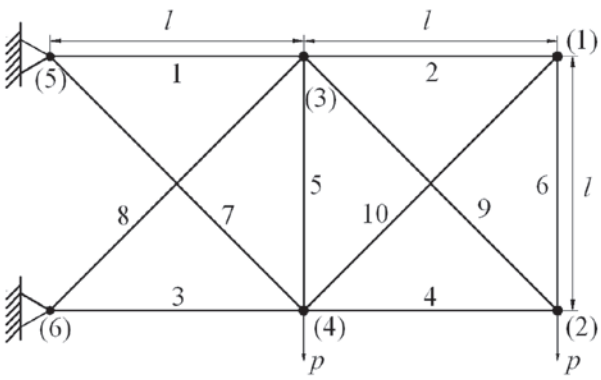


Fig. 1. A ten-bar truss structure with random cross-sectional areas

A ten-bar, linear-elastic, truss structure, shown in Fig. 1, was studied to examine the accuracy and efficiency of the proposed system reliability analysis method. Two concentrated forces are applied at nodes 2 and 4. In order to build the limit state function of the structural system, three failure modes of the system analyzed by Huang [17] are shown: The stress failure of bar 3 indicates that the stress applied on bar 3 is larger than that of the allowable stress, is expressed as:

$$g_1 = \frac{pl}{A_1 A_3 e d_{allow}} \cdot \frac{4\sqrt{2}A_1^3\left(24A_2^2 + A_3^2\right) + A_3^3\left(7A_1^2 + 26A_2^2\right) + 4A_1A_2A_3\left[\left(20A_1^2 + 76A_1A_2 + 10A_3^2\right) + \sqrt{2}A_3\left(25A_1 + 29A_2\right)\right]}{4A_2^2\left(8A_1^2 + A_3^2\right) + 4\sqrt{2}A_1A_2A_3\left(3A_1 + 4A_2\right) + A_1A_3^2\left(A_1 + 6A_2\right)} - 1; \tag{28}$$

The stress failure of bar 7, where the stress applied is larger than the allowable stress, is given by:

$$g_2 = \frac{p}{A_1\sigma_{allow}}\left[\frac{A_1A_2A_3\left(2\sqrt{2}A_1 + A_3\right)}{4A_2^2\left(8A_1^2 + A_3^2\right) + 4\sqrt{2}A_1A_2A_3\left(3A_1 + 4A_2\right) + A_1A_3^2\left(A_1 + 6A_2\right)} + 2\right] - 1; \tag{29}$$

And the displacement failure of the node 2, demonstrating that the maximum displacement occurred at node 2 exceeds to the allowable one, is presented by:

$$g_3 = \frac{p}{A_3\sigma_{allow}}\left[\frac{\sqrt{2}A_1A_2A_3\left(2\sqrt{2}A_1 + A_3\right)}{4A_2^2\left(8A_1^2 + A_3^2\right) + 4\sqrt{2}A_1A_2A_3\left(3A_1 + 4A_2\right) + A_1A_3^2\left(A_1 + 6A_2\right)} + \sqrt{2}\right] - 1. \tag{30}$$

Then the system limit state function of the ten-bar structure is given by:

$$G_s = min\left[g_1, g_2, g_3\right]; \tag{31}$$

The system which composed of three failure modes with the corresponding three component limit state functions is a series system. And this system is used to demonstrate the accuracy and efficiency of the proposed method. Properties of the input random variables, denoted as $X_1 - X_{11}$, are list in Table.1. All variables are normally distributed.

Table 1. Distribution details of input random variables

Variable	Description	Distribution	Mean	Standard deviation
X_1	A_1 (cm^2)	Normal	13	1
X_2	A_2 (cm^2)	Normal	2	0.5
X_3	A_3 (cm^2)	Normal	9	0.5
X_4	p (Kg)	Normal	10^4	500
X_5	l (cm)	Normal	360	1.2
X_6	e (GPa)	Normal	100	5
X_7	σ_{allow} (GPa)	Normal	2	0.2
X_8	d_{allow} (cm)	Normal	4.25	0.2

The reliability results, from PLSRRSM, C-PLSRRSM, and D-MCS, corresponding to the change of number of sample points used to develop the surrogate model, are given in Fig. 2. When the number of sample points are $4 \times n = 4 \times 8 = 32$, reliabilities of the ten-bar truss structure system obtained by PLSRRSM, C-PLSRRSM, and D-MCS, are 0.5767,0.9443 and 0.9315, respectively. The probability of reliability calculated by D-MCS is selected as the benchmark, and then the percentages of reliability result errors from PLSRRSM and C-PLSRRSM are 38.1% and 1.37%, respectively. It is shown that the accuracy of the reliability given by the correction model is improved by 36.74%. Moreover, when the number of sample points are more than 32, the accuracy of the reliability probabilities estimated by the two proposed methods stand still with the increase of the sample points. In other words, the accuracy of the

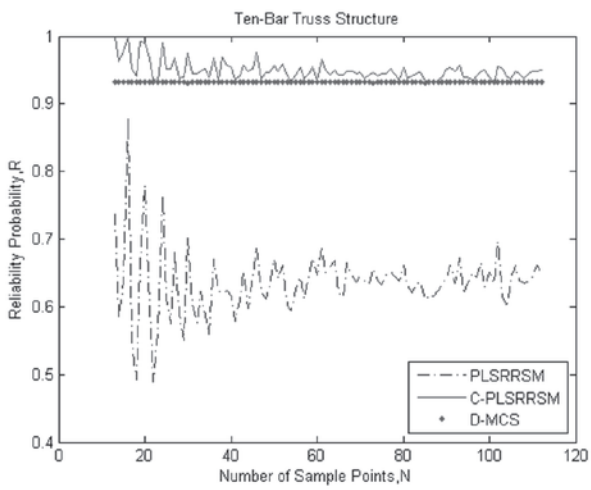


Fig. 2. Reliability of the structure corresponding to the sample points

proposed method cannot be improved by giving a large number of sample points. From the discussion, it is verified that the proposed correction model requires only a few original function evaluations to generate an accurate result.

5.2. Example 2—A cantilever beam system

This second test problem involves the reliability analysis of a cantilever beam as shown in Fig. 3. Two external forces F_1 and F_2, two external moments M_1 and M_2, and external distributed loads represented by (q_{L1}, q_{R1}) and (q_{L2}, q_{R2}), are applied on the cantilever beam. A total of twenty-one random variables, such as, dimensions, the yield strength S, the maximum allowable shear stress $\tau_{\max}$, are involved in this example, as shown in Table. 2. The system limit state function composed of three component limit sate functions [7] will be used to describe the accuracy of the proposed method.

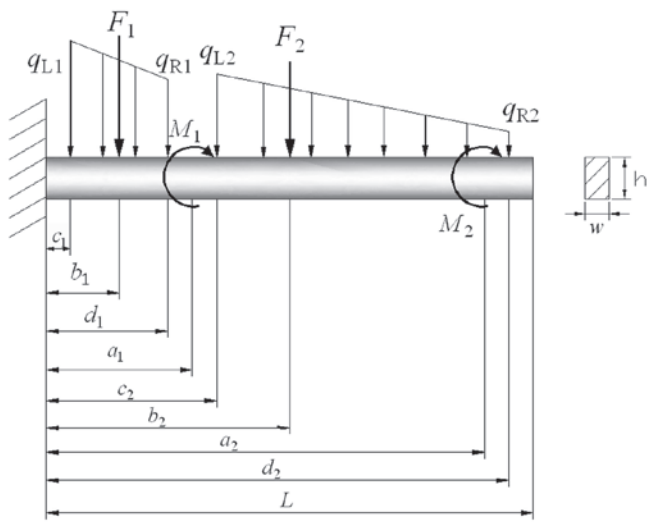


Fig. 3. Cantilever beam

The system limit state function consists of: the first component limit state function, representing the difference between the maximum normal stress and the yield strength S, is given by:

$$g_1 = S - 6\frac{M}{wh^2} = S - 6\frac{\sum_{i=1}^{2} M_i + \sum_{i=1}^{2} F_i b_i + \sum_{i=1}^{2} q_{Li}(d_i - c_i)(d_i + c_i)/2 + \sum_{i=1}^{2}\left[(q_{Ri} - q_{Li})(d_i - c_i)/2\right]\left[c_i + 2(d_i - c_i)/3\right]}{wh^2}. \tag{32}$$

The second component limit state function expresses that the deflection v_{tip} of the tip of the beam should be less than the allowable deflection $v_{\max}$ and is defined as:

$$g_2 = v_{max} - v_{tip} = v_{max} - \left\{\frac{1}{EI}\left[-\sum_{i=1}^{2}\frac{q_{Li}(L-c_i)^4}{24} - \sum_{i=1}^{2}\frac{(q_{Ri}-q_{Li})(L-c_i)^5}{120(d_i-c_i)} + \sum_{i=1}^{2}\frac{q_{Ri}(L-d_i)^4}{24}\right] + \frac{1}{EI}\left[\frac{ML^2}{2} + \frac{RL^3}{6} + \sum_{i=1}^{2}\frac{M_i(L-a_i)^2}{2} - \sum_{i=1}^{2}\frac{F_i(L-b_i)^3}{6}\right] + \frac{1}{EI}\sum_{i=1}^{2}\frac{(q_{Ri}-q_{Li})(L-d_i)^5}{120(d_i-c_i)}\right\}. \tag{33}$$

Table 2. Distributions of random variables

variable	description	Mean value	Standard deviation	Distribution type
X_1	M_1 (Nm)	50.0×10^3	5.0×10^3	Normal
X_2	M_2 (Nm)	30.0×10^3	3.0×10^3	Normal
X_3	F_1 (N)	18.0×10^3	4.0×10^3	Extreme value type I
X_4	F_2 (N)	30.0×10^3	3.0×10^3	Normal
X_5	q_{L1} (N/m)	30.0×10^3	1.0×10^3	Normal
X_6	q_{R1} (N/m)	20.0×10^3	1.0×10^3	Normal
X_7	q_{L2} (N/m)	20.0×10^3	1.0×10^3	Normal
X_8	q_{R2} (N/m)	1.0×10^3	10	Normal
X_9	a_1 (m)	1.5	0.005	Normal
X_{10}	a_2 (m)	4.5	0.005	Normal
X_{11}	b_1 (m)	0.75	0.001	Normal
X_{12}	b_2 (m)	2.5	0.001	Normal
X_{13}	c_1 (m)	0.25	0.0005	Normal
X_{14}	c_2 (m)	1.75	0.001	Normal
X_{15}	d_1 (m)	1.25	0.001	Normal
X_{16}	d_2 (m)	4.75	0.001	Normal
X_{17}	L (m)	5	0.01	Normal
X_{18}	w (m)	0.2	0.0001	Normal
X_{19}	h (m)	0.4	0.0001	Normal
X_{20}	S (Pa)	80.0×10^6	8.0×10^6	Normal
X_{21}	$\tau_{\max}$ (Pa)	3.5×10^6	0.5×10^6	Normal

Where R is the reaction force at the fixed end; The Young's modulus E is 200×10^9 Pa; the moment of inertia is $I = (1/12)\omega h^3$, and the allowable deflection is v_{max}=0.025 .

The third limit state function is given by:

$$g_3 = \tau_{\max} - \tau = \tau_{\max} - \frac{3}{2wh}\left\{\sum_{i=1}^{2} F_i + \sum_{i=1}^{2} q_{Li}(d_i - c_i) + \sum_{i=1}^{2}\frac{(q_{Ri}-q_{Li})(d_i-c_i)}{2}\right\}. \tag{34}$$

Where τ is the shear stress at root, and the term in curly brackets is the shear force at the root.

The system limit state function is defined as $G_s = min[g_1, g_2, g_3]$, and the system reliability is given as $pRs = \Pr\{G_S > 0\}$.

With the increase of the sample points, Fig. 4 presents reliability probabilities of the cantilever beam, predicted by PLSRRSM and C-PLSRRSM, as well as by D-MCS. The reliability probabilities from PLSRRSM, C-PLSRRSM and D-MCS, when the number of sample points is $4 \times n = 4 \times 21 = 84$, are 0.967, 0.9542 and 0.9537, respectively. The absolute error percentages of the PLSRRSM and C-PLSRRSM to D-MCS are 1.39% and 0.052%, respectively. Therefore, the probabilities calculated by both of the two proposed method are accurate, and the results obtained by C-PLSRRSM closes to the benchmark almost without error. As can be seen in Fig. 4, when the number of the sample points is more than 84, both the PLSRRSM and C-PLSRRSM provide stable reliability results with small fluctuations. The same conclusion is derived from the results that C-PLSRRSM is more accurate than PLSRRSM. The effectiveness of the proposed method is also demonstrated by the example.

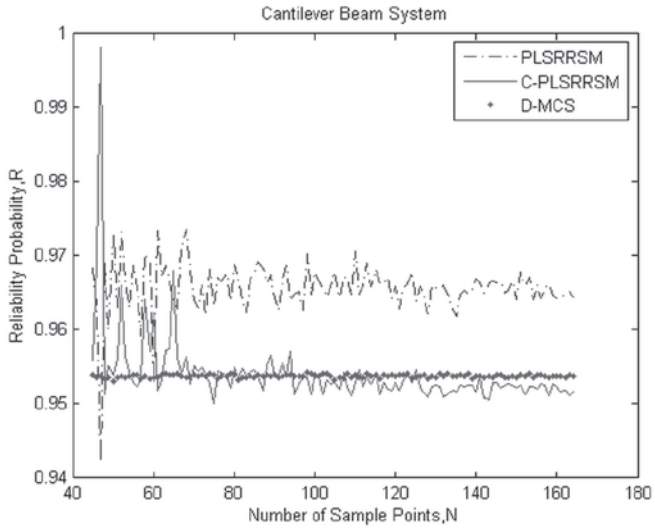


Fig. 4. Reliability of cantilever beam vs. sample points

5.3. Example 3—Vehicle side impact

The final test problem investigates the side impact crash-worthiness of a vehicle subjecting to variations in the sizes and material properties of several key components. This problem has been investigated by many researchers in the fields of reliability based design optimization and robust design optimization. However, the reliability of each component is treated separately without considering their failure dependency. Actually, all failure modes are being the potential failure mode and strong dependencies are contained between them. When any of the components fails, the entire vehicle as the series system is said to be failed. The limit state system function constructed by Bichon [3] with ten failure modes will be used to test the proposed method.

Ten equations corresponding to failure modes are considered:

the abdomen load

$$L = 1.16 - 0.3717X_2X_4 - 0.00931X_2X_{10} - 0.484X_3X_9 + 0.01343X_6X_{10}\ ; \quad (35)$$

the pubic symphysis force

$$F = 4.72 - 0.5X_4 - 0.19X_2X_3 - 0.0122X_4X_{10} + 0.009325X_6X_{10} + 0.000191X_{11}^2\ ; \quad (36)$$

the rib deflections at upper

$$D_u = 28.98 + 3.818X_3 - 4.2X_1X_2 + 0.0207X_5X_{10} + 6.63X_6X_9 - 7.7X_7X_8 + 0.32X_9X_{10}\ ; \quad (37)$$

the rib deflections at middle

$$D_m = 33.86 + 2.95X_3 + 0.1792X_{10} - 5.057X_1X_2 - 11.0X_2X_8 - 0.0215X_5X_{10} - 9.98X_7X_8 + 22.0X_8X_9; \quad (38)$$

the rib deflections at lower

$$D_l = 46.36 - 9.9X_2 - 12.9X_1X_8 + 0.1107X_3X_{10}\ ; \quad (39)$$

the viscous criteria at upper

$$VC_u = 0.261 - 0.0159X_1X_2 - 0.188X_1X_8 - 0.019X_2X_7 + 0.0144X_3X_5 + 0.0008757X_5X_{10} + 0.08045X_6X_9 + 0.00139X_8X_{11} + 0.00001575X_{10}X_{11}\ ; \quad (40)$$

the viscous criteria at middle

$$VC_m = 0.214 + 0.00817X_5 - 0.131X_1X_8 - 0.0704X_1X_9 + 0.03099X_2X_6 - 0.018X_2X_7 + 0.0208X_3X_8 + 0.121X_3X_9 - 0.00364X_5X_6 + 0.0007715X_5X_{10} - 0.0005354X_6X_{10} + 0.00121X_8X_{11}\ ; \quad (41)$$

the viscous criteria at lower

$$VC_l = 0.74 - 0.61X_2 - 0.163X_3X_8 + 0.001232X_3X_{10} - 0.166X_7X_9 + 0.227X_2^2\ ; \quad (42)$$

the velocity the B-pillar

$$V_B = 10.58 - 0.674X_1X_2 - 1.95X_2X_8 + 0.02054X_3X_{10} - 0.0198X_4X_{10} + 0.028X_6X_{10}; \quad (43)$$

the velocity at the door

$$V_D = 16.45 - 0.489X_3X_7 - 0.843X_5X_6 + 0.0432X_9X_{10} - 0.0556X_9X_{11} - 0.000786X_{11}^2; \quad (44)$$

Combined with the corresponding allowable values, then the system limit state function of the vehicle side impact problem is defined as:

$$G_s = min[g_1 = 1.0 - L; g_2 = 4.01 - F; g_3 = 32.0 - D_u; g_4 = 32.0 - D_m; g_5 = 32.0 - D_l; g_6 = 0.32 - VC_u; g_7 = 0.32 - VC_m; g_8 = 0.32 - VC_l; g_9 = 9.9 - V_B; g_{10} = 15.69 - V_D], \quad (45)$$

and the system reliability can be expressed by:

$$pRs = \Pr\{G_S > 0\}. \quad (46)$$

The distribution information of the random variables, denoted as X_1-X_{11}, involving the thickness and material properties of critical structures in the vehicle and the location of the impact, are described in Table. 3.

Table 3. Random variables in the vehicle side impact problem

Variable	Description	Distri-bution	Mean	Std.dev
Member thickness (mm)				
X_1	B-pillar inner	Normal	0.500	0.030
X_2	B-pillar reinforcement	Normal	1.310	0.030
X_3	Floor side inner	Normal	0.500	0.030
X_4	Cross members	Normal	1.395	0.030
X_5	Door beam	Normal	0.875	0.030
X_6	Door belt line rein-forcement	Normal	1.200	0.030
X_7	Roof rail	Normal	0.400	0.030
Material properties (GPa)				
X_8	B-pillar inner	Normal	0.345	0.006
X_9	Floor side inner	Normal	0.192	0.006
Deviation of impact location (mm)				
X_{10}	Barrier height		0.0	10.0
X_{11}	Barrier hitting posi-tion		0.0	10.0

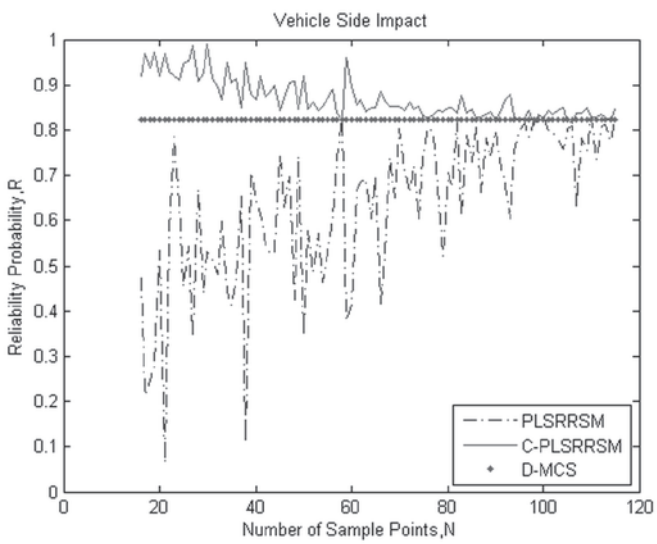


Fig. 5. Reliability of vehicle vs. sample points

The reliability was calculated by using the proposed method and compared with the D-MCS, as shown in Fig. 5. The reliability with the increase of the sample points are also described in Fig. 5. Corresponding to the number of sample points, $4\times n=4\times 11=44$, the calculated reliability results of the vehicle are 0.747, 0.8441 and 0.8231 for PLSRRSM, C-PLSRRSM, D-MCS respectively. The error percentages of the proposed method to D-MCS are 9.25% and 2.55%.

Different from the above examples, the accuracy of the results improved by the corrected model is not significant and the reliability results, computed by PLSRRSM and C-PLSRRSM are close to each other, when the number of the sample points is more than 82. This may be due to the fact that the original functions of the vehicle system are derived from RSM. Nevertheless, the accuracy and efficiency of the proposed method is obvious.

6. Conclusion

Based on the ability of the PLSR to analyze the dependent relationship between the same input variables and the corresponding different responses, a new response surface modeling method for the structural or mechanical system was developed. To improve the accuracy of the surrogate model, a correction method by adding a coefficient to each component meta-model of the system surrogate model was presented. The coefficient is defined as the difference between the exact response value and the surrogate one at the mean values of the input variables. Then the corrected surrogate model combined with MCS method was used to analyze the system reliability, named as Corrected-PLSR-RSM based system reliability analysis (C-PLSRRSM-SRA). As to the sampling method to build response surface model, LHS is selected and a proper coefficient $f=4.5$ to bound the sampling region of the input random variables was chosen.

Due to a small number of original function evaluations, the proposed method is effective, particularly when a response evaluation entails costly finite-element, mesh-free, or other numerical analysis, whose limit state function is implicit. By using the surrogate model, it is also an effective way to solve the problem composed of the complex and high nonlinear explicit limit state functions, which saves computational expense explicitly. The numerical examples tested in the paper indicate that the proposed method provides accurate and computationally efficient estimates of reliability. As Compared to the PLSRSM method, the C-PLSRRSM method makes considerable improvements from the perspective of accuracy, efficiency, and stability, with only one more time of calculating the original limit state functions. The C-PLSRRSM method could be more accurate to solve the reliability of the system with highly nonlinear limit state function composed of several component limit sate functions involving a large number of input variables, and provide a moderate accurate value of the system reliability which fulfils the requirements of engineering applications. With the proposed sampling method, another advantage of this new method is that a more accurate surrogate model can be built with the least number of sample points.

However, the C-PLSRRSM method needs a large number of original function evaluations to get the surrogate model when the number of the input random variables is large, which will decrease the efficiency of proposed methods. In addition, the C-PLSRRSM method may producing an error for large probability levels (e.g., more than 99.9%). Because the response surface model is regressed without mixed terms, the surrogate model cannot represent the original performance functions in the whole distribution region of the input random variables. To deal with these issues, developing a more accurate correction method will be a future work.

Acknowledgement

The support from the High-End Talents Recruitment Program of Liaoning Province of China is gratefully acknowledged. The authors would like to thank the anonymous referees and the editor valuable comments and suggestions leading to an improvement of this article.

References

1. Au S K, Beck J L. Estimation of small failure probabilities in high dimensions by subset simulation. Probabilistic Engineering Mechanics 2001; 16(4): 263-277, http://dx.doi.org/10.1016/S0266-8920(01)00019-4.
2. Bichon B J, Eldred M S, Swiler L P, Mahadevan S, McFarland J M. Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions. Aiaa Journal 2008; 46(10): 2459-2468, http://dx.doi.org/10.2514/1.34321.
3. Bichon B J, McFarland J M, Mahadevan S. Efficient surrogate models for reliability analysis of systems with multiple failure modes. Reliability Engineering & System Safety 2011; 96(10): 1386-1395, http://dx.doi.org/10.1016/j.ress.2011.05.008.
4. Cai G Q, Elishakoff I. Refined second-order reliability analysis. Structral safety 1994; 14(4): 267-276, http://dx.doi.org/10.1016/0167-4730(94)90015-9.
5. Campos-Requena V H, Rivas B L, Perez M A, Wilhelm M. Application of design of experiments, response surface methodology and partial least squares regression on nanocomposites synthesis. Polymer Bulletin 2014; 71(8): 1961-1982, http://dx.doi.org/10.1007/s00289-014-1166-6.
6. Ditlevsen O. Narrow Reliability Bounds for Structural Systems. Journal of Structural Mechanics 1979; 7(4): 453-472, http://dx.doi.org/10.1080/03601217908905329.
7. Du X P. System reliability analysis with saddle point approximation. Structural and Multidisciplinary Optimization 2010; 42(2): 193-208, http://dx.doi.org/10.1007/s00158-009-0478-x.
8. Du X P, Sudjianto A. First-order saddle point approximation for reliability analysis. Aiaa Journal 2004; 42(6): 1199-1207, http://dx.doi.org/10.2514/1.3877.
9. Echard B, Gayton N, Lemaire M. AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. Structural Safety 2011; 33(2): 145-154, http://dx.doi.org/10.1016/j.strusafe.2011.01.002.
10. Faravelli L. Response-Surface Approach for Reliability-Analysis. Journal of Engineering Mechanics-Asce 1989; 115(12): 2763-2781, http://dx.doi.org/10.1061/(ASCE)0733-9399(1989)115:12(2763).
11. Garthwaite P H. An Interpretation of Partial Least-Squares. Journal of the American Statistical Association 1994; 89(425): 122-127, http://dx.doi.org/10.1080/01621459.1994.10476452.
12. Gayton N, Bourinet J M, Lemaire M. CQ2RS: a new statistical approach to the response surface method for reliability analysis. Structural Safety 2003; 25(1): 99-121, http://dx.doi.org/10.1016/S0167-4730(02)00045-0.
13. Geladi P, Kowalski B R. Partial Least-Squares Regression - a Tutorial. Analytica Chimica Acta 1986; 185: 1-17, http://dx.doi.org/10.1016/0003-2670(86)80028-9.
14. Gupta S, Manohar C S. An improved response surface method for the determination of failure probability and importance measures. Structural Safety 2004; 26(2): 123-139, http://dx.doi.org/10.1016/S0167-4730(03)00021-3.
15. Hohenbichler M, Gollwitzer S, Kruse W, Rackwitz R. New light on first-and second-order reliability methods. Structural safety 1987; 4(4): 267-284, http://dx.doi.org/10.1016/0167-4730(87)90002-6.
16. Huang B Q, Du X P. Probabilistic uncertainty analysis by mean-value first order Saddle point Approximation. Reliability Engineering & System Safety 2008; 93(2): 325-336, http://dx.doi.org/10.1016/j.ress.2006.10.021.
17. Huang X Z, Zhang Y M, Wu M C. Research on the method for reliability analysis of structural systems based on the dimensional-reduction integration. Chinese journal of theoretical and applied mechanics 2013; 45(3): 456-460.
18. Huang Z, Turner B J, Dury S J, Wallis I R, Foley W J. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. Remote Sensing of Environment 2004; 93(1-2): 18-29, http://dx.doi.org/10.1016/j.rse.2004.06.008.
19. Mahadevan S, Raghothamachar P. Adaptive simulation for system reliability analysis of large structures. Computers & Structures 2000; 77(6): 725-734, http://dx.doi.org/10.1016/S0045-7949(00)00013-4.
20. Malthouse E C, Tamhane A C, Mah R S H. Nonlinear partial least squares. Computers & Chemical Engineering 1997; 21(8): 875-890, http://dx.doi.org/10.1016/S0098-1354(96)00311-0.
21. McDonald M, Mahadevan S. Design optimization with system-level reliability constraints. Journal of Mechanical Design 2008; 130(2): 1-10, http://dx.doi.org/10.1115/1.2813782.
22. Mckay M D, Beckman R J, Conover W J. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 1979; 21(2): 239-245.
23. Melchers R E. Importance Sampling in Structural Systems. Structural Safety 1989; 6(1): 3-10, http://dx.doi.org/10.1016/0167-4730(89)90003-9.
24. Mutingi M. System reliability optimization: a fuzzy multi-objective genetic algorithm approach. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2014; 16(3): 400-406.
25. Nie J, Ellingwood B R. Directional methods for structural reliability analysis. Structural Safety 2000; 22(3): 233-249, http://dx.doi.org/10.1016/S0167-4730(00)00014-X.
26. Niederreiter H, Spanier J. Monte Carlo and Quasi-Monte Carlo Methods. Berlin: Springer-Verlag, 2000.
27. Park J, Towashiraporn P. Rapid seismic damage assessment of railway bridges using the response-surface statistical model. Structural Safety 2014; 47: 1-12, http://dx.doi.org/10.1016/j.strusafe.2013.10.001.
28. Rackwitz R. Reliability analysis - a review and some perspectives. Structural Safety 2001; 23(4): 365-395, http://dx.doi.org/10.1016/S0167-4730(02)00009-7.
29. Rausand M, Hoyland A. System reliability theory: models, statistical methods, and application. 2nd edn. ed. New York: Wiley-interscience, 2000.
30. Schmidtlein S, Sassin J. Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. Remote Sensing of Environment 2004; 92(1): 126-138, http://dx.doi.org/10.1016/j.rse.2004.05.004.
31. Shi L, Yang R J, Zhu P. An Adaptive Response Surface Method Using Bayesian Metric and Model Bias Correction Function. Journal of Mechanical Design 2014; 136(3), http://dx.doi.org/10.1115/1.4026095.

32. Shieh M D, Yeh Y E. Developing a design support system for the exterior form of running shoes using partial least squares and neural networks. Computers & Industrial Engineering 2013; 65(4): 704-718, http://dx.doi.org/10.1016/j.cie.2013.05.008.
33. Song J H, Kiureghian A D. Bounds on system reliability by linear programming. Journal of Engineering Mechanics-Asce 2003; 129(6): 627-636, http://dx.doi.org/10.1061/(ASCE)0733-9399(2003)129:6(627).
34. Szybka J, Broniec Z, Pilch R. Forecasting the failure of a thermal pipeline on the basis of risk assessment and exploitation analysis. Eksploatacja i Niezawodnosc -Maintenance and Reliability 2011; (4): 5-10.
35. Wang H. Partial least squares regression method and application. Beijing: National defense industry press, 1999.
36. Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 2001; 58(2): 109-130, http://dx.doi.org/10.1016/S0169-7439(01)00155-1.
37. Won J, Choi C, Choi J. Improved dimension reduction method (DRM) in uncertainty analysis using kriging interpolation. Journal of Mechanical Science and Technology 2009; 23(5): 1249-1260, http://dx.doi.org/10.1007/s12206-008-0721-1.
38. Youn B D, Wang P F. Complementary Intersection Method for System Reliability Analysis. Journal of Mechanical Design 2009; 131(4): 1-15, http://dx.doi.org/10.1115/1.3086794.
39. Zhao W, Wang W. Application of non-linear partial least squares regression method to response surface method with uniform design. Acta aeronautica et astronautica sinica 2012; 33(5): 839-847.
40. Zhao W, Wang W. Application of partial least squares regression in response surface for analysis of structural reliability. Engineering mechanics 2013; 30(2): 272-277.
41. Zou T, Mahadevan S, Mourelatos Z, Meernik P. Reliability analysis of automotive body-door subsystem. Reliability Engineering & System Safety 2002; 78(3): 315-324, http://dx.doi.org/10.1016/S0951-8320(02)00178-3.

Huahan LIU
Wei JIANG
Zahid Hussain HULIO
Qiuzhi WANG

School of Mechanical Engineering
Dalian University of Technology
Linggong Road, No 2
Ganjingzi District, Dalian, Liaoning Province, China

E-mails: liuhh1987@126.com, jiangwei@dlut.edu.cn, hussainafrasiyab@gmail.com, 875961226@qq.com

Article citation info:
JAWORSKI J, KLUZ R, TRZEPIECIŃSKI T. Operational tests of wear dynamics of drills made of low-alloy high-speed HS2-5-1 steel. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 271–277, http://dx.doi.org/10.17531/ein.2016.2.15.

Jan JAWORSKI
Rafał KLUZ
Tomasz TRZEPIECIŃSKI

OPERATIONAL TESTS OF WEAR DYNAMICS OF DRILLS MADE OF LOW-ALLOY HIGH-SPEED HS2-5-1 STEEL

BADANIA EKSPLOATACYJNE DYNAMIKI ZUŻYCIA WIERTEŁ Z NISKOSTOPOWEJ STALI SZYBKOTNĄCEJ HS2-5-1*

To determine the effect of drill wear on the value of the axial force and cutting torque, a series of durability tests of drills with a diameter of 10 mm made of high-speed steel HS2-5-1 were carried out. The investigations were conducted during the durability period and at constant values of cutting parameters. The tests were carried out while drilling holes in samples made of C45 steel and EN-GJS-500-7 cast iron. The dynamics of wear on all parts of the drill was also determined. It has been shown that while drilling with different values of cutting parameters, there is a loss of machinability for different values of the wear indicators. While drilling with high cutting speeds and with small feeds, there is a loss of cutting ability in the area of accelerated wear. The application of TiN coating does not change the controlled wear locations. TiN coating only reduces the intensity of wear on the tool flank, which increases the durability of the drill.

***Keywords**: wear dynamics, cutting torque, high-speed steel, drilling, wear.*

W celu określenia wpływu zużycia wierteł na wartość siły osiowej oraz momentu skrawania przeprowadzono serię badań trwałościowych wierteł o średnicy 10 mm wykonanych ze stali szybkotnącej HS2-5-1, przy stałych parametrach skrawania, w czasie jednego przyjętego okresu trwałości. Badania prowadzono podczas obróbki otworów na próbkach ze stali C45 oraz z żeliwa EN-GJS-500-7. Określono również dynamikę zużycia na wszystkich częściach skrawających wiertła. Wykazano, że podczas eksploatacji wierteł z różnymi parametrami skrawania, utrata ich skrawności następuje dla różnych wartości wskaźników zużycia. Podczas wiercenia z dużymi prędkościami skrawania i małymi posuwami, utrata skrawności wiertła następuje w obszarze przyspieszonego ich zużycia. Naniesienie powłoki TiN nie zmienia kontrolowanych miejsc zużycia a tylko zmniejsza intensywność zużycia na powierzchni przyłożenia, co powoduje wzrost trwałości wiertła.

***Słowa kluczowe**: dynamika zużycia, moment skrawania, stal szybkotnąca, wiercenie, zużycie.*

1. Introduction

A cutting tool with certain parameters of the initial state starts working at the parameters for which it was designed. With time, in the process of cutting under the influence of mechanical and thermodynamic loads the geometric parameters of the blade change in value. Tool life is characterized by the period of its operation, taking into account the need for sharpening to ensure the state of availability for the job [1, 7]. All of the quantitative factors of tool life (e.g., probability of reliable operation of the tool, durability, average durability, the density function of durability) can be determined only experimentally on the finished new tool which is characterized by certain indicators of the input state, and based on the statistical observations of tools in operation.

Tribological processes that occur at the contact point of the cutting blade with the workpiece lead to wear and then to a sudden or gradual loss of the tool's cutting ability [18, 24, 26]. Among the symptoms of wear of a blade we can list [5, 11, 23, 29]: changes in the blade geometry, changes in physico-chemical properties due to chemical changes in the surface layer of the material, cracks and chipping of the material of the cutting tool.

Changes in the blade stereometry connected with its wear affect the drilling process [14]. The effect of drill wear on the accuracy of drilling of deep holes was investigated by Wieczorowski and Matuszak [28]. Catastrophic tool wear in tools made of high-speed steel is associated with an increase in the temperature of the cutting part of the tool to a value which causes changes in the basic properties of the tool [5, 29]. An analysis of the results carried out by Meena and El Mansori [21] showed that high-speed machining of cast iron combined with low feed values causes an increase of the value of cutting forces and the energy of the cut. For a drill with a steady specificity of wear, the temperature can be increased only by the increase of the heat source on the primary tool flanks. From these heat sources in a stage of steady wear, the cutting part of the drill and the drill margin flanks heat up [4]. Under production and laboratory conditions, the characteristic sound of „scratch" is assumed as the beginning of the catastrophic wear [32]. This signal is associated only with tribological changes in the contact surface of the drill and the workpiece. The beginning of this process reflects an increase in the amplitude of the acoustic signal and an increase in friction torque generated at margins of the drill [6]. The degree of wear of the blade is assessed using multiple criteria, which can be divided into four groups [27]: technological, physical, economic and geometric.

According to Pancielejko [22], when determining the wear of an uncoated drill and drills with hard coatings, it is desirable to use a number of wear indicators. Operational tests of twist drills made of high-speed steel HS6-5-2, uncoated and titanium carbonitride Ti (C, N) coated showed that the applied wear factors characterizing drill wear at the corners, at the drill margin and at the chisel edge of a drill characterize well the wear of tested drills. Studies by Liu et al. [20]

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

showed that the most important criterion of wear of the blade should be considered the wear in the corner of the drill.

To ensure suitable durability of drills it is necessary to examine the tool state in specific conditions of its operation and to determine which changed wear factor can be used as a criterion for its evaluation. To determine the effect of wear of the drill during operation on the value of the thrust force and cutting moment, a series of durability tests of drills while drilling holes in samples made of C45 steel and cast iron EN-GJS-500-7 were carried out. The assessment of the dynamics of wear is made according to the maximum linear wear value on all cutting parts of the drill.

2. Justification of the study

In order to increase the efficiency and reliability of machining processes it is possible to select cutting parameters for which, in the conditions of normal wear, the required probability of meeting all the quality requirements of the workpiece is ensured [25]. The reliability of the cutting process is conditioned by the period of assumed tool life, cutting forces, and the quality and precision of machined parts [15, 30].

In the cutting process, with the passage of time under thermodynamic loading, the geometric parameters of the blade change in value. To ensure efficient machining and the necessary efficiency and to calculate the cost of tool wear, factors that can be used as a criterion of tool wear need to be determined. Therefore, to ensure reliable diagnosis, it is necessary first of all to examine the state of the tool in specific conditions of operation [6, 9].

Damage caused by wear of cutting tools is the dominant factor determining the tool life in the optimal operating conditions [16, 17]. As an input criterion for assessing the tool state, a wear indicator should be selected that will clearly describe the change in the geometry and the properties of the tool material that result from the wear of the blade. Simultaneously, the value of the wear indicator should increase monotonically in a zone of normal wear and should allow the realization of accurate measurement [3, 10, 25]. During tool operation, it is very difficult to determine the optimal, economically justified moment to replace the tool. The permissible wear value reported by manufacturers of tools for specific machining conditions may differ from the value of wear that ensures full use of its cutting abilities [11].

Monitoring and surveillance of the blade state can be performed by an operator, or these tasks can be realized automatically. In the case of surveillance of natural wear by the man, the operator monitors the tool state during breaks in machining, based on the assessment of the quality of the workpiece surface obtained after machining, phenomena that accompany the cutting process and observation of the cutting edge geometry. Currently on the market there are systems for monitoring the state of the blades of cutting tools by using direct and indirect methods. They are diverse and have different degrees of effectiveness [13]. A system for monitoring the assumed state of tool wear which takes into account the possibility of its further development for the creation of an integrated adaptive control system of the process to ensure the maximum cutting efficiency is proposed in [2].

3. Research methodology

3.1. Determination of cutting force and cutting torque

To determine the effect of wear of drills during their operation on the value of the thrust force and cutting torque a series of durability tests was carried out on drills with a diameter D = 10 mm made of HS2-5-1 high-speed steel, at constant values of cutting parameters during one adopted period of tool life under cut. The investigations were carried out during the processing of initial holes with a diameter d = 4 mm on specially prepared samples (Fig. 1) made of C45 steel. The thrust force and cutting torque values were measured with a Kistler 9123 CQ05 dynamometer. The dynamometer was set in the drilling machine so that the axis of the drill coincided with the centre of the table of the drilling machine.

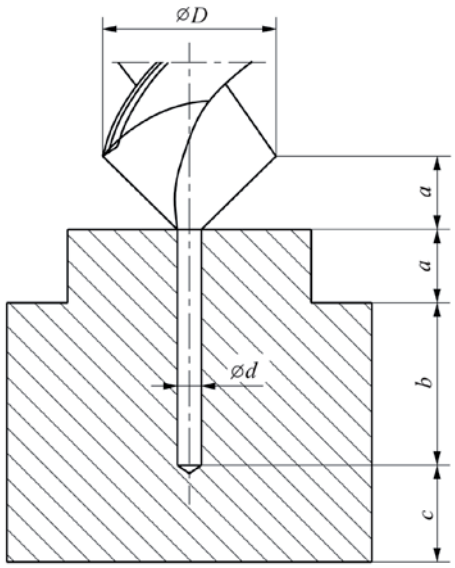


Fig. 1. Sample for the measurement of the machining resistances: a - section of the machine with main cutting edges, b - section of the machine with main cutting edges and the drill margins, c - section of the machine with all the cutting edges

The cutting torque value attributable to the margin of the drill is determined as the difference between the torque values recorded at sections *a* and *b* (Fig. 1). At section *a* (Fig. 1) only the lip of the drill participates in the cutting process and at section *b* (Fig. 1) the lip of the drill and the end cutting edge both participate. Observations of wear were carried out in all cutting parts of the drill. The evaluation of the wear of relevant surfaces was made on the basis of the maximum linear wear values of these surfaces.

3.2. Investigations of drill wear

In the literature there is no information on the location of drill wear during the drilling of steel materials that can be taken as an unequivocal criterion for its wear. Therefore, the aim of the research was to determine which controlled area on a drill made of HS2-5-1 steel may constitute such a criterion. Wear parameters which were controlled during machining of samples made of both C45 steel and cast iron EN-GJS-500-7 are shown in Fig. 2a and Fig. 2b, respectively. Research conducted by Pancielejko [22] has shown that the process of drill wear when machining cast iron has a different character from the case of machining C45 steel. Often used as the sole factor, *VB* characterizing the wear of the tool flank during machining of grey cast iron does not reflect the actual degree of wear of drills and can lead to excessive wear of cutting edges in other areas [22]. The different character of tool wear for drilling in steel and cast iron is the reason for developing drills intended only for drilling in cast iron, with a special geometry that ensures the reduction of both blade wear and the cutting force [31].

When machining steel, the tool flank is the main area subjected to wear. Wear of the drill margin causes catastrophic wear of the blade. When machining cast iron, tool flank wear is the reason for taking the drill out of service.

The tests on wear of drills during drilling in steel C45 were carried out in a column drilling machine PK 203, with the following parameters: cutting speed v_c = 40 m/min, feed rate f = 0.06 mm/rev, as well as v_c = 20 m/min and f = 0.2 mm/rev, ensuring the tool life $T_c \approx$

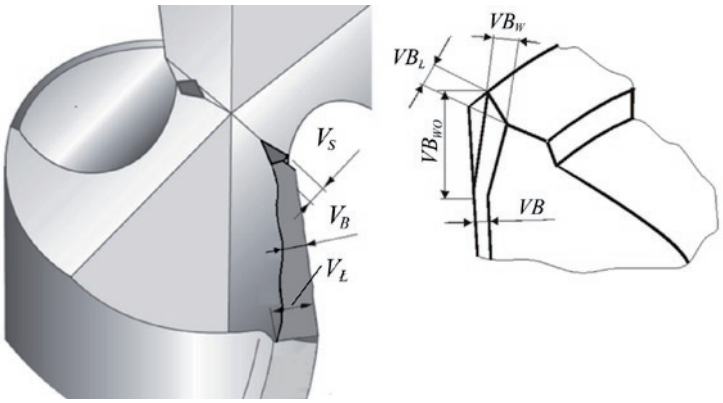


Fig. 2. Areas of wear controlled during durability testing of drills while turning C45 steel (a) and EN-GJS-500-7 cast iron (b): V_S – wear of flank of chisel edge; $V_Ł$ – wear of drill margin; VB, V_B – wear of tool flank; $VB_Ł$ – wear of drill margin flank; VB_W – wear of blade corner; VB_{WO} – wear of corner flank [12]

20 min. Changes in the value of wear of the flank of the chisel edge V_S, wear of the drill margin $V_Ł$ and wear of the tool flank V_B were investigated.

Wear of the drill margin $V_Ł$ was measured by measuring the diameter of the drill at the corners. During machining with the mentioned cutting parameters, ten drills were investigated. This allowed us to obtain statistically significant relationships of the expected values for the analyzed drill wear parameters.

Durability tests of drills during drilling in cast iron EN-GJS-500-7 were carried out at cutting speed v_c = 20 m / min, and three values of feed rate f = 0.1 mm/rev, f = 0.3 mm/rev and f = 0.4 mm/rev, ensuring the tool life of T_c = 15 min, which corresponded to machining 80–100 holes. In investigations of the tool life during drilling in cast iron, the change in the value of the following factors of tool wear was analyzed: wear of the tool flank *VB*, wear of the drill margin flank $VB_Ł$, wear of the blade corner VB_W and wear of the corner flank VB_{WO}.

In the investigations a TiN-coated drill and a drill without coating were used. To determine the effect of the TiN coating on the wear of the drill a series of durability tests were carried out. The tool life tests were conducted using specially prepared drills with a diameter D = 10 mm, on which a TiN coating with a thickness of 0.008 mm was applied. The studies were conducted at v_c = 20 m/min and f = 0.2 mm/rev, to ensure the tool life of T_c = 25 min, which corresponded to machining 100–150 holes.

The chemical composition and the basic mechanical properties of the tested materials are shown in Tables 1 and 2.

Table 1. Chemical composition of the tested materials [8, 19]

Material	Chemical composition, % wt.									
	C	Mn	Si	Cr	Ni	S	Mg	P	Cu	Sn
C45	0.42-0.50	0.50-0.80	0.17-0.37	max. 0.30	max. 0.30	0.04	-	-	-	-
EN-GJS-500-7	3.78	0.32	2.5	0.03	-	0.065	0.05	0.038	0.01	0.004

Table 2. Basic mechanical properties of the tested materials [8]

Material	Yield stress R_e [MPa]	Tensile strength R_m [MPa]	Elongation A_5 [%]	Hardness HB
C45	420	670	16	241
EN-GJS-500-7	320	500	7	170-230

4. Results and analysis

4.1. The effect of wear of HS2-5-1 steel drill on the thrust force and cutting torque

Changes in the thrust force F_f and the cutting torque M_c depending on the wear on the lip of a drill are shown in Fig. 3. It should be noted that although on the axis of ordinates only the values of wear of the lip of the drill are marked, the value of the thrust force and cutting torque show the cumulative effect of all places of wear. When catastrophic wear of the drill occurs, an intense increase in the cutting torque is observed (Figs. 3, 4). Thus the cutting torque can be used to determine the time at which catastrophic wear of the drill will appear. The general tendency to increase the thrust force F_f and cutting torque M_c during drilling operation is observed. However, until catastrophic wear occurs, the values of the thrust force and cutting torque vary differently.

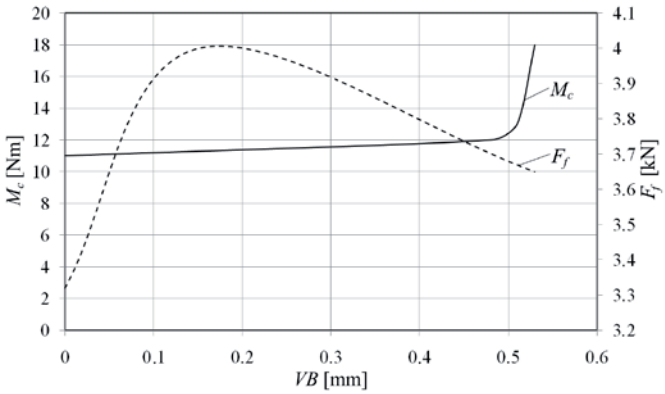


Fig. 3. Dependence of the axial force F_f and the cutting torque M_c vs. size of wear on the main tool flank VB while turning C45 steel with all parts of the drill at v_c = 20 m/min and f = 0.2 mm/rev

The lack of unequivocal correlation between the value of wear of the tool flank and thrust force and cutting torque can be explained by the fact that the drill wears not only on the primary tool flank, but also on the chisel edge of the drill, surfaces adjacent to the chisel edge and on the drill margin. The results of variation in the thrust force F_f and cutting torque *Mc* as a function of the wear of the primary tool flank during cutting by using the lips of the drill are presented in Fig. 5.

Increasing the wear of the primary tool flank *VB* leads to approximately linear increase in the thrust force F_f corresponding to the lips of the drill. After reaching a predetermined value of wear, a change in the shape of the chisel edge of the drill occurs, both on the chisel edge of the drill and on the surfaces adjacent to the chisel edge. Therefore, it is not possible to establish the relationship between the thrust force value corresponding to the chisel edge of the drill and the wear of surfaces adjacent to the chisel edge of the drill.

Although the value of the drill margin wear increases slightly during operation of the drill, a significant increase in cutting torque attributable to the primary tool flanks is observed. As the drill wear increases, the amount of heat released in the cutting zone and thermal deformation of the workpiece caused by this phenomenon also increases. With increase of the drill wear, the drilled holes deform in such a way that their diameters decrease [11]. Once the deforma-

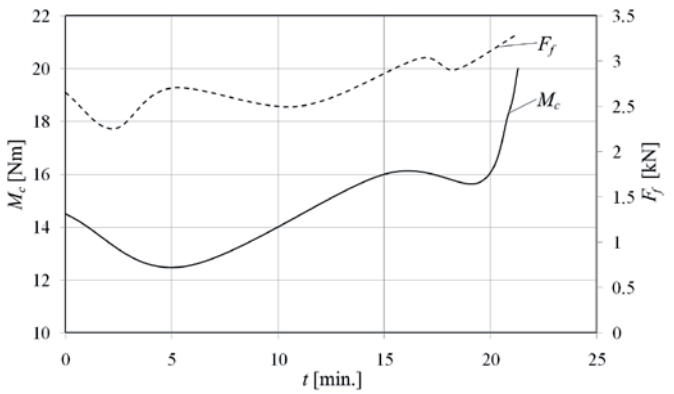


Fig. 4. Effect of machining time t on the value of the axial force F_f, the cutting torque M_c while turning C45 steel at v_c = 20 m/min and f = 0.2 mm/rev

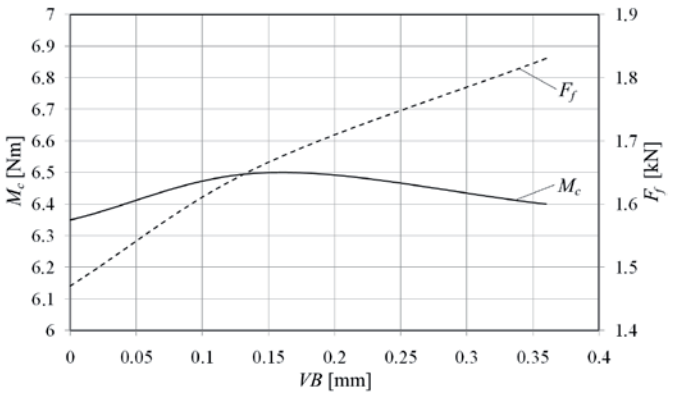


Fig. 5. Dependence of the axial force F_f and the cutting torque M_c vs wear of the main tool flank VB while turning with main cutting edges; cutting parameters: v_c = 20 m/min and f = 0.2 mm/rev

tion exceeds the clearance arising as a result of the reverse drill taper, the machined surface of the hole will cause a change in the contact area on the drill margins, their wear and the occurrence of catastrophic drill wear.

The observed values of the drill margin wear are insignificant, and therefore the chisel edge of a drill does not participate in the cutting process. It can be assumed that the increase in cutting torque attributable to the drill margin is caused by wear on the primary tool flanks.

4.2. Drill wear while drilling C45 steel

Changes in the value of the wear factors of drills made of HS2-5-1 steel without coating while processing C45 steel are shown in Figs. 6 and 7. The tool flank wear *VB* occurs in the case of both the cutting parameters applied. In the period of tool lapping, intensive wear of the chisel edge of the drill is observed. The wear of the drill margin is almost imperceptible, until the moment when catastrophic wear occurs.

The wear of both the tool flank V_B and the flank of the chisel edge V_S increases monotonically in the period of normal wear. None of these parameters reflects the moment when catastrophic wear occurs. The values of the V_B and V_S parameters at the moment of catastrophic wear depend on the values of the cutting parameters. The wear of the tool flank V_B at the moment of catastrophic wear at v_c = 40 m/min and *f = 0.6* mm/rev is 0.2 mm (Fig. 6), while at v_c = 20 m/min and *f* = 0.2 mm/rev it is about 0.55 mm (Fig. 7).

The results of investigations of the TiN-coated drills (Fig. 8) suggest that their wear is the same as drills without the coating, but the intensity of wear in controlled areas is smaller, which causes an

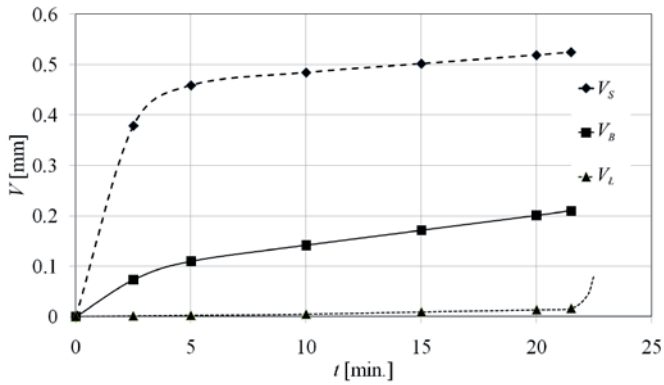


Fig. 6. Dynamics of the variation of the wear parameters while drilling C45 steel at v_c = 40 m/min and f = 0.06 mm/rev

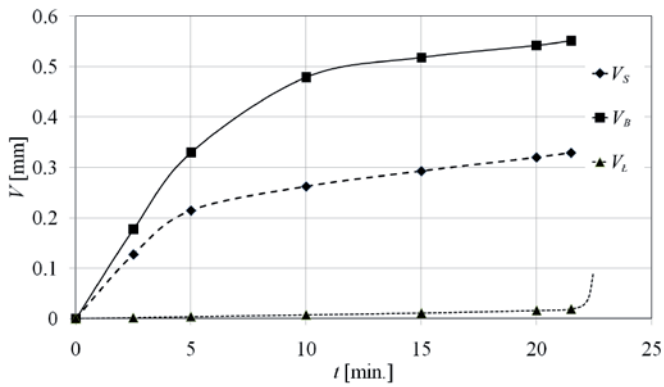


Fig. 7. Dynamics of the variation of the wear parameters while drilling C45 steel at v_c = 20 m/min and f = 0.2 mm/rev

increase of the tool life. Thus, applying the TiN coating does not change the areas of wear but only reduces the wear intensity.

The specificity of the wear of drills made of HS2-5-1 steel while drilling C45 steel primarily consists in the wear of the tool flank, and the loss of their cutting ability (catastrophic wear) occurs unexpectedly on the drill margin V_L. To prevent such unexpected wear it is necessary to know, on the basis of investigations, the value of permissible wear on the tool flank of the drill. The values of pressure during the drilling are the same order as in the case of friction welding, and lead to catastrophic wear of the drill margin. It is well known [10, 25] that the permissible value of the wear of the tool flank from which the loss of cutting ability of the drill begins depends on the cutting conditions and above all on the cutting parameters. It is not possible to determine the permissible value of wear V_B, from which catastrophic wear of the drill margin starts. This can be explained by the fact that the value of

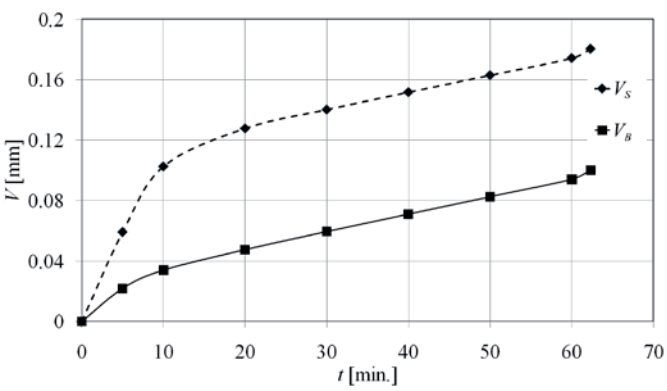


Fig. 8. Dynamics of the variation of the wear parameters while drilling C45 steel using TiN coated drill at v_c = 20 m/min and f = 0.2 mm/rev

deformation of the drilled holes under the influence of the temperature, leading to catastrophic drill wear, depends on many parameters, often mutually correlated [12].

4.3. Drill wear while drilling cast iron EN-GJS-500-7

An increase of the feed value from the value $f = 0.1$ mm/rev to $f = 0.4$ mm/rev causes an approximately two-fold increase in the values of VB_{WO}, VB_W and VB_L, parameters at which the period of normal wear of the drill begins (Figs. 9–11). Taking into account all the feed values, the zone of normal wear starts in the range of parameter values: VB = 0.07–0.1 mm VB_L = 0.15–0.3 mm VB_W = 0.2–0.46 mm and VB_{WO} = 0.3–0.54 mm. In the period of normal wear of controlled areas of the drill blade there is a monotonic increase in the value of all the tested parameters (Figs. 9–12). The value of feed has a significant impact on the moment when a fast increase of the wear of the blade corner VB_{WO} begins. The fast increase in the value of the parameter VB_{WO} during drilling at feed $f = 0.1$ mm/rev exists after making approximately $n = 100$ holes. After a three-fold increase of the feed rates, a fast increase of the wear of the corner flank occurs at the parameter value $VB_{WO} = 0.7$ mm after making 40 holes. This leads to breaking of the drill margin, preventing further control of the wear of the blade on the corner flank.

The intensification of the wear process means that after a certain time the wear VB_W encompasses the whole width of the drill margin. Under these conditions, after breaking the reference lines, measurement of the wear of the drill margin VB_L is impossible (Fig. 11). A rapid increase in the wear of the blade corner VB_W does not have a significant effect on the change of values of other wear parameters. Their values increase monotonically. Only the value of the VB factor obtained during the drilling at feed $f = 0.1$ mm/rev for the whole period of normal wear remains stable (Fig. 12). This means that there is no loss of the cutting ability of the blade and the rapid increasing

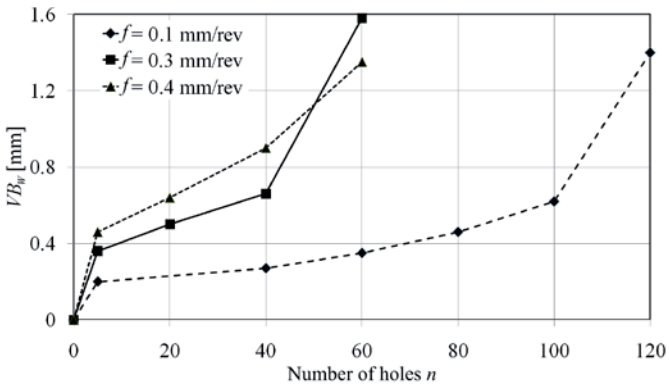


Fig. 9. Dynamics of the wear of the blade corner VB_W at v_c = 20 m/min

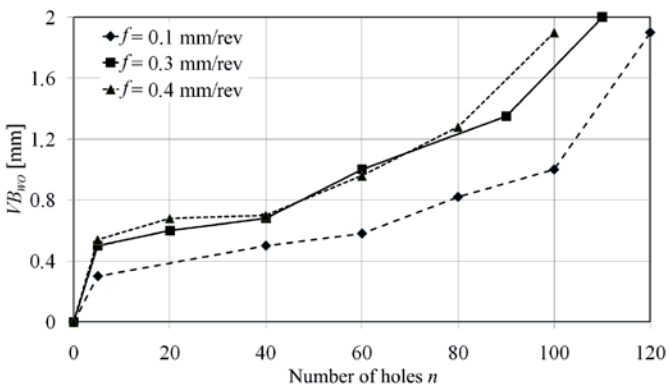


Fig. 10. Dynamics of the wear of the corner flank VB_{WO} at v_c = 20 m/min

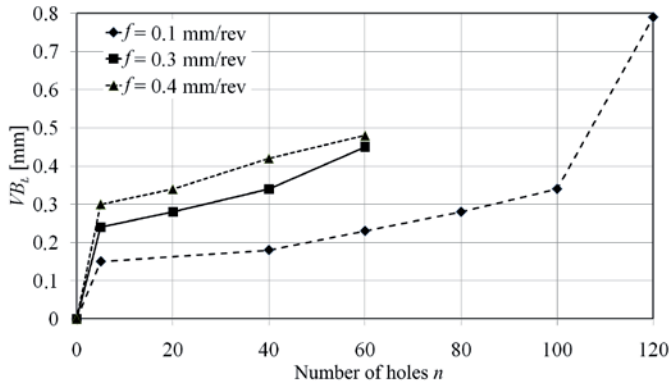


Fig. 11. Dynamics of the wear of drill margin flank VB_L at v_c = 20 m/min

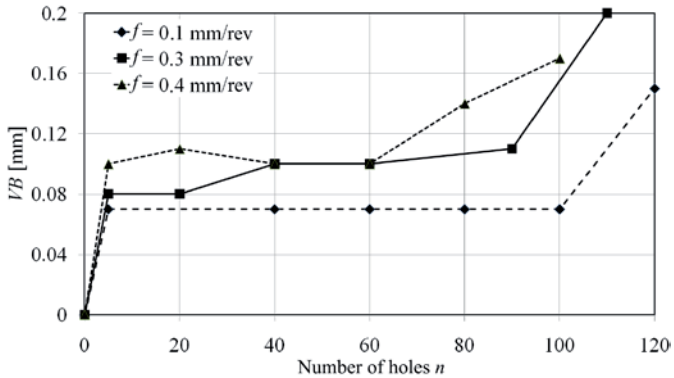


Fig. 12. Dynamics of the wear of the tool flank VB at v_c = 20 m/min

of the wear of the blade corner cannot be considered as the beginning of accelerated wear.

When operating the drill, the rapid increase in the value of the VB_W factor does not cause the loss of the cutting ability of the blade. Then increased intensity of the blade corner occurs, and thus the drill wear moves from the first to the second phase of normal wear. Loss of the cutting ability occurs as a result of the intensification of the wear of the corner flank. The moment when the cutting ability is lost depends on the feed and occurs after making about 80–100 holes (Fig. 10).

The character of the variation of the wear of the corner flank VB_{WO} at the first (to 40 holes) and second (after 40 holes) phases of normal wear changes slightly (Fig. 10). The zone of accelerated wear when working at feeds $f = 0.3$ mm/rev and $f = 0.4$ mm/rev begins after the realization of 80 holes at the VB_{WO} value of about 1.25 mm. At the feed of 0.1 mm/rev the value of the VB_{WO} factor is about 1 mm. During machining of 40–80 holes the values of the blade corner wear and wear of the drill margin flank change slightly (Figs. 9 and 11).

5. Conclusions

In catastrophic drill wear an intense increase in the cutting torque is observed. So the cutting moment can be used as a diagnostic criterion for determining the moment at which catastrophic wear occurs.

The application of a TiN coating does not change the places on the drill where wear occurs but only reduces the intensity of the wear of the tool flank. It allows the tool life of the drill to be increased.

In the case of drilling C45 steel there is a lack of such places of wear that can be used as a criterion of drill wear before its cutting ability is lost (catastrophic wear). In the period of tool lapping, intensive wear of the chisel edge of the drill is observed. Wear of the drill margin is almost imperceptible, until the moment when catastrophic wear occurs.

An increase in the feed value from the value $f = 0.1$ mm/rev to $f = 0.4$ mm/rev causes an approximately two-fold increase in the values of parameters VB_{WO}, VB_W and VB_L, at which the period of normal wear of the drill begins. In the period of normal wear of controlled areas of the drill blade there is a monotonic increase in the value of all tested parameters.

When drilling holes in samples of cast iron EN-GJS-500-7, a rapid increase in the wear of the blade corner VB_W does not have a significant effect on changing the values of other wear parameters. Their values increase monotonically. This means that no loss of the blade's cutting ability occurs and the rapid increasing of the wear of the blade corner cannot be considered as the beginning of accelerated wear. When operating the drill, the rapid increase in the value of the VB_W factor does not cause the loss of the blade's cutting ability. Then increased intensity of the blade corner occurs, and thus the drill wear goes from the first to the second phase of normal wear.

References

1. Cichosz P. Narzędzia skrawające. Warszawa: WNT, 2006.
2. Baldan A, Cuppini D, D'errico G, Rutelli G. Tool wear monitoring based on cutting power measurement. Wear 1990; 139: 303-311, http://dx.doi.org/10.1016/0043-1648(90)90052-C.
3. Czechowski K, Profeska-Filip I, Fedaczyński A. PVD coatings on ceramic materials cutting inserted. Surface Engineering 2005; 2: 19-24.
4. Dolinšek S. Work-hardening in the drilling of austenitic stainless steels. Journal of Materials Processing Technology 2003; 133: 63-70, http://dx.doi.org/10.1016/S0924-0136(02)00245-5.
5. Gawlik J. Prognozowanie stanu zużycia ostrzy narzędzi w procesie skrawania. Kraków: Wydawnictwo Politechniki Krakowskiej, 1988.
6. Gómez M P, Hey A M, Ruzzante J E, D'Attellis C E. Tool wear evaluation in drilling by acoustic emission. Physics Procedia 2010; 3: 819-825, http://dx.doi.org/10.1016/j.phpro.2010.01.105.
7. Grzesik W, Podstawy skrawania materiałów metalowych. Warszawa: WNT, 1998.
8. Grzesik W, Żak K. Investigations of surface textures produced by oblique machining of different workpiece materials. Archives of Materials Science and Engineering 2011; 52:46-53.
9. Huang B W, Kuang J H, Yu P P. Effect of crack on drilling vibration. Journal of Sound and Vibration 2009; 322: 1100-1116, http://dx.doi.org/10.1016/j.jsv.2008.11.048.
10. Jaworski J. Development tendency of cutting tool materials. Przegląd Mechaniczny 2005; 11: 33-36.
11. Jaworski J, Trzepieciński T. Możliwości zapewnienia jakościowego wykonania narzędzi z oszczędnościowej stali szybkotnącej. Rzeszów: Oficyna Wydawnicza Politechniki Rzeszowskiej, 2014.
12. Jaworski J, Kluz R, Trzepieciński T. Wpływ parametrów obróbki na miejsce powstawania zużycia i intensywność procesu zużywania wierteł. Tribologia 2014; 45: 81-90.
13. Jemielniak K. Automatyczna diagnostyka stanu narzędzia i procesu skrawaniem. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej, 2002.
14. Jurko J, Berdis A., Gecák J, Nemcova J. Study of changes the tool wear of the cutting tool part of a cast iron GTW 35-04 while drilling. Applied Mechanics and Materials 2013; 404: 82-85, http://dx.doi.org/10.4028/www.scientific.net/AMM.404.82.
15. Kilickap E. Optimization of cutting parameters on delamination based on Taguchi method while drilling of GFRP composite, Expert Systems with Applications 2010; 37: 6116-6122, http://dx.doi.org/10.1016/j.eswa.2010.02.023.
16. Królczyk G M, Niesłonny P, Legutko S. Determination of tool life and research wear while duplex stainless steel turning. Archives of Civil and Mechanical Engineering 2015; 15: 347-354, http://dx.doi.org/10.1016/j.acme.2014.05.001.
17. Królczyk G, Gajek M, Legutko S. Predicting the tool life in the dry machining of duplex stainless steel. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2013; 15: 62-65.
18. Kuczmaszewski J, Pieśko P. Wear of milling cutters resulting from high silicon aluminium alloy cast AlSi21CuNi machining. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2014; 16: 37-41.
19. Leppert T, Peng R L. Surface residual stresses in dry turning of 0.45% C steel. International Centre for Diffraction Data 2009, 2009: 304-311.
20. Liu H S, Lee B Y, Tarng Y S. In-process prediction of corner wear in drilling operations. Journal of Materials Processing Technology 2000; 101: 152-158, http://dx.doi.org/10.1016/S0924-0136(00)00434-9.
21. Meena A, El Mansori M. Drilling performance of green austempered ductile iron (ADI) grade produced by novel manufacturing technology. International Journal of Advanced Manufacturing Technology 2012; 59: 9-19, http://dx.doi.org/10.1007/s00170-011-3469-1.
22. Pancielejko M. Analiza mechanizmów zużycia wierteł ze stali HS6-5-2 z powłokami Ti(C, N) wytworzonymi metodą PVD, po badaniach eksploatacyjnych. Inżynieria Materiałowa 2010; 31: 1147-1152.
23. Przestacki D, Jankowiak M. Surface roughness analysis after laser assisted machining of hard to cut materials. Journal of Physics: Conference Series 2014: 483: 012019, http://dx.doi.org/10.1088/1742-6596/483/1/012019.
24. Semotiuk L. An analysis of the operational characteristics of innovative tool structures used in high speed rough milling processes. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2009; 11: 46-53.
25. Sinopalnikov W A, Grigoriev S N. Reliability and diagnosis of technological systems. Moscow: MGU Stankin, 2003.
26. Twardowski P, Legutko S, Krolczyk G M, Hloch S. Investigation of wear and tool life of coated carbide and cubic boron nitride cutting tools in high speed milling. Advances in Mechanical Engineering 2015; 7: 1-9, http://dx.doi.org/10.1177/1687814015590216.
27. Wartacz A, Świć A, Zubrzycki J. Zużycie ostrza narzędzia, a zmiana parametrów toczenia w kolejnych zabiegach. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2004; 4: 58-61.
28. Wieczorowski K, Matuszak A. Wpływ zużycia wiertła na dokładność wiercenia głębokich otworów. Archiwum Technologii Maszyn i Automatyzacji 1997; 17: 67-71.
29. Yan J, Murakami Y, Davim J P. Tool design, tool wear and tool life, [in:] K. Cheng (ed.), Machining dynamics. Fundamentals, applications and practices. London: Springer-Verlag, 2009: 117-149, http://dx.doi.org/10.1007/978-1-84628-368-0_5.

30. Zabel A, Heilmann M. Deep hole drilling using tools with small diameters process analysis and process design, CIRP Annals - Manufacturing Technology 2012; 61: 111-114.
31. Zheng X, Zhang H, Xue K, Chen M, Zhang Y, Optimization of twist drill's geometry for cast iron. Key Enginnering Materials 2010; 431-432: 555-558, http://dx.doi.org/10.4028/www.scientific.net/KEM.431-432.555/
32. Zoriktuev V T, Nikitin Y A, Sidorov A S. Monitoring and prediction of cutting-tool wear. Russian Engineering Research 2008; 28: 88-91.

Jan JAWORSKI
Rafał KLUZ

Department of Manufacturing and Production Engineering
Rzeszow University of Technology
Al. Powstańców Warszawy 12, 35-959 Rzeszów

Tomasz TRZEPIECIŃSKI
Department of Materials Forming and Processing
Rzeszow University of Technology
Al. Powstańców Warszawy 12, 35-959 Rzeszów

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

Article citation info:
IWANEK M, KOWALSKA B, Hawryluk E, Kondraciuk K. Distance and time of water effluence on soil surface after failure of buried water pipe. Laboratory investigations and statistical analysis. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 278–284, http://dx.doi.org/10.17531/ein.2016.2.16.

Małgorzata IWANEK
Beata KOWALSKA
Elżbieta HAWRYLUK
Kamil KONDRACIUK

DISTANCE AND TIME OF WATER EFFLUENCE ON SOIL SURFACE AFTER FAILURE OF BURIED WATER PIPE. LABORATORY INVESTIGATIONS AND STATISTICAL ANALYSIS

ODLEGŁOŚĆ I CZAS WYPŁYWU WODY NA POWIERZCHNIĘ TERENU PO AWARII PODZIEMNEGO WODOCIĄGU. BADANIA LABORATORYJNE I ANALIZY STATYSTYCZNE

One solution to limit the inconvenience caused by suffosion processes following pipe breakages is retaining so-called protection zones near the pipes, the utilization of which would be handled by the system operator. Due to the fact that to determine the size of such zones is a challenging task, the analysis should be performed gradually, based on successive field studies, laboratory and numerical research. The present article is the outcome of the first stage of the laboratory research eventually aiming at the determination of the protection zone around a potential leakage in a water supply pipe. The first stage of the investigations was devoted to (1) the assessment of an average distance between the place of water effluence on the soil surface and the place of the water failure for 4 different areas of leak and 11 values of hydraulic pressure head in the pipe, (2) the initiatory assessment of the protection zone dimensions for analysed soil conditions, (3) the analysis of dependence between the time of water effluence on the soil surface after a failure of a buried water pipe and the leak area as well as the hydraulic pressure head in the pipe. The scope of the works comprises laboratory study and statistical analysis. The research was carried out preserving geometrical and kinematic similarity. The obtained results should be considered initial, oriented towards further stages of laboratory research comprising dynamic similarity.

Keywords: *water supply, breakage, maintenance, water effluence.*

Jedną z propozycji ograniczenia uciążliwości spowodowanych zjawiskami sufozyjnymi po awarii wodociągu jest zachowanie w pobliżu przewodów tzw. stref ochronnych, o zagospodarowaniu których decydowałby eksploatator sieci. Określenie wymiarów takich stref jest bardzo trudnym zadaniem, dlatego stosowne analizy powinny odbywać się stopniowo, na bazie kolejnych etapów badań terenowych, laboratoryjnych i numerycznych. W ramach niniejszej pracy przedstawiono wyniki pierwszego etapu badań laboratoryjnych, których ostatecznym celem jest wyznaczenie strefy ochronnej wokół ewentualnej nieszczelności rury wodociągowej. Pierwszy etap badań objął określenie przeciętnej odległości wypływu wody na powierzchnię terenu od miejsca awarii podziemnego wodociągu dla 4 różnych powierzchni nieszczelności przewodu oraz 11 wysokości ciśnień w przewodzie, wstępne oszacowanie wielkości strefy ochronnej dla analizowanych warunków gruntowych oraz analizę zależności między czasem wypływu wody na powierzchnię terenu po awarii podziemnego wodociągu a powierzchnią nieszczelności i wysokością ciśnienia w przewodzie. Zakres pracy obejmował badania laboratoryjne i analizy statystyczne. Badania przeprowadzono z zachowaniem podobieństwa geometrycznego i kinematycznego. Uzyskane wyniki należy więc traktować jako wstępne, ukierunkowujące dalsze etapy badań laboratoryjnych, uwzględniające również podobieństwo dynamiczne.

Słowa kluczowe*: wodociąg, awaria, eksploatacja, wypływ wody.*

1. Introduction

The use of water supply pipes all over the world has always been accompanied by breakages and leakages. What can reduce a number of damages is skilful management of a water supply system and proper maintenance. It is, however, impossible to entirely eliminate such incidents as, most often, they occur randomly [5, 7, 24]. They can result in financial and social losses [7, 8, 25, 28]. Moreover, leakages can pose a threat to the safety of people and property particularly in urban agglomerations, where water supply systems are located within roadway, constituting an element of an underground utility, as well as in areas of compact settlement [13]. The threat emerges as a result of the particles being washed out from the soil skeleton during the breakage of an underground pipe which can lead to the formation of empty spaces beneath the ground surface and contribute to the creation of depression or holes in the Earth's surface (suffosion processes) [1, 4, 9]. Such incidents took place worldwide and produced detrimental social and economical effects [22]. Occurrence of internally unstable soils, especially in the range of the loess plateau [2] as well as a high failure intensity rate of water supply systems, compared to other countries [15, 16, 18, 19], are factors which increase the risk of the emergence of such a problem in Poland.

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

One of the proposals on how to reduce negative consequences of suffosion phenomena following a potential water supply damage is retaining so-called protection zones near the pipes, with all the decisions regarding their use taken exclusively by the system operator [11, 14]. To define the size of such zones seems a difficult task, taking into account the complex character of the phenomenon in question [3, 10, 27]. During the breakage of an underground water supply there are three fundamental interconnected processes: pressure water flow in a closed pipe, water effluence through a leak and a water flow in a porous medium. They can be characterised by multiple parameters comprising values which are independent (e.g. water pipe diameter), mutually interdependent within one process (e.g. dependence between the volume of the water flow in a water pipe and the water pipe pressure level), as well as affecting each other in various processes (e.g. the speed of water flowing out through a leak affects the speed of the water flow in the ground). Another obstacle is the fact that certain parameters are variable in space or time (e.g. soil hydraulic conductivity) or depend on external factors (e.g. water viscosity index depending on the temperature). Furthermore, connections between the parameters of the processes can be described by means of complex dependencies – for instance water movement in the ground can be represented with the Richards equation [23], which is a second order partial differential equation. What can pose another problem is the process of the soil particles being washed out and transferred with water flowing out of a water pipe after its breakage. Due to the obstacles mentioned above, the analysis aiming at determining the size of protection zones should be performed gradually based on successive field studies, laboratory and numerical research.

The article presents the results of the first stage of the research eventually focused on determining the size of a protection zone around the area where a water pipe is particularly exposed to leakage. The first stage of the investigations covers (1) the assessment of an average distance between the place of water effluence on the soil surface and the place of the water failure for 4 different areas of leak and 11 values of hydraulic pressure head in the pipe, (2) the initiatory assessment of the protection zone dimensions for analysed soil conditions, (3) the analysis of dependence between the time of water effluence on the soil surface after a failure of a buried water pipe and the leak area as well as the hydraulic pressure head in the pipe. The scope of the article comprises laboratory tests and statistical analysis. The research was carried out preserving geometrical and kinematic similarity [12]. Obtained results are initial and oriented towards further stages of laboratory research comprising dynamic similarity.

2. Methodology

The research presented in the article is twofold. The first part consists in laboratory simulation of water supply system breakage, conducted on a laboratory setup and reflecting natural operation conditions scaled to 1:10. The second part embraces statistical analysis of the laboratory test results including the average distance between the point of water effluence on the sand surface and the leak of the pipe as well as time of this effluence.

Laboratory tests required constructing the setup (Fig. 1). It consisted of a water supply pipe 1 buried in medium sand filling a 1.5 m × 1.5 m × 0.5 m box 2. The sand was manually compacted in two-centimetre layers according to closely prearranged processes. During the laboratory tests the following sand parameters were determined (tab. 1 and 2):

- article-size distribution – using a sieve analysis on the basis of the standard [21],
- degree of compaction – using the standard Proctor test – the method No 1 given in the standard [21] (a small cylindrical mould – 113 mm in the inner diameter, 3 soil layers, 25 drops of 2,5-kilogram hammer, 320 mm of drop distance),

Table1. Results of a sieve analysis

Particle size *d*, mm	Content, %
$4 < d$	8.97
$3.15 < d \le 4$	2.17
$2 < d \le 3.15$	4.2
$1.4 < d \le 2$	3.66
$1 < d \le 1.4$	3.63
$0.8 < d \le 1$	3.83
$0.5 < d \le 0.8$	17.74
$0.25 < d \le 0.5$	35.67
$0.125 < d \le 0.25$	15.06
$d \le 0.125$	5.03

Table 2. Parameters of compacted sand

Parameter	Value of parameter
Degree of compaction	0.93÷0.94
Porosity	0.26
Saturated conductivity, m/s	$1.4 \cdot 10^{-4}$
Volumetric water content, %	3.89÷5.20

- porosity – using a Le Chatelier flask [22],
- saturated conductivity – using the GeoN permeameter (Geonordic AB, Sweden),
- volumetric water content – measured before each simulation using a TDR-meter (EASY-TEST, Lublin, Poland).

One end of the pipe 1 (Fig.1) extending from the box 2 was linked through a rubber hosepipe 3 to a collection container 6 placed at a assumed height. The other end of the pipe 1 directed water to a floor drain through a rubber hosepipe 8. The pipe consisted of two equal length parts connected by a bell-and-spigot joint (9). Drain 10 installed at the bottom of the box 2 enabled the outflow of excess water after each trial.

The experiment consisted in introducing water under pressure into a damaged water supply pipe so as to produce controlled leakage. First, all valves were closed and water was poured into the container 6 above the assumed level. Next, the valves were opened and the deaeration process took place. When water in container 6 reached the assumed level, valve 7 was closed and the end of pipe 1 from the side of feeding point was pulled in the opposite direction resulting in the loosening of the bell-and-spigot joint 9. After the emergence of water on the sand surface valve 4 was closed. The next step was to determine the size of the soil surface cavity and its location in relation to the leak in the water supply pipe. An important element of the laboratory setup was the holder 11, installed inside the box on the supply side of the pipe 1 (Fig. 2). The holder enabled the same width of the leak in every repetition of an experiment.

The tests were conducted for 4 different areas of leaks ensuing due to loosening of the pipe connection – 4.71 cm^2, 5.58 cm^2, 9.42 cm^2 and 12.25 cm^2. The width of the leak between a spigot end and a socket end of the pipe equalled 15 mm for each experiment repetition, while the inner pipe diameter changed (10 mm, 13 mm, 20 mm and 26 mm). Internal water pressure in the pipe varied from 31.9 to 58.8 kPa (from 3.25 to 6.0 m H_2O) depending on the height of container 6 and the water level in it.

Moist sand being in the area around a leak and between a leak and a place of water effluence on the sand surface was replaced by dry sand and compacted after each experiment. Moreover, replacement of the whole sand was performed after a series of experiments for the same leak area.

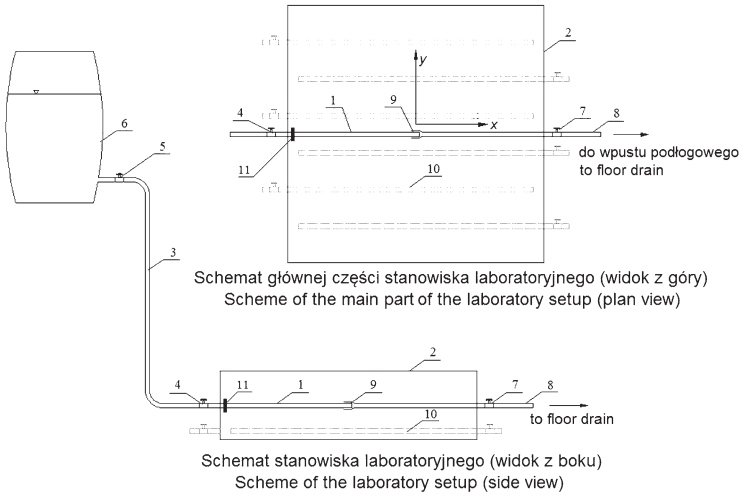


Fig. 1. Scheme of laboratory setup for physical simulation of water supply damage [11]: 1 – water supply pipe, 2 – box filled with sand, 3 – rubber hosepipe from the feeding side, 4,5 – cut-off valve on the feeding side, 6 – water container, 7 – cut-off valve on the outflow side, 8 – rubber hosepipe inserted into floor drain, 9 – bell-and-spigot joint, 10 – drainage, 11 –holder

Fig. 2. Holder enabling the same width of leak after each loosening of pipe connection

The entire experiment, starting from the deareation to the moment of water effluence on the soil surface, was recorded by a video camera and documented photographically. Altogether, there were 105 successful trials carried out.

Obtained laboratory test results allowed to determine anticipated values of the distance between the place of the water effluence on the soil surface and pipe leak after the breakage of the water supply pipe, based on statistical estimations calculated with Statistica 10 software (StatSoft, Inc.). The first step of the analysis was descriptive statistics [17, 26] comprising measures of central tendency (arithmetic mean, mode and median) as well as measures of dispersion and the shape of distribution (standard deviation, skewness, kurtosis). In the cases where values obtained in laboratory trials were not close to normal distribution, it was examined whether the laboratory test result logarithms can be characterised by it. Distribution normality of the results as well as their logarithms was verified with the Kolmogorov-Smirnov normality test modified by Liliefors and Shapiro-Wilk [6]. Calculations were conducted for all the data obtained in laboratory trials (excluding extreme values) without any divisions and for the data divided per leak area and hydraulic pressure in the pipes.

The results provided the base for determining the indicative range of the protection zone around the leakage through defining ranges of tolerance with the confidence level of 95% and 99%. All the results regarding the distance between the water effluence on the sand surface and the place on the sand surface located directly above the leak were taken into consideration. With regards to statistical calculations, the lower tolerance limit can take a negative number, though it is inconsistent with the definition of distance. That is why in calculations 0 value was assumed for the lower tolerance limit, *i.e.* the place of water effluence on the soil surface located closest to the leak is the area situated directly above the leak, which corresponds to the real conditions. The determined upper tolerance limit marks the radius of the circular protection zone, whose centre is located directly above the place of the leak.

The next research stage was the analysis of the results of measurements of time between the loosening of a pipe connection and an effluence of water on the soil surface. As in the case of a distance, data distributions were verified with the Kolmogorov-Smirnov normality test modified by Liliefors as well as with the Shapiro-Wilk normality test. Next, an influence of the leak area and the hydraulic pressure head in the pipe on time of water effluence was evaluated on the basis of a linear, exponential and power regression model.

3. Results and discussion

Summary statistics of distances between the water effluence on the sand surface and the place on the sand surface located directly above the leak obtained as a result of laboratory study are compiled in Tables 3-5.

Measures compiled in Table 3 indicate that all the values of the analysed distance examined together with no division are characterised by the distribution different from normal. Distribution of logarithmized distance values, however, is approximate to normal (Fig. 3) which is proved by normality tests with 95 per cent confidence level. The average distance between the water effluence on the soil surface and the place of the leakage obtained in laboratory study does not amount to 30.60 cm, as implied by Table 3, but to 28.18 cm, constituting the result of number 10 raised to the power of the logarithm arithmetic mean of measured distances.

Table 3. Summary statistics of all distances registered on the soil surface between water effluence and the place of the breakage

Number of measurements *n*	Distance			Mode	Median	Standard deviation	Skewness	Kurtosis
	max	mean	min					
-	cm	cm	cm	cm	cm	cm	-	-
105	59.00	30.60	9.00	Multiple	29.00	12.64	0.52	-0.60

None of the distributions whose measures are presented in Table 4 were found normal. Dependencies between the mean, median and mode as well as positive skewness values indicate that the obtained

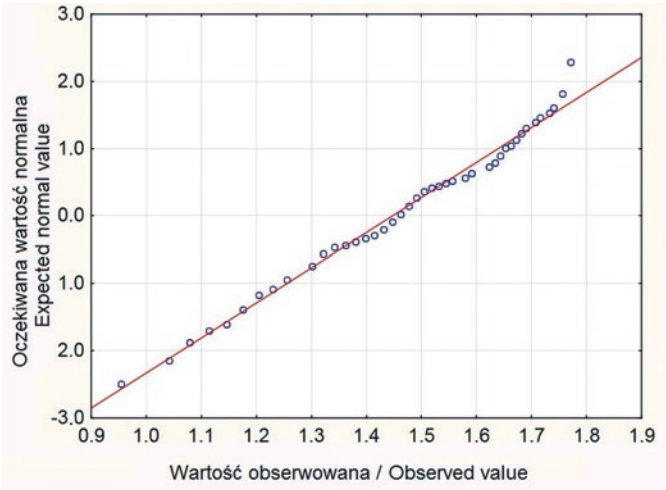


Fig. 3. Normality graph of the logarithm distribution for measured distances

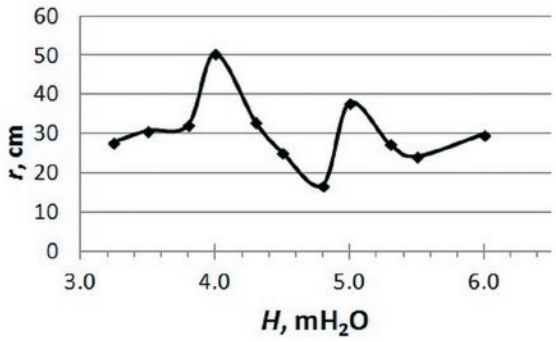


Fig. 4. Correlation between water pressure in pipe H and distance r on the ground surface between water effluence and pipe leak

Table 4. Summary statistics of distances recorded on the soil surface between the water effluence and breakage for different leak area F

F	Number of measur. *n*	Distance			Mode	Median	Standard deviation	Skewness	Kurtosis
		max	mean	min					
cm^2	-	cm	cm	cm	cm	cm	cm	-	-
4.71	25	48.90	28.28	16.00	21.00	25.10	10.34	0.72	-0.74
5.58	25	35.20	20.92	9.00	17.90	17.90	8.57	0.14	-1.38
9.42	27	58.80	38.74	18.10	27.00	38.00	10.34	0.09	-0.78
12.25	24	59.00	33.41	15.80	20.00	28.50	13.41	0.73	-0.80

results are of a positively asymmetric distribution for all leak areas analysed in laboratory tests. Therefore, it is the mode value that should be assumed as the average distance value obtained for particular leak areas in tests, not the arithmetic mean, which is common the case in normal distribution. No regularities in the relation between the leak area and the distance to the effluence point occurred were observed.

Among 11 distributions whose measures are presented in Table 5, only 2 are close to normal (for $H = 3.25$ m H_2O and $H = 5.30$ m H_2O), whereas 2 are close to normal after logarithmizing data (for $H = 3.50$ m H_2O and $H = 3.80$ m H_2O). The average distance between the water effluence and the water pipe breakage in laboratory tests for the pressure values of $H = 3.25$ m H_2O and $H = 5.30$ m H_2O is in accordance with the mean presented in Table 5 (28.00 cm and 27.45 cm respectively), whereas in the case of values of $H = 3.50$ m H_2O and $H = 3.80$ m H_2O it is 30.90 cm and 32.36 cm respectively, which is a result of the data being expressed as logarithms. As far as the remaining pressure values are concerned, it can be assumed that the average effluence distance is in accordance with the arithmetic mean, but since the results obtained for those values are not close to normal (being irregular in all the cases), it should be borne in mind that the average will not be an efficient estimator [17]. Similarly to the analysis of the influence of leak area on the effluence distance after potential breakage, no regularities in the relations between the pipe water pressure and the distance itself were observed (Fig. 4).

In 105 physical simulations of water supply failures carried out in a laboratory there were no cases of water effluence directly above the leakage. This necessitated determining the zone on the soil surface of probable water effluence after an underground water system failure through the estimation of the upper tolerance limit. Due to the fact that previous analysis showed that the logarithms of all distances obtained during laboratory tests, examined without division, are of a normal distribution, tolerance intervals were determined for the logarithms of those values (Table 6).

Table 5. Summary statistics of distances recorded on the soil surface between the water effluence and breakage for different pressure levels H

H	Number of measur. *n*	Distance			Mode	Median	Standard deviation	Skewness	Kurtosis
		max	mean	min					
m H_2O	-	cm	cm	cm	cm	cm	cm	-	-
3.25	6	36.10	28.00	17.10	Multiple	28.40	6.46	-0.78	1.27
3.50	7	21.00	32.28	46.80	21.00	28.20	10.39	0.30	-1.73
3.80	7	49.10	33.71	21.00	30.00	31.90	8.74	0.57	1.25
4.00	7	62.00	50.43	39.30	42.00	51.70	9.29	-0.03	-2.23
4.30	7	44.00	33.14	18.00	44.00	34.00	4.00	-0.59	-1.27
4.50	7	44.50	25.29	15.00	Multiple	22.30	10.86	0.09	0.90
4.80	8	27.00	17.00	9.00	Multiple	16.00	6.46	0.24	-0.82
5.00	10	55.00	37.90	18.00	Multiple	45.50	14.73	-0.41	-1.98
5.30	11	35.30	27.45	20.00	27.00	27.80	4.35	-0.22	0.39
5.50	10	31.90	24.30	13.20	Multiple	26.00	6.49	-0.67	-1.02
6.00	11	44.00	30.00	20.00	Multiple	28.00	9.42	0.60	-1.22

Conducted measurements show that the protection zone radius, depending on accepted statistical assumptions, is found within the range from 46.5 cm to 77.1 cm in laboratory conditions (Fig. 5), *i.e.* the zone of the radius 46.5 cm (4.46 m in real conditions) will cover 70 per cent of the water effluence points with a 95 per cent confidence level, whereas the zone of a 77.1 cm long radius (7.71 m in real conditions) will cover 95 per cent of the ef-

Table 6. Upper tolerance limit for distances measured on soil surface between water effluence and breakage

Tolerance level [%]	70	75	80	85	90	95
Confidence level [%]	95					
Logarithm of the upper tolerance level [log cm]	1.66	1.69	1.72	1.75	1.80	1.86
Upper tolerance limit [cm]	46.5	49.2	52.5	56.7	62.7	73.2
Confidence level [%]	99					
Logarithm of the upper tolerance level [log cm]	1.68	1.70	1.73	1.77	1.81	1.89
Upper tolerance limit [cm]	47.8	50.73	54.3	58.9	64.6	77.1

Analysis of the time measurements results indicated that the obtained values are not characterised by the normal distribution in any cases – neither for data examined together with no division nor for data divided according to hydraulic pressure, nor for data divided according to leak area. The normal distribution of logarithmized data was obtained for rare cases – for $F = 4.71$ cm^2 and $F = 9.42$ cm^2 for data divided according to leak area and for $H = 4.30$ m H_2O and $H = 5.50$ m H_2O for data divided according to hydraulic pressure.

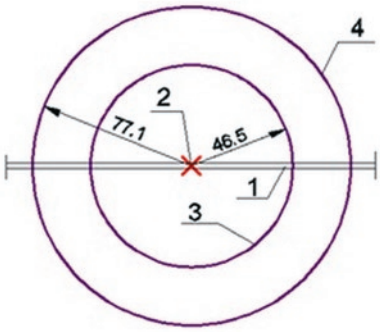


Fig. 5. Minimal and maximal (depending on statistical assumptions) protection zone range: 1 – water supply, 2 – leak, 3 – protection zone range for a 95 per cent confidence level and 70 per cent tolerance level, 4 – protection zone range for a 99 per cent confidence level and 95 per cent tolerance level

fluence points with a 99 per cent confidence level. The increase of the confidence level from 95% to 99% results in the growth of the zone radius by about 1 ÷ 4 cm in laboratory conditions. It ought to be mentioned, however, that the difference increases with the increase of the tolerance level. It is the change of the tolerance level which is of bigger impact on the value of the zone radius, as with the increase of the parameter in question from 70% to 95%, the radius rises by 26.7 cm with a 95% confidence level and by 29.3 cm with a 99% confidence level.

Table 7. Time of water effluence on sand surface starting from the moment of failure occurrence, in dependence on leak area and hydraulic pressure head

H [m H_2O]	Effluence time *t* [s] for area of leak			
	4.71 cm^2	5.58 cm^2	9.42 cm^2	12.25 cm^2
3.3	22.10	-	65.67	91.67
3.5	116.03	26.15	41.08	81.25
3.8	65.98	117.23	15.50	3.01
4.0	13.51	21.50	51.11	37.98
4.3	14.49	4.07	29.18	82.53
4.5	7.50	8.57	61.50	19.50
4.8	26.00	15.89	18.51	4.00
5.0	61.55	1.00	10.00	1.54
5.3	10.01	1.99	8.12	3.87
5.5	5.50	37.20	9.03	2.06
6.0	9.47	15.00	2.52	2.50

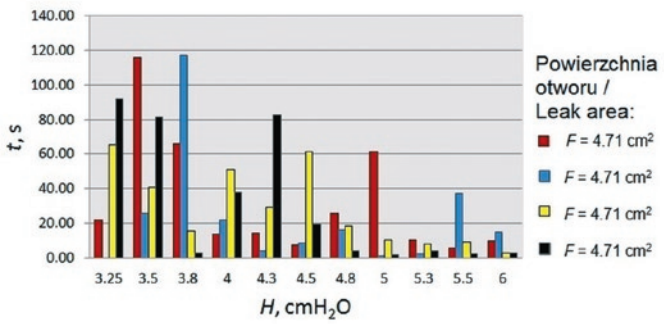


Fig. 6. Dependence between time of water effluence on soil surface after pipe loosening and hydraulic pressure head (H) for different leak areas (F)

The arithmetic means of time of water effluence measurement values for different hydraulic pressures in a pipe and leak areas are shown in Table 7. The means are not efficient estimators because of the data distribution irregularity, so they should be treated as indicative only.

Analyzing the obtained results (Tab. 7) it can be noticed as per the expectations, that higher place of the water supply container results in tendency of time of water effluence on the sand surface to be lower. The mentioned dependence is clearly visible in the bar chart (Fig. 6). However it should be emphasized, that unambiguous evaluation of kind of trend is difficult. The coefficient of determination for all considered models (linear, exponential and power) was low (Tab. 8).

Table 8. Coefficient of determination R^2 for linear, exponential and power regression model defining function t(H)

Leak area *F*	R^2 for regression model		
	linear	exponential	power
4.71 cm^2	0.23	0.32	0.25
5.58 cm^2	0.15	0.13	0.20
9.42 cm^2	0.56	0.68	0.52
12.25 cm^2	0.53	0.57	0.53

Any regularity between the time of water effluence and the leak area in a loosening pipe was noticed during the analysis of the data shown in Table 7 and in Fig. 6.

4. Summary and conclusions

Failures and leakages are inextricably linked to the use of water supply systems and, apart from economic loss, they generate a real risk of dangerous suffosion phenomena in the soil. In order to minimise their harmful effects, it has been suggested that protection zones

of a limited land development should be retained around water supply elements which are particularly subject to leakages. Currently there are no guidelines on how to determine the range of such zones, even though water supply system operators demonstrate keen interest in this sort of data.

Determination of the size of protection zones is a significantly time-consuming task of considerable difficulty due to the complexity of the effects accompanying underground water supply system failure. The sheer number of parameters influencing the direction as well as the speed of water movement in the ground together with interconnections between those parameters necessitate the division of the research into stages, beginning with the most basic. Research constituting the first stage presented in this article proved to be highly valuable with respect to further study. What it did was enabling the recognition of the obstacles that had to be overcome during the physical simulation of a water supply system breakage and setting direction for further research.

Statistical analysis of the results obtained at the first stage of the laboratory tests performed with the use of geometric and kinematic probability was also found promising and allowed preliminary estimation of the range of the protection zone. The range in question is based predominantly on the obtained distances between the water effluence on the surface of the soil and breakage, and to a lesser extent, on statistical assumptions. Moreover, the conducted analyses confirmed clear dependence between the time of water effluence on soil surface after a waterpipe failure and the hydraulic pressure in a pipe. However, values obtained as a result of time measurements were not characterized by the normal distribution, so they should be treated as indicative only. Thus, in order to achieve the best possible results of estimating both the zone range and parameters influencing the time of water effluence on soil surface, the quality of the laboratory research should not be neglected above all. It is, therefore, suggested that in order to reach normal result distribution, dimensional analysis preceding laboratory study should be repeated, and ought to focus on extending the number of parameters influencing the studied process, together with increasing the number of simulated breakages repetitions in steady conditions.

Acknowledgements

The article has been financed through statutory activities of Department of Water Supply and Wastewater Disposal of Environmental Engineering Faculty of Lublin University of Technology

References

1. Bendahmane F., Marot D., Rosquoët F., Alexis A. Experimental parametric study of suffusion and backward erosion. International Journal of Geotechnical and Geoenvironmental Engineering 2008; 134 : 57-67, http://dx.doi.org/10.1061/(ASCE)1090-0241(2008)134:1(57).
2. Bernatek A. Rola sufozji w rozwoju rzeźby – stan i perspektywy badań. Przegląd Geograficzny 2014; 86: 53–76, http://dx.doi.org/10.7163/PrzG.2014.1.4.
3. Błażejewski R., Maćkowski S. Eksfiltracja, infiltracja i sufozja przez szczeliny uszkodzonych kanałów ściekowych. In: J. Dziopak, D. Słyś, A. Stec (eds.): Materiały II Ogólnopolskiej Konferencji Naukowo-Technicznej Infrastruktura Komunalna a Rozwój Zrównoważony Terenów Zurbanizowanych. INFRAEKO 2009: 19-31.
4. Bonelli S., Marot D. On the modelling of internal soil erosion. The 12th International Conference of International Association for Computer Methods and Advances in Geomechanics (IACMAG) 1–6 October, Goa, India 2008.
5. Denczew S. The reliability, safety and risks of water supply system operation versus critical infrastructure. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2007; 2: 15-21.
6. Dobosz M. Wspomagana komputerowa analiza statystyczna wyników badań. Warszawa: Akademicka Oficyna Wydawnicza EXIT, 2001.
7. Hotloś H. Analiza uszkodzeń i kosztów naprawy przewodów wodociągowych w okresie zimowym. Ochrona Środowiska 2009; 31: 41-48.
8. Hotloś H. Metodyka i przykłady prognozowania kosztów naprawy przewodów wodociągowych. Ochrona Środowiska 2006; 28: 49-54.
9. Indraratna B., Nguyen V. T., Rujikiatkamjorn C. Assessing the Potential of Internal Erosion and Suffusion of Granular Soils. Journal of Geotechnical and Geoenvironmental Engineering 2011; 137: 550-554, http://dx.doi.org/10.1061/(ASCE)GT.1943-5606.0000447.
10. Iwanek M. Zjawisko sufozji jako skutek awarii infrastruktury wodociągowej lub kanalizacyjnej. Przegląd literatury. In: K. Kuś, F. Piechurski F. (eds.): Nowe Technologie w Sieciach i Instalacjach Wodociągowych i Kanalizacyjnych. Gliwice 2014, 57-78.
11. Iwanek M., D. Kowalski D., Kowalska B., Hawryluk E., Kondraciuk K. Experimental investigations of zones of leakage from damaged water network pipes. In: C. A. Brebbia, S. Mambretti (eds.): Urban Water II. WIT Transactions on the Built Environment 2014; 139; Southampton, Boston, UK: WIT Press: 257-268, http://dx.doi.org/10.2495/uw140221.
12. Iwanek M., Malesińska A. Zastosowanie teorii podobieństwa w modelowaniu awarii sieci wodociągowych. Gaz, Woda i Technika Sanitarna 2015; 3: 82-86.
13. Khomenko V. P. Suffosion hazard: today's and tomorrow's problem for cities. In: Culshaw, M. G., Reeves, H. J., Jefferson, I. & Spink, T. W. (eds.) Engineering geology for tomorrow's cities, Geological Society, London: Engineering Geology Special Publication, 2009.
14. Kowalski D., Jaromin K. Metoda wyznaczania zasięgu strefy ochrony wodociągowych przewodów tranzytowych. Proceedings of EC Opole 2011; 4: 419-424.
15. Kowalski D., Kowalska B., Kwietniewski M., Sygacz – Adamska J. Analiza bilansu wody oraz awaryjności wybranego systemu dystrybucji wody. In: Z. Dymaczewski, J. Jeż – Walkowiak (eds.): Zaopatrzenie w wodę, jakość i ochrona wód, tom 2, Polskie Zrzeszenie Inżynierów i Techników Sanitarnych, Oddział Wielkopolski, Poznań 2012: 77-92.
16. Kowalski D., Miszta-Kruk K. Failure of water supply networks in selected Polish towns based on the field reliability tests. Engineering Failure Analysis 2013; 35: 736-742, http://dx.doi.org/10.1016/j.engfailanal.2013.07.017.
17. Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M. Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, cz. II Statystyka matematyczna. Warszawa: Wydawnictwo Naukowe PWN, 2011
18. Kutyłowska M., Hotloś H. Failure analysis of water supply system in the Polish city of Głogów. Engineering Failure Analysis 2014; 41: 23-29, http://dx.doi.org/10.1016/j.engfailanal.2013.07.019.
19. Kwietniewski M. Awaryjność infrastruktury wodociągowej i kanalizacyjnej w Polsce w świetle badań eksploatacyjnych. XXV Konferencja Naukowo-Techniczna Awarie Budowlane, 24-27 maja 2011: 127-140.
20. Myślińska E. Laboratoryjne Badania Gruntów. Warszawa: Wydawnictwo Naukowe PWN, 1998.

21. PN-88/B-04481 Grunty budowlane. Badania próbek gruntu – Building soils. Laboratory tests.
22. Ragozin A. L. Basic principles of natural hazard risk assessment and management. In: R. Oliveira, L. F. Rodrigues, A. G. Coehlo, A. P. Cunha (eds.): Proceedings of the 7th International Congress of the International Association of Engineering Geology 3, Lisbon, Portugal, A. A. Balkema, Rotterdam 1994: 1277-1286.
23. Richards L.A. Capillary conduction of liquids through porous mediums. Physics 1931; 1: 318-333, http://dx.doi.org/10.1063/1.1745010.
24. Romano M., Kapelan Z. Geostatistical techniques for approximate location of pipe burst events in water distribution systems. Journal of Hydroinformatics 2013; 15.3: 634-635, http://dx.doi.org/10.2166/hydro.2013.094.
25. Siwoń Z., Cieżak W., Cieżak J. Praktyczne aspekty badań strat wody w sieciach wodociągowych. Ochrona Środowiska 2004; 26/4: 25-30.
26. Sobczyk M. Statystyka. Warszawa: Wydawnictwo Naukowe PWN, 2014.
27. Vlahovič M. Dispersion of results during determination of possibilities of internal erosion phenomenon using the empirical methods. Proceedings of 5th International Congress International Association of Engineering Geology 2, Buenos Aires 1986: 1037-1040.
28. Żaba T., Langer A. Monitoring strat wody elementem ograniczenia kosztów działalności przedsiębiorstwa. Napędy i sterowanie 2012; 4:100-103.

Małgorzata IWANEK
Beata KOWALSKA
Department of Water Supply and Wastewater Disposal
Faculty of Environmental Engineering
Lublin University of Technology
ul. Nadbystrzycka 40 B, 20-816 Lublin, Poland

Elżbieta HAWRYLUK
Kamil KONDRACIUK
graduates
Lublin University of Technology, Lublin, Poland

E-mails: M.Iwanek@wis.pol.lublin.pl, B.Kowalska@pollub.pl

Article citation info:
KOSOBUDZKI M, STAŃCO M. The experimental identification of torsional angle on a load-carrying truck frame during static and dynamic tests. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 285–290, http://dx.doi.org/10.17531/ein.2016.2.17.

Mariusz KOSOBUDZKI
Mariusz STAŃCO

THE EXPERIMENTAL IDENTIFICATION OF TORSIONAL ANGLE ON A LOAD-CARRYING TRUCK FRAME DURING STATIC AND DYNAMIC TESTS

IDENTYFIKACJA EKSPERYMENTALNA KĄTA SKRĘCENIA USTROJU NOŚNEGO POJAZDU PODCZAS TESTU STATYCZNEGO I DYNAMICZNEGO*

The underframe of a truck is one of the most loaded parts of a vehicle. It is a spatial unit and it must be strong enough to withstand random loading within many years of maintenance. The most severe form of deformation is in torsion. So, frame side members are often made from elements with channel sections, rigid for bending and flexible for torsion. Authors have conducted the research of 6x6 high mobility wheeled vehicle assigned to 20-feet container. Their load-carrying structure is made from two separate underframes: longitudinal and auxiliary connected with bolted joints. The goal of the research was to check if the torsional angle of deformation of the underframe during static and dynamic tests is within an acceptable range. The static test was carried out for the main underframe first to assess the characteristic of torsional stiffness without the auxiliary frame. After connecting both frames together the measure was conducted again. In the experiment the diagonal wheels were lifted up and the resulting displacement of the ends of the frame side members was recorded. Simultaneously the strain at chosen points of the underframe was measured with a system of turned half bridge strain gauges. After calibrating the measuring system a second part of experiment was conducted within proving ground tests when the vehicle was fully loaded. The collected strain data at chosen points allowed for calculating the resultant displacement of the ends of the frame side members in function of sort of road and to indicate the influence of auxiliary frame on increasing the torsional stiffness of the underframe.

Keywords: *underframe of vehicle, experimental measurements, numerical modeling, vehicle testing.*

Ustrój nośny pojazdu jest jednym z jego najbardziej obciążonych zespołów konstrukcyjnych. Jest to zespół o złożonej budowie przestrzennej, który musi być wystarczająco wytrzymały by wytrzymać zmienne obciążenia przez wiele lat eksploatacji pojazdu. Najbardziej obciążające są te obciążenia które wywołują skręcanie ustroju nośnego. Stąd ustrój nośny składa się najczęściej z podłużnic połączonych poprzeczami co w efekcie zapewnia dużą sztywność na zginanie i podatność na skręcanie. W artykule przedstawiono badania podwozia pojazdu kołowego wysokiej mobilności 6x6 przeznaczonego do połączenia z kontenerem 20-stopowym. Ustrój nośny pojazdu składa się z ramy głównej połączonej za pomocą połączeń podatnych z ramą pośrednią. Celem badań było sprawdzenie czy kąt skręcenia ustroju nośnego pojazdu w badaniach statycznych i dynamicznych nie wywołuje naprężeń wykraczających poza zakres dopuszczalny. Test statyczny został przeprowadzony najpierw tylko do ramy głównej w celu wyznaczenia jej sztywności skrętnej. Następnie ramy zostały połączone i wyznaczenie sztywności zostało powtórzone. W ramach testu koła znajdujące się w pojeździe po przekątnej zostały podniesione aż do utraty kontaktu z podłożem. Równocześnie rejestrowano przemieszczenie końców podłużnic ramy i odkształcenia w wybranych punktach, w których naklejono tensometry. Po skalibrowaniu układu pomiarowego przeprowadzono szereg testów przebiegowych z pojazdem całkowicie obciążonym ładunkiem. Zarejestrowane wartości odkształceń wykorzystano do wyznaczenia odkształcenia wypadkowego końców podłużnic ramy w funkcji rodzaju drogi oraz wpływu zamocowania kontenera na wypadkową sztywność skrętną ustroju pojazdu.

Słowa kluczowe*: ustrój nośny pojazdu, badania eksperymentalne, modelowanie numeryczne, testy pojazdu.*

1. Introduction

A loads that have an influence on a lorry during its ride are various in general and depends on a road condition [6,8,12]. The designer has been use the proper factor of safety that allows the structure of a vehicle to tolerate random loads if the level of loads are known in advance [7, 14]. So the sort of road for planned rides has to be established at the beginning of the design process. The factor of safety should be limited to avoid oversizing the structure. In this case the permissible sorts of road have to be limited as well. From this perspective the quality assessment of design the underframe of high mobility lorry was conducted [11]. The vehicle's chassis taken into investigation was 6x6 and was designed to carry the 20-feet container (Fig. 1). The rigid container was design to be connected with the chassis in four points with use an quick disconnect coupling. Because of that the susceptibility of underframe to torsion was significantly limited.

The lorry was built by JELCZ-KOMPONENTY Ltd. on high-mobility, heavy-weights, all-wheel-drive, 3-axle chassis version. The vehicle is adjusted to drive on and off roads and carry a 20-feet container. The wheel base is 4400 + 1400 [mm]. The approach and departure angles are as follow: 36 and 29 [°]. The main frame of the chassis is made from two longitudinal members of a frame with channel sections with additional pads attached to the longitudinal and seven crossbars with open and close sections. The torsional section modulus is J_m = 363 [cm^3]. The container is a rigid spatial object, so only in the middle part of a frame the torsional deformation in elastic range should be accessible. To make this real the tubular crossbeams were used (Fig. 2).

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

Fig. 1. The general view of investigated vehicle

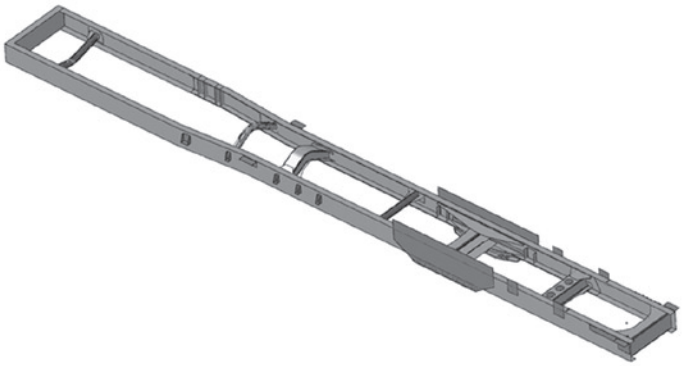


Fig. 2. Main chassis frame of the vehicle

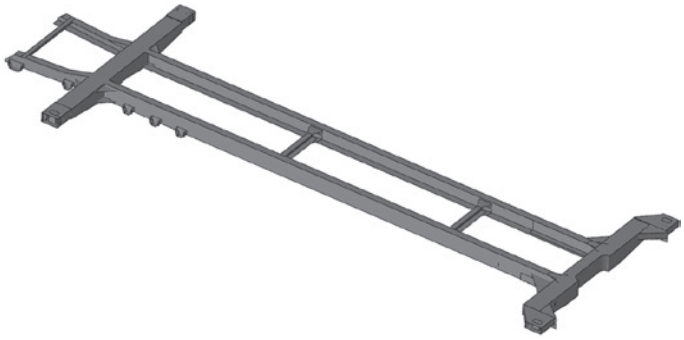


Fig. 3. The auxiliary frame

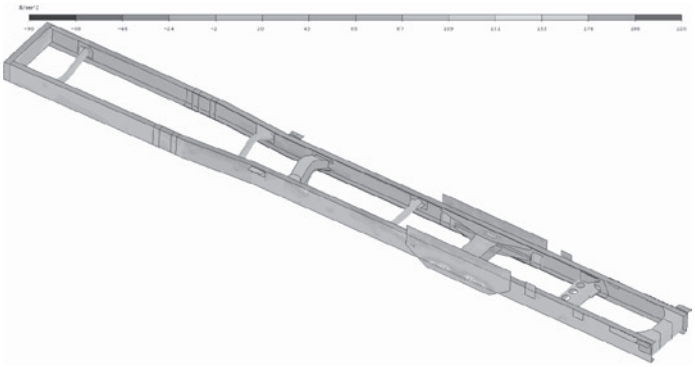


Fig. 4. Principal stress value within a main frame

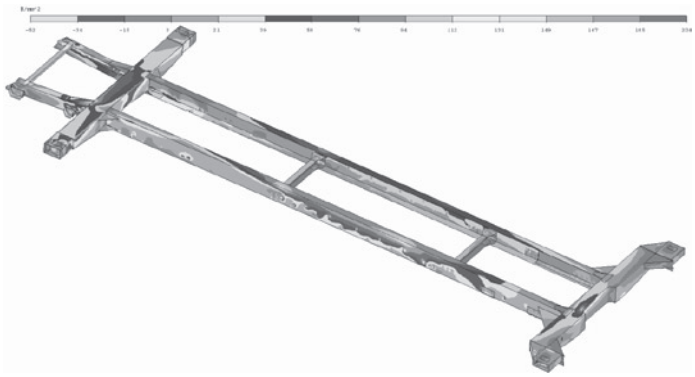


Fig. 5. Principal stress value within a auxiliary frame

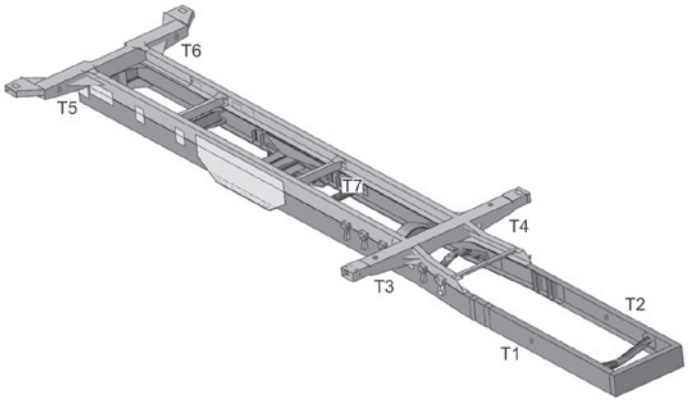


Fig. 6. Points of location the gauges on main and auxiliary frame

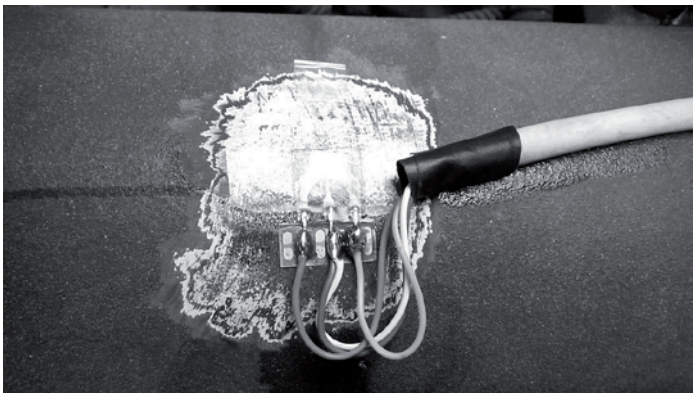


Fig. 7. Strain gauge in measure point T7

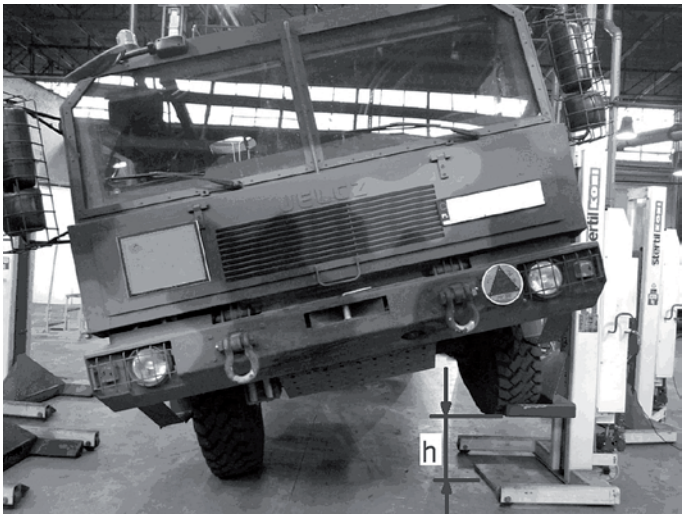


Fig. 8. A view of vehicle during a measurement

To make the part of a main frame rigid enough to join it with the container the auxiliary frame was created which the longitudinal members were with channel section (Fig. 3). Ends of longitudinal members of auxiliary frame are jointed to each other with bolts. The close-section beams have been specially designed to achieve a nonlinear section modulus of bending. This auxiliary frame increases the torsion section modulus by J_a = 149 [cm^3]. Those frames have been connected to each other with use the flexible joints in the front and rigid in the rear part of the auxiliary frame.

2. Selection of measurement points

Firstly, in order to determine the measurement points of deformation of jointed frames the analysis of technical documentation was conducted. Any points were torsion or bending appears were selected. Then a shell FEM model [10, 15] of frames with a 256000 nodes of freedom was created and analysed. As a result of the analysis, the zones of uniform main stresses were found [5]. An examples of stress value within a main and an auxiliary frame were presented in figure 4 and 5 accordingly.

Finally, the measurements of deformation were conducted at the seven points presented in figure 6. The three of them (T1, T2, T7) were placed on the main frame and the others (T3, T4, T5, T6) on the auxiliary frame [1, 3, 4] (Fig. 6).

3. Measurement's arrangement

To measure the deformation of frames the electric resistance wire strain gauges TFpxy-5/350 and TFx-6/350 were used [2] (Fig. 7). The resistance of them was R = 350,5 [Ω] ± 0,25 [%], with a constant k_{sg}=2,15±0,5 [%]. As a consequence of the use of inverted half bridge strain gauges is a decrease of their sensitivity w_{CH} caused by a resistance of wires. This quantity was taken into account and a range of stress $\Delta\sigma_i$ was calculated as follow:

$$\Delta\sigma_i = \Delta\varepsilon_i \cdot E \tag{1}$$

$$\Delta\varepsilon_i = \frac{4 \cdot \Delta U_i}{U_{zm} \cdot w_{CH} \cdot k_{sg}} \tag{2}$$

where: $\Delta\varepsilon_i$ – range of recorded strain, E – Young's module (E = 210 [GPa]), ΔU_i - range of voltage [V], U_{zm} - power supply voltage [V], k_{sg} – constant of gauge, w_{CH} – sensitivity of gauge.

4. Experimental results

The experimental measures of torsion on the underframe of the analyzed vehicle were investigated in three steps [9, 13, 16]:

Static measures:
- the main frame only – step 1,
- the joined frames – step 2,
- Dynamic measure:
- the joined frames during a ride with a container – step 3.

As an external force in static measures a column hydraulic elevators were used that lifted a wheel or wheels until the unlifted wheels had a contact with a ground. During an experiment the front part of an underframe was loaded by a cabin and a rear part was loaded by a mass of a lifted axles, suspension and wheels (Fig. 8).

4.1. Static measurements of main underframe

Within a static measures of stress caused by a torsional deflection of underframe wheels of front and rear wheels were lifted up to a height of 600 [mm]. The algorithm of measure was as follows:

Table 1. The value of torsional angle of main underframe of vehicle (α – angle of twist the front bumper of a vehicle, β - angle of twist the rear bumper of a vehicle, FL – a front left wheel lifted 600 [mm], FR – a front right wheel lifted 600 [mm])

Nr of measure	α[deg]	β[deg]	α+β[deg]	Notice
1	-15.99	0	15,99	only FL lifted
2	1.34	18.33	16,99	only FR lifted
3	8.31	12.1	20,41	lifted LF and RR

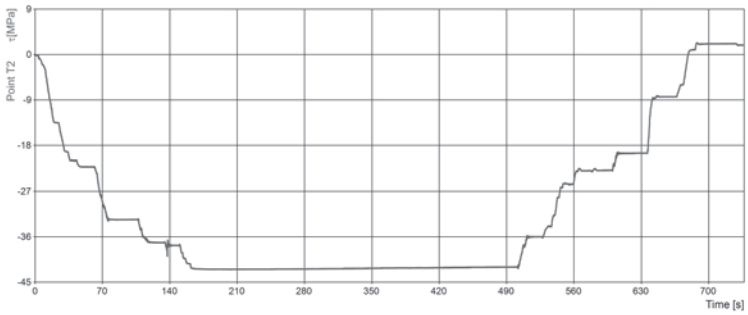


Figure 9. The range of shear stress in point T1

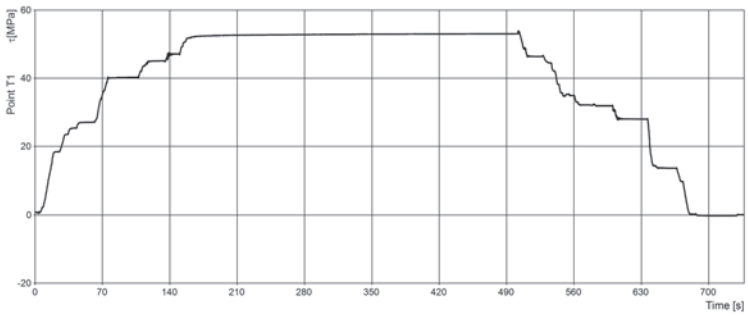


Figure 10. The range of shear stress in point T2

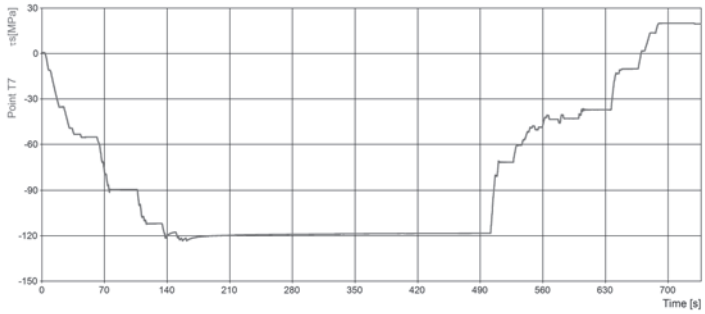


Figure 11. The range of torsional stress in point T7

- front left wheel was lifting up to a 600 [mm],
- the middle right wheel was lifting up to 600 [mm] when a front left wheel was lifted,
- rear right wheel was lifting up to 600 [mm] when a front and a middle wheel were lifted,
- finally all wheels were lowered to the start position.

Then the measure started from lifting a front rear wheel according to the algorithm described above. The main goal of experiment was to determine the maximum angle of torsion of the main underframe without an auxiliary frame. Results of the tests are presented in table 1.

Table 2. The list of maximum sheer stress in chosen points of twisted jointed frames

Nr of measure	Point T1 τ[MPa]	Point T2 τ[MPa]	Point T7 τ[MPa]
1	38	−22	−85
2	43	−39	−111
3	53	−42	−118

The range of a shear stress recorded in point T1, T2 and torsional stress in point T7 (measure nr 3) is presented in Fig. 9-11.

Maximum value of twist angle of main underframe was 20,4 [deg], and a maximum sheer stress in points T1, T2 and T7 was as follow: 83, −42 and 118 [MPa]. The list of recorded values of stress was presented in table 2.

The calculated twist susceptibility of a jointed frames between front bumper and the last rear crossbar was 2,1 [deg/m].

4.2. Static measurements of jointed main and auxiliary frames

The next part of experiment was to measure the stress caused by a torsional deflection of the jointed main and auxiliary frame in eight points placed in the front, middle and rear part of jointed frames. The positions of gauges are presented in figure 12. Points A and B were placed on the front metal bumper, C, D, G and H were placed on the rear container beam and E and F were placed on the front container beam. The algorithm of measure was described in paragraph 4.1. The vehicle was unloaded. Results of tests are presented in table 3.

The total maximum twist angle was decreased by 37% in comparison to the case described in point 4.1 because of the auxiliary frame that was twisted of max. 6,8 [deg]. The twist angle between front bumper and front container beam was max. 5,4 [deg]. The main frame was twisted in maximum range near the front container beam. That was caused because the engine rigid for twisting is jointed to the frame in the front part of it. The maximum twist angle of auxiliary frame was 9,4 [deg] when only one front wheel is lifted. The front container beam was loaded not only by bending, but also by torsion. So, the joint of frames should be rigid at the end of frames and susceptible in the middle zone (connection between front container beam and main frame).

Table 3. The value of torsional angle of jointed main and auxiliary underframe of vehicle (α – angle of twist the front bumper of a vehicle, β – angle of twist the rear bumper of a vehicle, γ – angle of twist the front container beam, δ – angle of twist the rear container beam, FL – a front left wheel lifted 600 [mm], FR – a front right wheel lifted 600 [mm])

Nr	α [deg]	β [deg]	γ [deg]	δ [deg]	α– β [deg]	γ– δ [deg]	
1	−13.2	−0.9	−8	1.4	−12.3	−9.4	LF↑
2	4.7	15	−8.8	14.5	−10.3	−5.7	RR↑
3	−3.6	9.3	1.8	8.6	−12.9	−6.8	LF↑ –RR↑

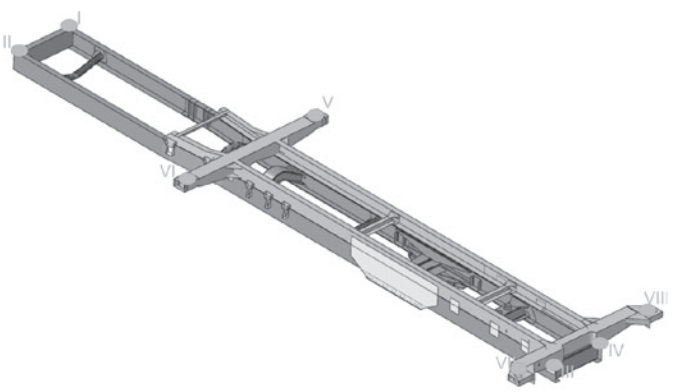


Fig. 12. Points of measurement the twist angle

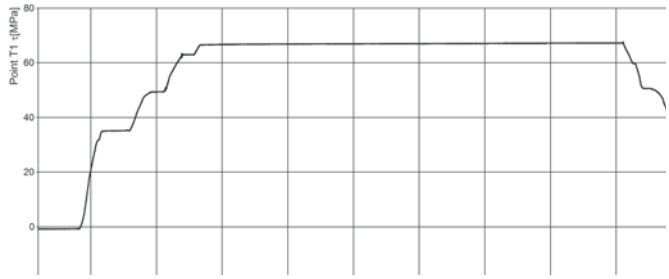


Fig. 13. The changes of shear stress in point T1 when frames were maximally twisted

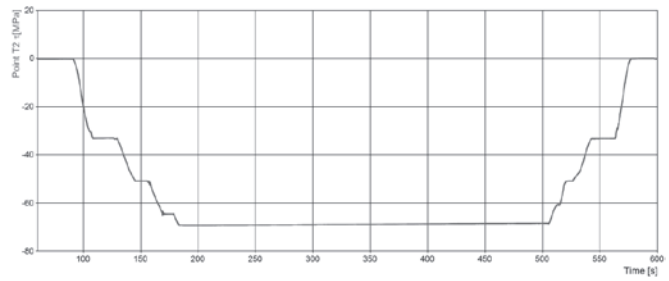


Fig. 14. The changes of shear stress in point T2 when frames were maximally twisted

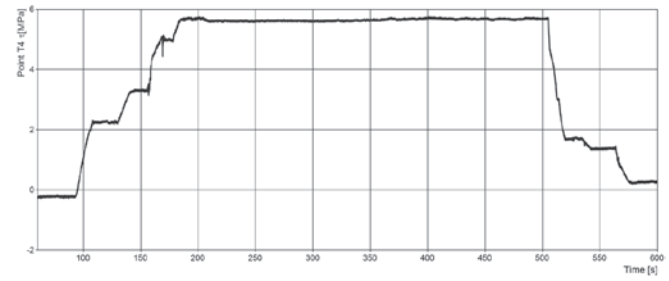


Fig. 15. The changes of shear stress in point T4 when frames were maximally twisted

The stress changes caused by twisting crossbar are presented in figures 13÷16.

The maximum values of stress in chosen points are presented in table 4.

Recorded value of stress at points T1 and T2 was higher by about 50 [%] then when only the main frame was twisted. In point T7 the value of stress was smaller by about 35 [%]. This is the reinforcement effect of auxiliary frame. The twist susceptibility of jointed frames between the front and rear container

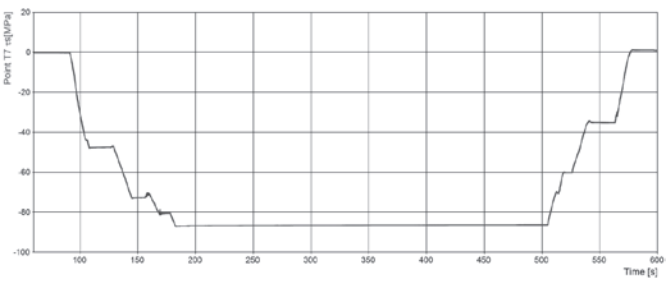


Fig. 16. The changes of torsional stress in point T7 when frames were maximally twisted

Table 4. Examples of stress value in chosen points when frames were maximally twisted

Nr of measurement	Point T1 τ[MPa]	Point T2 τ[MPa	Point T4 τ[MPa	Point T7 τ[MPa
1	59	−56	5	−81
2	59	−61	4	−65
3	67	−69	6	−86

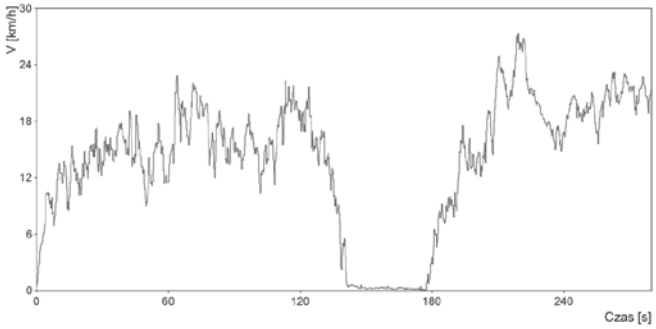


Fig. 17. An example of speed variability in off-road test

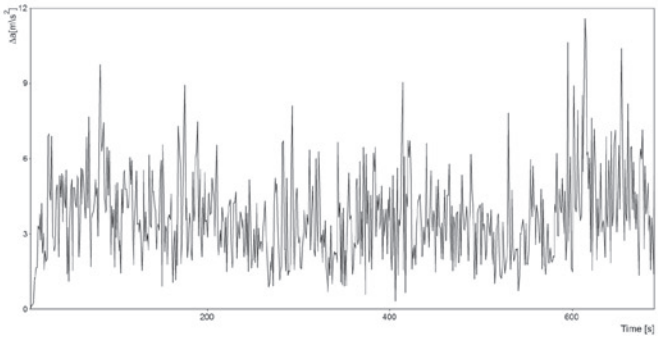


Fig. 18. The outcome value of vertical acceleration of container

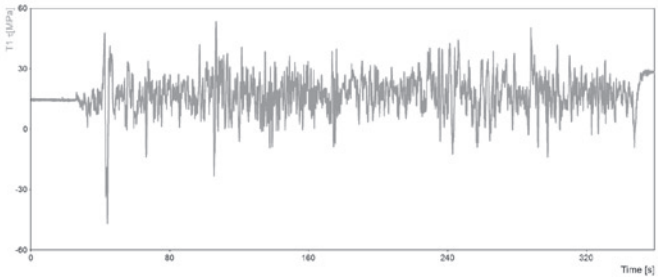


Fig. 19. Value of τ in point T1

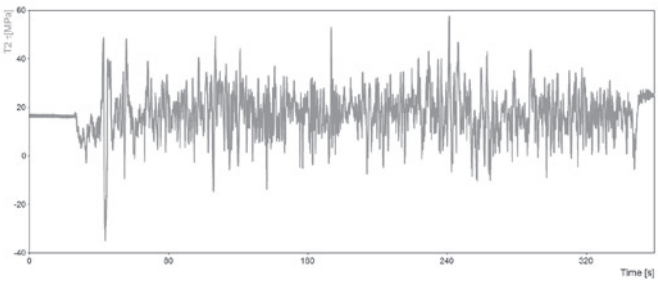


Fig. 20. Value of τ in point T2

beam was 1,3 [deg/m]. The total twist susceptibility of the jointed frames between front and rear bumper was 3,8 [deg/m].

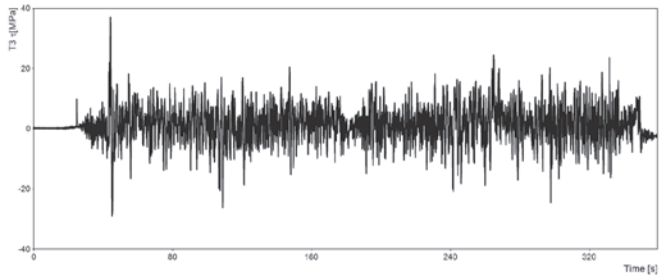


Fig. 21. Value of τ in point T3

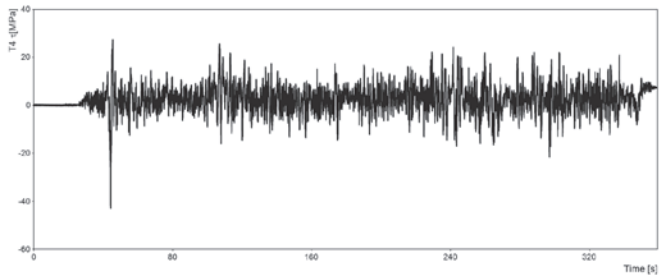


Fig. 22. Value of τ in point T4

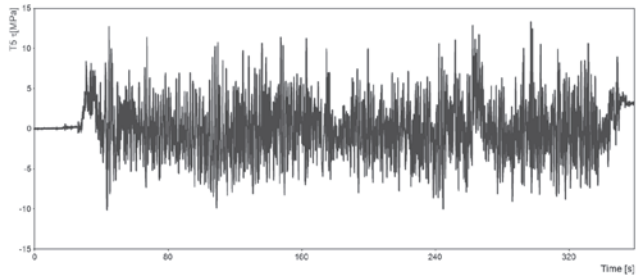


Fig. 23. Value of τ in point T5

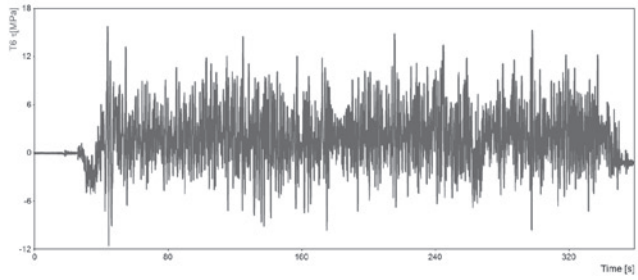


Fig. 24. Value of τ in point T6

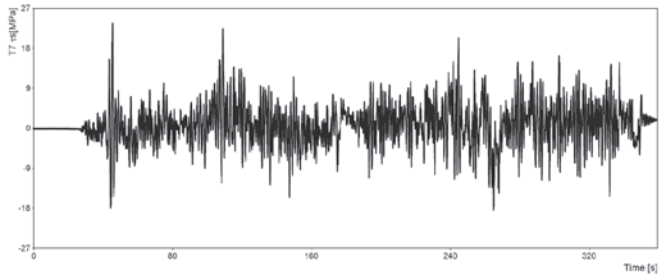


Fig. 25. Value of τ in point T7

4.3. Dynamic measurements of jointed main and auxiliary frames

Measurements of acceleration during the off-road ride with the vehicle at eleven points were conducted. The vehicle in the first part of dynamic test was without container and in the second part was with 20-feet container fully loaded. The total mass of container was

Table 5. En example of maximum, minimum and the range of sheer stress value during off-road ride

Measurement point	τ_{min} [MPa]	τ_{max} [MPa]	$\Delta\tau$ [MPa]
T1	–47	53	100
T2	–35	58	93
T3	–29	37	66
T4	–43	27	70
T5	–10	13	23
T6	–12	16	28
T7	–18	24	42

8.72Mg. The rigid container was connected to the chassis at four points by quick disconnect couplings. The stresses at chosen points of frames pointed in paragraph 3 were conducted as well as acceleration in eleven points of the vehicle in three dimensions.

In off-road condition the accelerations were measured when the vehicle was moving at an average speed of 16 [km/h]. An example of speed variability was presented in figure 17.

The outcome value of vertical acceleration recorded in four upper corner of container was up to 11,3 [m/s^2] and side acceleration was up to 4,8 [m/s^2]. An example of vertical accelerations was presented in figure 18.

An examples of stress value in chosen points of jointed frames were presented in figures 19 ÷ 25. The list of recorded sheer stress value for off-road ride is presented in table 5.

The range of twist angle of the auxiliary frame was very small because the container helps make the frame rigid. So, the connections between the main and auxiliary frame are prone to lengthening in the elastic range of deformation and this must be taken into account.

5. Conclusions

As pointed out above the range of twist angle of jointed frames depends on the stiffness of a body. The total twist angle of the main frame was 20,4 [deg] and was twice larger than the twist angle of jointed frames. This information is crucial because the vehicle is a high mobility class and in this case the jointed frames should be rather flexible than rigid [6,8]. So, if the twist stiffness of jointed frames between the front and rear container beam is quite high, the value of sheer stress in the front part of jointed frames is getting higher. That was confirmed in the tests and a necessary change of joining the two frames had to be done.

References

1. Bocian M., Jamroziak K., Kulisiewicz M., The identification of nonlinear damping of the selected components of MDOF complex vibratory systems, Proceedings of the 9th International Conference on Structural Dynamics, EURODYN 2014 Porto, Portugal, 30 June - 2 July 2014, A. Cunha, E. Caetano, P. Ribeiro, G. Müller (eds.) Book Series: EURODYN-International Conference on Structural Dynamics 2014: 3365- 3372.
2. Buchacz A., Płaczek M., Damping of mechanical vibrations using piezoelements, including influence of connection layer's properties on the dynamic characteristic. Solid State Phenomena 2009: 147-149: 869-875, http://dx.doi.org/10.4028/www.scientific.net/SSP.147-149.869.
3. Buchacz A., Płaczek M., The analysis of a composite beam with piezoelectric actuator based on the approximate method. Journal of Vibroengineering 2012: 14; 111-116.
4. Buchacz A., Płaczek M., The approximate Galerkin's method in the vibrating mechatronic system's investigation, New face of TMCR proceedings Book series. Proceedings of the International Conference ModTech 2010: 147-150.
5. Głowacz A., Głowacz A., Korohoda P., Recognition of monochrome thermal images of synchronous motor with the application of binarization and nearest mean classifier. Archives of Metallurgy and Materials 2014: 59(1); 31-34, http://dx.doi.org/10.2478/amm-2014-0005.
6. Jamroziak K., Kosobudzki M., Determining the torsional natural frequency of underframe of off-road vehicle with use of the procedure of operational modal analysis. Journal of Vibroengineering 2012: 14(2): 472-475.
7. Lee Y. L., Pan J., Hathaway R., Barklay M., Fatig testing and analysis. Theory and practice. Burlington: Elsevier-Heinemann, 2005.
8. Reimpell J., Betzler J., Podwozia samochodów. Podstawy konstrukcji. Warszawa: WKŁ, 2008.
9. Rusiński E., Czmochowski J., Kowalczyk M., Pietrusiak D., Numerical and experimental modal modes identification methods of steel construction. Journal of Science of the Gen. Tadeusz Kosciuszko Military Academy of Land Forces 2010: 4; 208-218.
10. Rusiński E., Czmochowski J., Smolnicki T. Advanced Finite Element Method, Wroclaw: Wroclaw University of Technology, 2000.
11. Rusiński E., Koziołek S., Jamroziak K., Quality assurance method for the design and manufacturing process of armoured vehicles. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2009: 3; 70-77.
12. 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.
13. Smolnicki T., Karlinski J., Derlukiewicz D., Identyfication of internal stress in bolted flanged joints. Solid State Phenomena 2010: 165: 352-358, http://dx.doi.org/10.4028/www.scientific.net/SSP.165.353.
14. Wong J. Y., Terramechanics and off-road vehicle engineering. Terrain behaviour, off-road vehicle performance and design. Butterworth-Heinemann: Elsevier, 2010.
15. Zienkiewicz O.C., Taylor R.L. The Finite Element Method, London: McGraw Hill Book Company, 1991.
16. Żółkiewski S., Numerical application for dynamical analysis of rod and beam systems in transportation. Solid State Phenomena 2010: 164: 343-348, http://dx.doi.org/10.4028/www.scientific.net/SSP.164.343.

Mariusz KOSOBUDZKI
Logistics Departament
Gen. Tadeusz Kosciuszko Military Academy of Land Forces
ul. Czajkowskiego 109, 51-150 Wrocław, Poland
E-mail: m.kosobudzki@wso.wroc.pl

Mariusz STAŃCO
Department of Machine Design and Research
Wroclaw University of Technlogy
ul. Lukasiewicza 5 B-5, 50-370 Wrocław, Poland
E-mail: mariusz.stanco@pwr.edu.pl

Article citation info:
WANG H, DENG G, LI Q, KANG Q. Research on bispectrum analysis of secondary feature for vehicle exterior noise based on nonnegative tucker3 decomposition. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 291–298, http://dx.doi.org/10.17531/ein.2016.2.18.

Haijun WANG
Guoyong DENG
Qinglin LI
Qiang KANG

RESEARCH ON BISPECTRUM ANALYSIS OF SECONDARY FEATURE FOR VEHICLE EXTERIOR NOISE BASED ON NONNEGATIVE TUCKER3 DECOMPOSITION

BADANIA NAD ANALIZĄ BISPEKTRUM CECH DRUGORZĘDNYCH HAŁASU ZEWNĘTRZNEGO POJAZDÓW W OPARCIU O NIEUJEMNĄ DEKOMPOZYCJĘ TUCKERA3

Nowadays, analysis of external vehicle noise has become more and more difficult for NVH (noise vibration and harshness) engineer to find out the fault among the exhaust system when some significant features are masked by the jamming signals, especially in the case of the vibration noise associating to the bodywork. New method is necessary to be explored and applied to decompose a high-order tensor and extract the useful features (also known as secondary features in this paper). Nonnegative Tucker3 decomposition (NTD) is proposed and applied into secondary feature extraction for its high efficiency of decomposition and well property of physical architecture, which serves as fault diagnosis of exhaust system for an automobile car. Furthermore, updating algorithm conjugating with Newton-Gaussian gradient decent is utilized to solve the problem of overfitting, which occurs abnormally on traditional iterative method of NTD. Extensive experimen results show the bispectrum of secondary features can not only exceedingly interpret the state of vehicle exterior noise, but also be benefit to observe the abnormal frequency of some important features masked before. Meanwhile, the overwhelming performance of NTD algorithm is verified more effective under the same condition, comparing with other traditional methods both at the deviation of successive relative error and the computation time.

Keywords: *feature extraction, vehicle exterior noise, NTD, updating algorithm.*

Obecnie inżynierowie NVH (zajmujący się problematyką hałasu, drgań i uciążliwości akustycznych) napotykają na coraz większe trudności przy analizie hałasu zewnętrznego pojazdów wynikające z faktu, że istotne cechy związane z nieprawidłowościami układu wydechowego są maskowane przez sygnały zakłócające, szczególnie hałas wibracyjny związany z pracą nadwozia. Niezbędna jest zatem nowa metoda, która pozwoli rozkładać tensory wysokiego rzędu i wyodrębniać przydatne cechy (zwane w tym artykule także cechami drugorzędnymi). Do ekstrakcji cech drugorzędnych wykorzystano w prezentowanej pracy metodę nieujemnej faktoryzacji tensorów znaną także jako nieujemna dekompozycja Tuckera 3 (NTD) , która cechuje się wysoką efektywnością dekompozycji i może być wykorzystywana w diagnostyce uszkodzeń układu wydechowego samochodów. Problem nadmiernego dopasowania, który występuje w tradycyjnej metodzie iteracyjnej NTD rozwiązano przy pomocy algorytmu aktualizacyjnego sprzężonego z gradientem prostym Newtona-Gaussa. Wyniki doświadczeń pokazują, że bispektrum cech drugorzędnych nie tylko pozwala doskonale interpretować stan hałasu zewnętrznego pojazdu, ale również umożliwia wykrywanie wcześniej maskowanych nieprawidłowych częstotliwości odpowiadających niektórym ważnym cechom. Badania potwierdzają, że algorytmu NTD jest bardziej efektywny, w tych samych warunkach, w porównaniu z innymi tradycyjnymi metodami zarówno w zakresie odchyleń błędu względnego jak i czasu obliczeń.

Słowa kluczowe*: ekstrakcja cech, hałas zewnętrzny pojazdu, NTD, algorytm aktualizacyjny.*

1. Introduction

Recently, nonnegative Tucker3 decomposition (NTD), also known as a generalized form of nonnegative tensor factorization (NTF) model, has received hot attention by a considerable amount of people for its overwhelming performance of decomposition efficiency on multi-way dataset decomposition [1]. Both NTD and NTF here can be viewed as high-order extensions of non-negative matrix factorization (NMF) method as referred in some literatures about relationship between NTF/NTD and NMF, all of whose factors are based on an alternating minimization of cost functions incorporating distances or divergences measures with its application in environmental data analysis and can be found in reference therein [2]. NTF with its form of hierarchical alternative least square (ALS) is applied into blind signal separation, of whose alpha and beta divergences methods both are involved in as well [3]. Then much more reseach has been done with NTF itself and gradually evolved to be NTD model.

Mostly, NTD is explored and applied to decompose larger-scale tensors in data analysis. Just one of crucial usages of NTD is of feature extraction from high-order datasets ranging from signals, images, speech, neuroscience, systems biology, chemometrics, or texts [4-10], and also used for designing complex systems as it is the case of wireless communication systems as the publication of the novel paper [11]. However, some useful features usually masked by jamming signals are still a serious problem to judge one state. For example, overfitting appearing in the iterative procedure would generate a large number of harmonic waves as interrupting signal before feature extraction [12], which brings in hard difficulty in detecting the useful features for fault diagnosis and usually occurs in the frequency analysis of vehicle exte-

rior noise. Thus, how to avoid overfitting and extract the secondary features to interpret the special state in physical property has become an extremely urgent scheduler for NTD [13].

Physical singal existing in the real-world data is non-negative. Incorporating constraints such as sparseness, smoothness or orthogonality on NTD have been the object of significant works for feature extraction during the last years [14, 15]. Actually, NTD for multi-way data analysis results from the large volume of current data to be analyzed under non-negative constraint on the factors of Tucker3 model for the secondary features to be estimated as well, when only non-negative parameters are physically interpretable [16]. The bispectrum feature of vehicle exterior noise is similar to be non-negative in itself. The hybrid noise by exhaust system is not easy to be certained only depending on the frequency existing on the exhaust pipe, since some coupled noise such as conjugating engine noise with resonant noise of transmission shaft always arises in the process of transferring. So using NTD methodology to extract the secondary features is necessary and practicable in fault diagnosis of exhaust system, whose bispectrum directly extracted from the original noise without interference will be a vital way for NVH engineer. Secondary features extracted by NTD are of physical sparseness in data analysis of vehicle exterior noise and to be shown in the later experiments.

Besides, non-negative constraint forced on all the factors of Tucker3 decompositon is able to radically solve the problem of iteration converge in the calculation procedure, or overcome the overfitting in the case of a large-scale tensor decomposition. Meanwhile, NTD may allow to relax the traditional updating form, and to develop a specialized updating algorithm that improves the performance both in terms of accuracy and computational cost, since it just lends itself to the iteration in the form of Newton-Gaussian gradient descent (NGGD). In fact, NGGD can be developed as a way of updating the factors all-at-once as well. This way will be not only used to reduce the complexity of iterative calculation, but be also an available solution to the robustness of NTD and more significant to the matrices and core tensors, which are crucial to the basis images for reconstructing the secondary feature to analyze the vehicle exterior noise of an automobile car.

2. NTD algorithm

2.1. Definition and notation

Several expressions are necessary to Tucker3 algorithm dealing with real dataset referred to throughout the paper. Meanwhile, some reviews are made for the notation and definitions that will be used as well. To facilitate the distinction between scalars, vectors, matrices, and higher order tensors, the type of a given quantity will be reflected by its representation: scalars are denoted by italic scripts, e.g., α, μ; vectors are written as bold italic latin lower-case letters, e.g., $\boldsymbol{a}$, $\boldsymbol{b}$; matrices correspond to bold italic latin capital letters, e.g., $\boldsymbol{A}$, $\boldsymbol{B}$; and tensors are written as bold italic euclid letters, e.g., $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{Y}}$. The Frobenius norm of a tensor $\boldsymbol{\mathcal{Y}}$ is denoted by:

$$\|\boldsymbol{\mathcal{Y}}\|_{\mathrm{F}} = \sqrt{\langle \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{Y}} \rangle}\,. \tag{1}$$

where $\|\bullet\|_{\mathrm{F}}$ denotes frobenius norm, which can be find in [17].

Definition (Matricization) Matricization, also known as unfolding, is the process of rearranging an N-way dataset as a set of matrices over product. For example, a tensor $\boldsymbol{\mathcal{Y}} \in \mathbf{R}^{I_1 \times I_2 \times \cdots \times I_N}$, of the partitioning of the set $\{I_n, n: \in 1,2,\cdots,N\}$, is matricized into two ordered subsets $\mathbf{R}_1$ and $\mathbf{R}_2$, there exists a subset $\{I_n, n: \in 1,2,\cdots,N\}$ of dimension indices with the length p and N-p, respectively, the matricization of Nth-order tensor can be described as:

$$\boldsymbol{Y}_{\mathbf{R}_1;\mathbf{R}_2} = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} \boldsymbol{y}_{i_1,\cdots,i_N} \left(\underset{n \in \mathbf{R}_1}{\otimes \boldsymbol{e}_{i_n}^{(n)}} \right) \left(\underset{n \in \mathbf{R}_2}{\otimes \boldsymbol{e}_{i_n}^{(n)}} \right)^T \in \mathbf{R}^{R_1 \times R_2}\,, \tag{2}$$

$$\text{s.t.} \qquad R_1 = \prod_{n=1}^{p} \mathrm{I}_n, R_2 = \prod_{n=N\text{-}p}^{N} I_n\,.$$

Where symbol $\otimes$ denotes kronecker product; $\boldsymbol{e}$ is a unit vector; round bracket $(\ )^T$ denotes that it returns a permutation vector or matrix.

As the multi-way dataset concered in practical application, it must be transited into a tensor. So the third-order tensor is considered herein and involved in the behind. For instance, matricizing a third- order tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ along each mode, three matrices can be obtained and expressed as following, respectively called horizontal, lateral, and frontal matrix slices (see reference [18]):

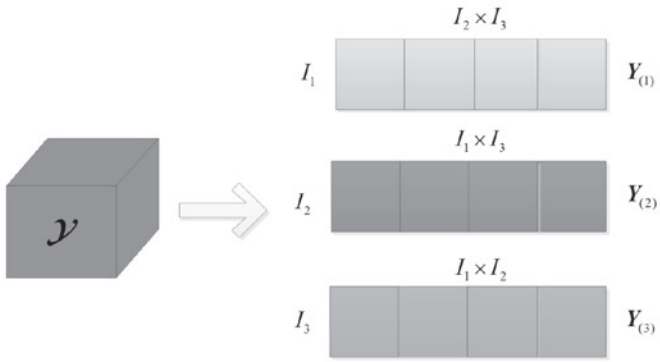


Fig. 1. Model of tensor matricization

$$\mathbf{Y}_{(n)} \in \mathbb{R}^{\mathrm{I}_n \times \mathrm{I}_s}\,, \mathrm{I}_n := \{I_1, I_2, I_3\},\ \mathrm{I}_s := \prod_{s \neq n} I_s\,, \mathrm{n} = \{1,2,3\}\,. \tag{3}$$

Furthermore, the three diffirent matrix slices of Eq.(3) can be illustrated as Fig. 1.

2. 2. Tucker model

Considering an N-way tensor $\boldsymbol{\mathcal{Y}} \in \mathbf{R}^{I_1 \times I_2 \times \cdots \times I_N}$, the generalized approximant of the Tucker model is presented as follows:

$$\boldsymbol{\mathcal{Y}} = \widehat{\boldsymbol{\mathcal{Y}}} + E \approx \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{A}^{(1)} \times_2 \boldsymbol{A}^{(2)} \cdots \times_N \boldsymbol{A}^{(N)},\ \ or\ \ \mathbf{Y}_{(n)} = \boldsymbol{A}^{(n)} \boldsymbol{G}_{(n)} (\boldsymbol{A}^{\otimes_{-n}})^T. \tag{4}$$

$$s.t\ \ \boldsymbol{A}^{\otimes_{-n}} = \boldsymbol{A}^{(N)} \otimes \cdots \boldsymbol{A}^{(n+1)} \otimes \boldsymbol{A}^{(n-1)} \otimes \cdots \otimes \boldsymbol{A}^{(1)};\ \{\boldsymbol{A}\} = \{\boldsymbol{A}^{(1)}, \cdots, \boldsymbol{A}^{(N)}\};\ n \leq N;$$

Where $\widehat{\boldsymbol{\mathcal{Y}}}$ is an approximation of the real-valued $\boldsymbol{\mathcal{Y}}$, symbol $\times_n$ denotes the product between mode-n matrix and tensor; $\boldsymbol{\mathcal{G}} \in \mathbf{R}^{J_1 \times J_2 \times \cdots \times J_r}$ is a core tensor in the Tucker3 model, and the parameters meet $J_1 \leq I_1, J_2 \leq I_2, \cdots, J_r \leq I_n, r \leq n \leq N$. The Tucker model of Eq. (4) can be rewritten in an element-wise form as:

$$\boldsymbol{y}_{i_1,\cdots,i_N} = \sum_{j_1=1}^{J_1} \cdots \sum_{j_r=1}^{J_r} \boldsymbol{g}_{j_1\cdots j_r} \prod_{n=1}^{N} \boldsymbol{a}_{i_n,j_r}^{(n)} \quad s.t. \quad \boldsymbol{a}_{i_n,r_n}^{(n)} \in \boldsymbol{A}^{(n)}, i_n \le I_n. \tag{5}$$

Where $\boldsymbol{g}_{j_1\cdots j_r}$ and $\boldsymbol{y}_{i_1,\cdots,i_N}$ are two different elements of the tensors $\mathcal{G}$ and $\mathcal{Y}$, respectively. Physical model of feature extraction consisting of several sub-tenors by NTD method can be represented as Fig. 2.

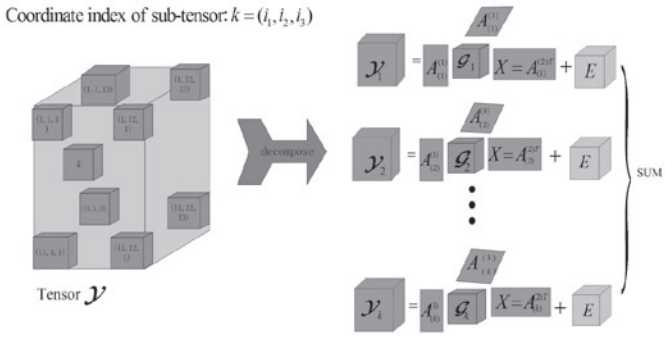


Fig. 2. Feature extraction by Tucker3 model for sub-tensors

Choosing $N^{th} \le 3$ as the order of Eq. (4) and Eq.(5) in the real applications. Therefore, the secondary sub-features extracted by Tucker3 algorithm to a third-order dataset are expressed as:

$$\boldsymbol{y}_{i_1,i_2,i_3} = \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \sum_{j_3=1}^{J_3} \boldsymbol{g}_{j_1 j_2 j_3} \prod_{n=1}^{3} \boldsymbol{a}_{i_n,j_r}^{(n)}, \quad \text{s.t.} \quad \mathrm{j_r = j_1, j_2, j_3 \le \min\{I_n\},\ i_n \le I_n, n \le N} \tag{6}$$

Some updating algorithms of Tucker3 decomposition for iterative calculation can be found in [19]. However, traditional updating algorithm usually takes much more computer cost when updating all the factors of NTD algrorithm, especially used to a large-scale tensor. Thus, a novel updating form will be explored for resolving this problem in the following section.

3. Updating algorithm of iterative calculation

The method of iteration algorithm via ALS with the way of calculation one-by-one has the advantages of simple mathematical model and lower computer storage requirement, but it has to solve the problems of slow convergence and overfitting in the computation procedure [20]. Herein, the solution of computing the factors all-at-once is applied to overcome these problems.

3.1. Updating algorithm based on NGGD

A real-value tensor $\mathcal{Y} \in \mathrm{R}^{\mathfrak{L}}$ is decomposed into N mode matrices $\mathbf{A}^{(n)} \in \mathrm{R}^{I_n \times J_n}$ and a core tensor $\mathcal{G} \in \mathrm{R}^{J_1 \times J_2 \times \cdots \times J_N}$. Then all the factors are integrated into a global matrix $\boldsymbol{M} = (\boldsymbol{A}^{(1)\mathrm{T}}, \cdots, \boldsymbol{A}^{(N)\mathrm{T}}, vec(\mathcal{G}))$. Operator $vec(\cdot)$ means a tensor $\mathcal{Y}$ stacks its column into a matrix $\boldsymbol{M}$, which is also known as matricization mentioned above. Besides, the Hessian matrix is often utilized to remedy data overfitting in the calculation procedures.

Therefore, the simultaneously updating algorithm based on the Gauss-Newton gradient descent is expressed in a common formula as:

$$\mathbf{M}_+ = (\mathbf{M} - \mathbf{H}^{-1}\mathbf{G})_+, \tag{7}$$

Where subscript + means adding nonnegative constraint on $\mathbf{M}$. The gradient $\mathbf{G}$ and the approximation Hessian matrix $\mathbf{H}$ are respectively computed by:

$$\mathbf{G} = \mathbf{K}^{\mathrm{T}}(\hat{\mathbf{Y}} - \mathbf{Y}),\ \mathbf{H} = \mathbf{K}^{\mathrm{T}}\mathbf{K}, \tag{8}$$

$$s.t. \quad \mathbf{K} = [\mathbf{K}_1, \mathbf{K}_2, \cdots, \mathbf{K}_{N+1}],\ \mathbf{K} \in \mathrm{R}^{\prod I_n \times \sum_{n=1}^{N} R_n I_n},$$

$$\mathbf{K}_n = \begin{cases} \mathbf{P}_n^{\mathrm{T}}(\{\mathbf{A}\}_{-n} \mathrm{G}_{(n)}{}^{\mathrm{T}} \otimes \mathbf{I}_{I_n}), & n = 1, 2, \cdots, N, \\ \{\mathbf{A}\}, & n = N+1. \end{cases}$$

Where $\boldsymbol{Y} = vec(\mathcal{Y})$, $\mathcal{Y} \in \mathbf{R}^{\mathfrak{L}}$; $\boldsymbol{P}$ is a permutation matrix and $\boldsymbol{K}$ is a Jacobian matrix which can be directly utilized in the iteration Eq. (7); $\{\boldsymbol{A}\}_{-n}$ denotes the Kronecker products of all mode matrices except $\boldsymbol{A}^{(n)}$. However, in the iteration procedure the Hessian matrix $\boldsymbol{H}$ is possible to achieve null point, which can easily cause overfitting and spontaneously enlarge the mean square error [17]. In the sequence, a solution for the iteration Eq. (8) will be presented.

For an Hessian matrix $\boldsymbol{H}$ closing to null point, take an approximate Hessian $\widehat{\boldsymbol{H}}$ instead of $\boldsymbol{H}$, that is, $\widehat{\boldsymbol{H}} = \boldsymbol{H} + u\boldsymbol{I}$, where $0 < u \ll 1$. Assuming $f(\cdot)$ is a quadratic function to the real numbers with Hessian of the new form $\widehat{\boldsymbol{H}} = \lambda \boldsymbol{I}$ where $\lambda > 0$. Given a point $\boldsymbol{M}^{t+1} = \boldsymbol{M}^t - \eta(\nabla f(\boldsymbol{M}^t))$, and a decent step $0 < \eta < 1/\lambda$ then $f(\boldsymbol{M}^{t+1}) < f(\boldsymbol{M}^t)$. The process of mathematical proof can also be seen in [19].

The Hessian matrix can alleviate the overfitting happening in the process of the factors calculation. However, the large-scale Hessian matrix demands higher computer cost but lower accuracy [21]. Thus, a more efficient way of computation is required to improve the iterative performance for NTD in the next following.

3.2. Operator optimization

From Eq.(8), the simpled function can be deduced as:

$$\begin{aligned} \boldsymbol{G} &= \boldsymbol{K}^{\mathrm{T}}(\hat{\boldsymbol{Y}} - \boldsymbol{Y}) = (\boldsymbol{G}_{(n)\,-n}\{\boldsymbol{A}\}^{\mathrm{T}})\boldsymbol{P}_n^{\mathrm{T}}(\hat{\boldsymbol{Y}} - \boldsymbol{Y}) \\ &= (\hat{\boldsymbol{Y}} - \boldsymbol{Y})(_{-n}\{\boldsymbol{A}\}^{\mathrm{T}}\boldsymbol{G}_{(n)}) \\ &= (vec(\mathcal{G})\{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\}^{\otimes_{-n}} - vec(\mathcal{Y})\{\boldsymbol{A}\}^{\otimes_{-n}\mathrm{T}}, \boldsymbol{G}_{(n)}) \\ &= vec(< \mathcal{G} \times_{-n} \{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\} - \mathcal{Y} \times_{-n} \{\boldsymbol{A}\}^{\mathrm{T}}, \mathcal{G} >_{-n}) \end{aligned}. \tag{9}$$

Where $<\bullet>_{-n}$ denotes inner product between two tensors along all the matrices except mode-*n*. the product between core tensor and mode matrices should be demonstrated:

$$\begin{aligned} \mathcal{G} \times_{-n} \{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\} &= \mathcal{G} \times_1 \boldsymbol{A}^{(1)\mathrm{T}}\boldsymbol{A}^{(1)} \times \cdots \times_{n-1} \boldsymbol{A}^{(n-1)\mathrm{T}}\boldsymbol{A}^{(n-1)} \times_{n+1} \boldsymbol{A}^{(n+1)\mathrm{T}}\boldsymbol{A}^{(n+1)} \times \cdots \times_N \boldsymbol{A}^{(N)\mathrm{T}}\boldsymbol{A}^{(N)} \\ \mathcal{Y} \times_{-n} \{\boldsymbol{A}\}^{\mathrm{T}} &= \mathcal{Y} \times_1 \boldsymbol{A}^{(1)\mathrm{T}} \times \cdots \times_{n-1} \boldsymbol{A}^{(n-1)\mathrm{T}} \times_{n+1} \boldsymbol{A}^{(n+1)\mathrm{T}} \times \cdots \times_N \boldsymbol{A}^{(N)\mathrm{T}} \end{aligned}. \tag{10}$$

$$s.t \quad \boldsymbol{A}^{(n)} \in \mathrm{R}^{I_n \times J_n}, \mathcal{G} \in \mathrm{R}^{J_1 \times \cdots \times J_n}, J_n \le I_n, n \le N.$$

The Eq.(10) only needs the length of computation space $I_n \times \prod_{k \neq n}^{N} J_k$ rather than $\prod_{k \neq n}^{N} I_k$ in Eq.(9), which consumes much more computer storage when $I_n \gg J_n$. Meanwhile, the Eq.(10) is not necessary to reconstruct an approximate tensor $\widehat{\mathcal{Y}}$ any more.

3.3. Methodology implement

Conjugating the traditional NTD and the new updating algorithm, the methodology of the NTD algorithm is carried out as following:

Input: $\mathcal{Y} \in \mathrm{R}^{I_1 \times I_2 \times I_3}$, core tensor $\mathcal{G} \in \mathrm{R}^{J_1 \times J_2 \times J_3}$, matrices $A^{(1)} \in \mathrm{R}^{I_1 \times J_1}$, $A^{(2)} \in \mathrm{R}^{I_2 \times J_2}, A^{(3)} \in \mathrm{R}^{I_3 \times J_3}, 0 < \alpha < 1$;

Output: $A^{(1)} \in \mathrm{R}^{I_1 \times J_1}, A^{(2)} \in \mathrm{R}^{I_2 \times J_2}, A^{(3)} \in \mathrm{R}^{I_3 \times J_3}$ and $\mathcal{G} \in \mathrm{R}^{J_1 \times J_2 \times J_3}, \mathcal{Y}$:.

Begin

Initializing A and $\mathcal{G}$

for $n = 1 : N$

$A^{(n)} = Y_{(n)}\left(vec\left(ttm(\mathcal{G} \otimes A, -n)\right)\right)_+$; /* Add nonnegative constraint on A

$\mathcal{G} = ttm(\mathcal{Y}, A)$;

/* A is Moore – Penrose of A

if (update > 0) Update; end

end

$\sigma = \left\| \mathcal{Y} - \widehat{\mathcal{Y}} \right\|_F^2$;

Update

for $n = 1 : N + 1$

$H = (\{A\}_{-n} \mathcal{G}_{(n)}{}^{\mathrm{T}} \otimes I_{I_n})^{\mathrm{T}} (\{A\}_{-n} \mathcal{G}_{(n)}{}^{\mathrm{T}} \otimes I_{I_n})$;

$M_H = (A^{(1)\mathrm{T}}, \cdots, A^{(N)\mathrm{T}}, vec(\mathcal{G}))$;

$K_n = \{A\}_{-n} \mathcal{G}_{(n)}{}^{\mathrm{T}} \otimes I_{I_n}$;

end

$K = [K_1, K_2, \cdots, K_{N+1}]$;

$G = K^{\mathrm{T}}(\hat{Y} - Y)$;

$M_{H+} = (M - H^{-1}G)_+$; /* Add nonnegative constraint on M_H

if ($error \leq 10^{-3}$ or $delta = 0$ or $iteration \geq 3000$) stop; end

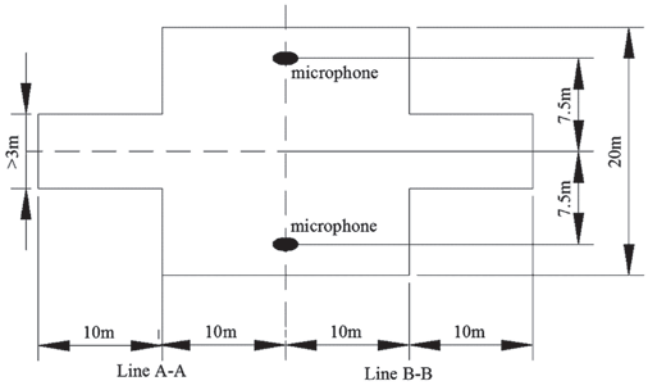
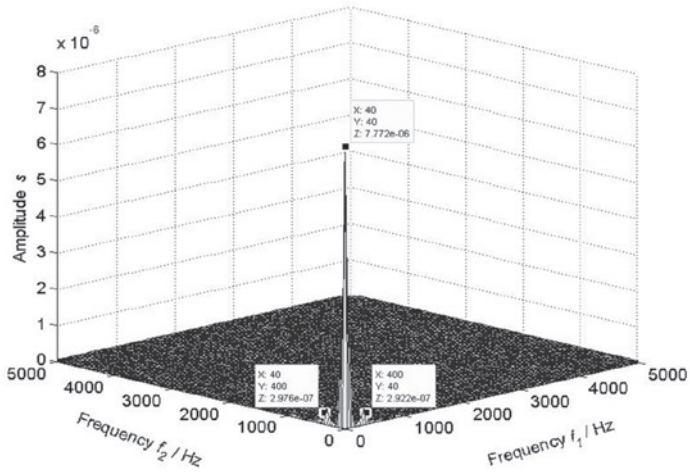


Fig. 3. Layout of the testing ground

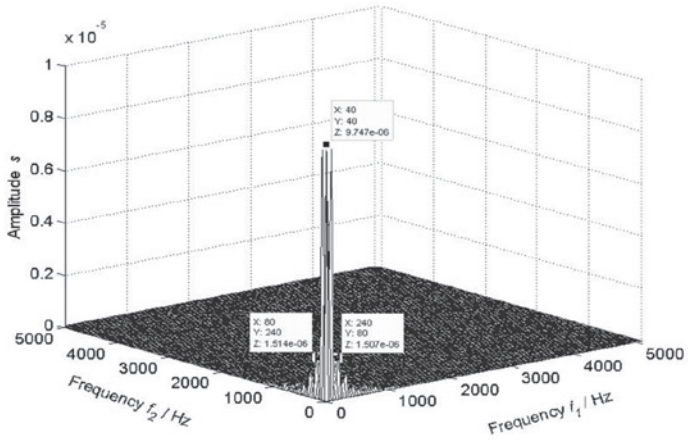
Fig. 4. Test ground with LMS test.lab

4. Bispectrum analysis of automobile vehicle exterior noise

The layout of the testing ground can be simply sketched shown as in Fig. 3. Line A-A and Line B-B are the two starting points of acceleration in the case of wide open throttle (WOT) via opposite directions, respectively. The data acquisition equipment of LMS test.lab must be fixed both on the points of the two microphones. Set the sample frequency as 10240 Hz and the sample time as 10 seconds. The real test ground and the system of LMS test.lab are shown as in Fig.4.

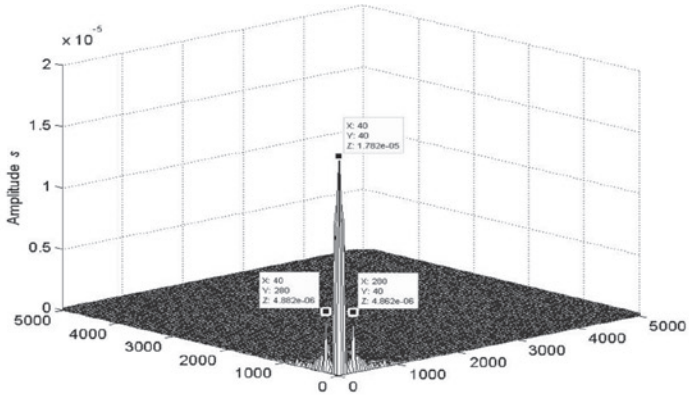


(a) Get rid of thermal shield with damping disk on exhaust pipe



(b) Add new sound package and get rid of thermal shield with damping disk on exhaust pipe

Fig. 5. Bispectrum expression of external vehicle noise



(c) Original state

Fig. 5. Bispectrum expression of vehicle exterior noise

Taking the computation time into account, the length of each test array data is chosen as 65536 or 6.4s sample time. Thus the bispecture is consist of the matrix with the size 256×256. Herein, we choose three different states of vehicle exterior noise as analytical object: (a) get rid of thermal shield with damping disk on exhaust pipe; (b) add new sound package and get rid of thermal shield with damping disk on exhaust pipe; (c) original state with no thermal shield, respectively. The bispectrums of three states are illustrated as Fig. 5, where the figures are plotted with the frequency f_1 on x-axis, the frequency f_2 on y-axis and the vibration displacement S on z-axis(the same below).

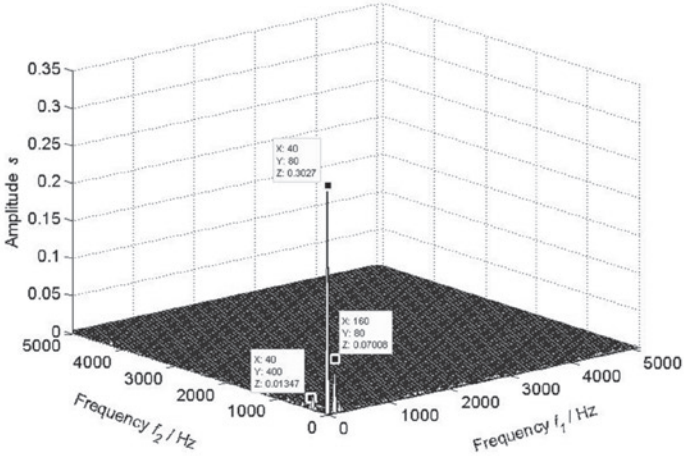
The different bispectrums of the vehicle exterior noise are shown as in Fig. 5. It is easy to find that the bispectrums may be confused from the peak values due to interference signals, or some useful signals are masked by other noise, which leads to seriously difficult to judge the state types of the vehicle exterior noise. Fine out the frequency of the noise property belonging to a harmonic pipe is the primary way to NVH engineer. Thus, the methodology of secondary feature extraction is necessary to develop for state recognition once more.

4.1. Secondary feature extraction

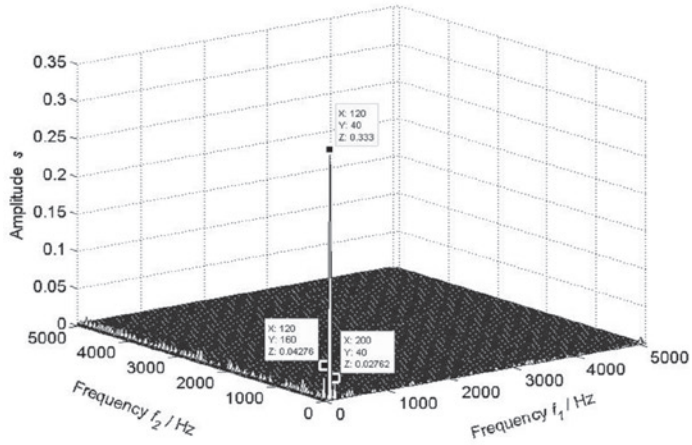
The experiments are implemented on MATLAB R2012b and partly use the tensor toolbox [22]. White Gaussian noise with the level 0:0.1:6.4 is added on the dataset to reconstruct a new tensor with the size 256×256×64. Thus, for a third order real tensor, the expression of the secondary feature involved in reference [23] and can be written as:

$$\mathcal{Y} := \mathcal{G} \times_1 A^{(1)} \times_2 A^{(2)} \times ((A_2^{(2)T} A_2^{(2)}) * (A_1^{(1)T} A_1^{(1)}), \qquad (11)$$

Where $\mathcal{Y}$: is a set basis images of secondary features. In the initializing phase, the size of the core tensor is set as (128, 128, 32), which refers to the conclusion about the arguments size of core tensor approximating to one half size of the original tensor referred in [19]. Two basis images of secondary features for each state are extracted from the reconstruction tensor of vehicle exterior noise shown as in Fig. 6.

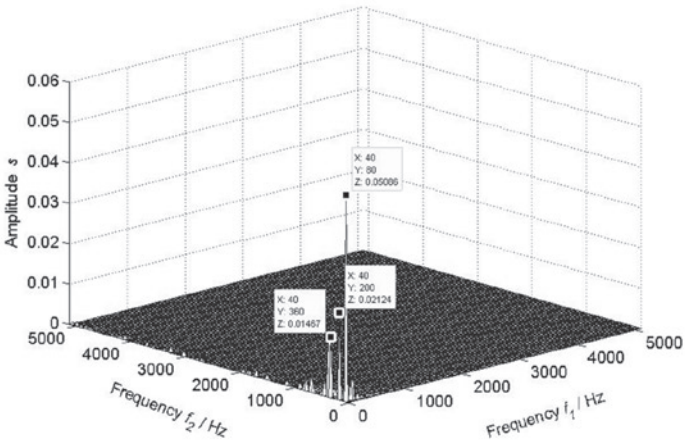


(1)

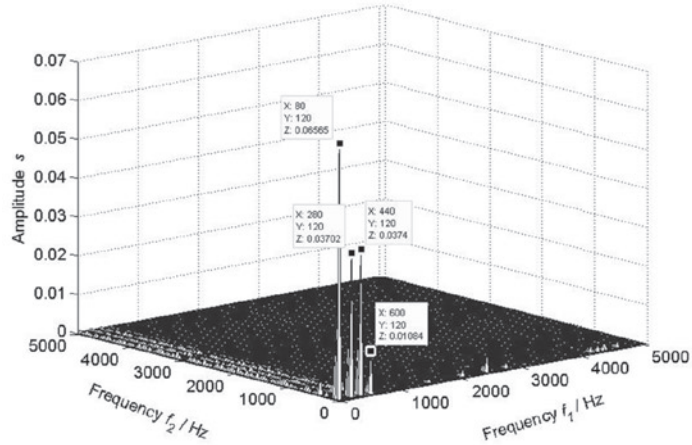


(2)

(a) Get rid of thermal shield with damping disk on exhaust pipe



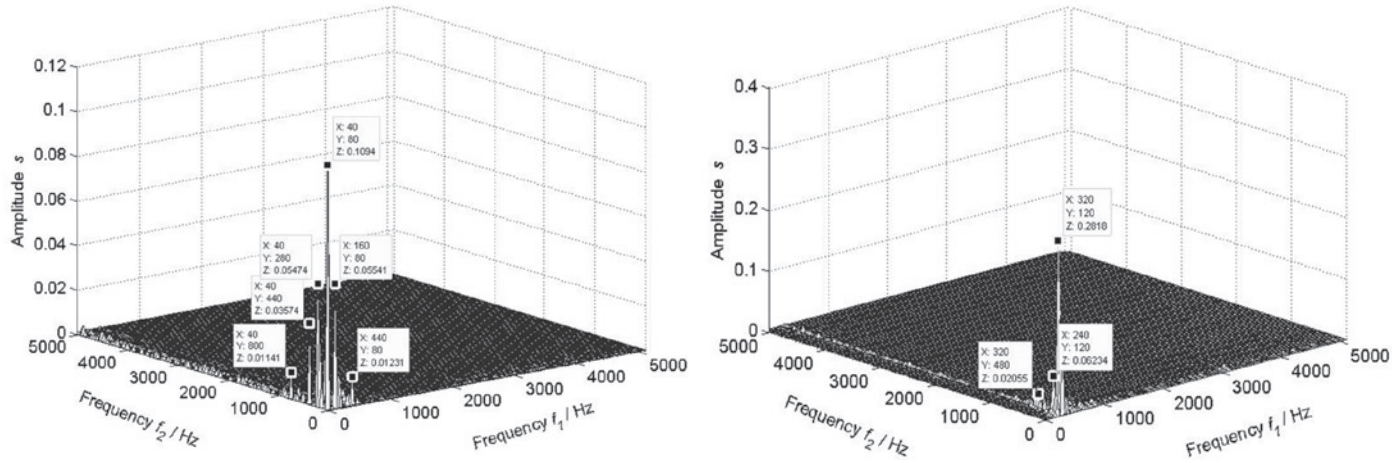
(3)



(4)

(b) Add new sound package and get rid of thermal shield with damping disk on exhaust pipe

Fig. 6. Two basis images of secondary features extracted by NTD from vehicle exterior noise (part images)



(5) (6)

(c) Original state

Fig. 6. Two basis images of secondary features extracted by NTD from vehicle exterior noise (part images)

From Fig. 6, the primary paired frequency (40, 80) (Hertz, the same below) denotes there exists some significant secondary features, which are masked probability before, and against only 40 Hz alone in Fig.5. Furthermore, along with some other paired frequency multiplication such as (120, 40), (80, 120), (220, 120) arises in the three states that possibly lead to an important argument of harmonic generation and keep the noise rising, which are not revealed absolutely in Fig. 5 as well. Consequently, the analytical way of bispectrum of secondary features extracted by NTD is of significance to expose out the masked signals availably for signal interpretation.

4.2. Efficiency comparison

In this section, the NTD method will be used to decompose the same tensor of vehicle exterior noise and compare with other typical algorithms, such as NTF, HNTF with hierarchical ALS and NMF mentioned above, take the pre-existing sample dataset as the object of analytic target. Theoretically, the complexity of each algorithm for a same tensor with the size $n \times n \times n$ is demonstrated as in Table.1 and to be verified in the following.

Table 1. Complexity for each method

Method	NTD	NTF	HNTF	NMF
Complexity	$j_1 j_2 j_3 \log n$	$j^3 \log n$	$j^3 \log n$	$n^3 \log n$

Note: $j_1 \le n, j_2 \le n, j_3 \le n;\ j_1, j_2, j_3 \le j \le n.$

Accordding to the Table 1, when under the same condition, the complexity of NTD method is much less comparing with other methods shown as in the columns. If the deviations of successive relative error (DSRE, dB) are marked as γ , that is:

$$\gamma = -20\log_{10}\frac{\left\|\widehat{\mathcal{Y}} - \mathcal{Y}\right\|_F}{\mathcal{Y}_F} \tag{12}$$

Herein, we adopt the DSRE and the computation time as two measure gauges to verify the effect of NTD. Choose (128, 128, 32) as the rank of core tensor for the different scale tensors. Results of all methods about DSRE (γ /dB) and time (t/s) are recorded in Table 2.

Table 2. Computation results of different methods from three dataset

Methods	$256\times256\times40$		$256\times256\times48$		$256\times256\times64$	
	DSRE	Time	DSRE	Time	DSRE	Time
NTF	14.32	1633.00	23.56	1906.64	25.10	2500.38
HNTF	21.04	1008.60	24.99	1362.12	26.54	1878.31
NMF	15.14	3837.60	18.76	4140.78	20.17	4785.40
NTD	28.36	987.26	28.78	1197.17	31.03	1604.00

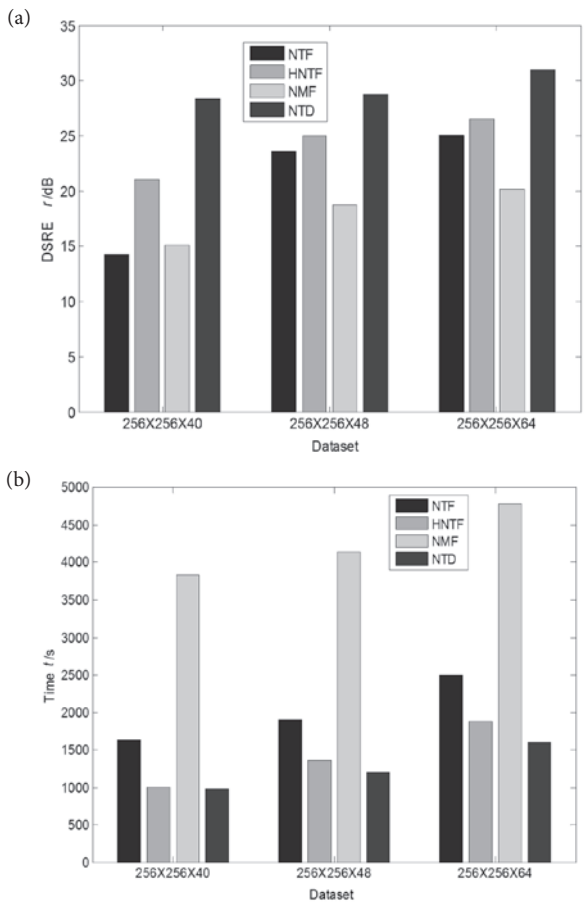


Fig. 7. Bar result comparison of different methods. (a) Computation accuracy, (b) Computation time

In order to observe the computation results, the bar diagrams are generated from Table.2 shown as in Fig.7.

Combing the bar diagram Fig.7 with the basic data in Table.2, it is easy to find the NTD can reach the highest DSRE with 31.03dB but with the least time against other methods under the same condition, shown as in Fig.7 (a). Particularly, the NMF algorithm needs to take the most time with 4 785.40 s to complete the calculation procedure as in Fig.7 (b) but the DSRE with only 20.17 dB, which means lower performance than other methods. The complexity for each algorithm mosltly meets the theoretical expression in Table.1. Thus, the NTD has overwhelming performance in tensor decomposition under the same condition.

4.3. Results and discussion

Features extracted by NTD in section 4.1 allow us to know that some secondary features are distinctly exposed out from the masked signals, which maybe generate destructive interference with the real features such as the paired frequency at the (120, 40), (80, 120) and (220,120) or near the around. Thus, secondary features help us analyzing the natural frequency of the components in the exhaust system. It is an effective way to lower the noise level of exhaust system and improve the NVH (noise, vibration and harshness) in automobile engineering as well. How to offer an approach to revise a component in exhaust system for avoiding the coupling of noise transmission is not involved in NTD, but NTD can provide with more information for the state diagnosis or maintenance. Above all, NTD is a successful tool to NVH engineer and will be a new method in the NVH analysis of an automobile car.

5. Conclusions

(1) NTD is proposed to extract the secondary feature for bispectrum analysis of vehicle exterior noise, and the basis images are able to interpret the new features masked before. Method of iteration calculation conjugating with updating algorithm based on NGGD improves the iterative performance of NTD.

(2) The more efficiency and higher accuracy of NTD are verified by different dimensions of the same tensor. Meanwhile, NTD is of success to overcome the problem of overfitting in theory. Related conclusions are also discussed in [24].

(3) Experiments show NTD less complexity comparing other typically methods, and advantages both at the DSRE and computation time.

Acknowledgement

This project is supported by the fund project of SAIC-GM-Wuling Automobile CO., LTD (Grant No.N310-N12). Thanks to all the engineers of project participants who give strong supports as well.

References

1. Kopriva I. 3D tensor factorization approach to single-frame model-free blind-image deconvolution, Optics letters 2009; 34(8): 2835-2837, http://dx.doi.org/10.1364/OL. 34.002835
2. Paatero P, Tapper U. Positive matrix factorization - a nonnegative factor model with optimal utilization of error-estimates of data values, Environmetrics 1994; 5(2): 111-126, http://dx.doi.org/10.1002/env.3170050203.
3. Cichocki A, Phan A H , Caiafa C. Flexible HALS algorithms for sparse non-negative matrix/tensor factorization, IEEE Workshop on Machine Learning for Signal Processing 2008: 73 – 78, http://dx.doi.org/10.1109/MLSP.2008.4685458.
4. Kopriva I, Cichocki A. Nonlinear Band Expansion and 3D Nonnegative Tensor Factorization for Blind Decomposition of Magnetic Resonance Image of the Brain. Latent Variable Analysis and Signal Separation 2010, Vigneron V, Zarzoso V, Moreau E, Gribonval R, and Vincent E, Editors, ed: Springer-Verlag Berlin 2010; 6365: 490-497, http://dx.doi.org/ 10.1007/978-3-642-15995-4_61.
5. Kırbız S, Günsel B. A multiresolution non-negative tensor factorization approach for single channel sound source separation. Signal Processing 2014; 105: 56-69, http://dx.doi.org/10.1016/j.sigpro.2014.05.019.
6. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 2004; 134: 9-21, http://dx.doi.org/10.1016/j.jneumeth. 2003.10.009.
7. Conesa A, Prats-Montalbán J M, Tarazona S, Nueda M J, Ferreret A. A multiway approach to data integration in systems biology based on Tucker3 and N-PLS. Chemometrics and Intelligent Laboratory Systems 2010; 104: 101-111, http://dx.doi.org/ 10.1016/j.chemolab.2010.06.004.
8. Opriva I, and Jerić I. Blind separation of analytes in nuclear magnetic resonance spectroscopy: Improved model for nonnegative matrix factorization. Chemometrics and Intelligent Laboratory Systems 2014; 137: 47-56, http://dx.doi.org/10.1016/j. chemolab.2014.06.004.
9. Ding H, Ji H. Application of chemometric methods to analyze the distribution and chemical fraction patterns of metals in sediment from a metropolitan river. Environmental Earth Sciences 2010; 61: 641-657, http://dx.doi.org/10.1007/s12665-009-0379-8.
10. Acar E, Yener B. Unsupervised multiway data analysis: A literature survey. IEEE Transactions on Knowledge and Data Engineering 2009; 21: 6-20, http://dx.doi.org/10.1109/TKDE.2008.112.
11. Sidiropoulos N D, Giannakis G B, Bro R. Blind PARAFAC receivers for DS-CDMA systems. IEEE Transactions on Signal Process 2000; 48(3): 810-823, http://dx.doi.org/10.1109/78.824675.
12. Cichocki A, Zdunek R, Phan A H, Amari S I. Alternating Least Squares and Related Algorithms for NMF and SCA Problems, in Nonnegative Matrix and Tensor Factorizations, ed: John Wiley & Sons, Ltd, 2009: 203-266, http://dx.doi.org/10.1002/ 9780470747278.ch4.
13. Kim H, Park H. Sparse non-negative matrix factorizations via alternating non-negativity constrained least squares for microarray data analysis. Bioinformatics 2007; 23(12): 1494-1402, http://dx.doi.org/10.1093/bioinformatics/btm134.
14. Cichocki A, Zdunek R, Phan A H, Amari S I. Introduction-Problem Statements and Models, in Nonnegative Matrix and Tensor Factorizations. ed: John Wiley & Sons, 2009:1-75, http://dx.doi.org/http://dx.doi.org/10.1002/ 9780470747278.ch1.
15. Kim Y D, Choi S J. Nonnegative tucker decomposition. IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2007; 1-8, http://dx.doi.org/ 10.1109/CVPR.2007.383405.
16. Bro R. Multi-way analysis in the food industry: Models, algorithms and applications. PhD thesis, Amsterdam: university of Amsterdam, 1998, http://curis.ku.dk/ ws/files/13035961/rasmus_bro.pdf.
17. Chen X R. Introduction to mathematical statistics in Linear model, Beijing : Chinese science press, 2007. (in Chinese)
18. Cichocki A, Zdunek R, Phan A H, Amari S I. Multi-Way Array (Tensor) Factorizations and Decompositions, in: Nonnegative Matrix and Tensor Factorizations. ed: John Wiley & Sons, 2009; 337-432, http://dx.doi.org/ 10.1002/9780470747278.ch7.

19. Wang H J, Xu F Y, Zhao J A, Jia M P,Hu J Z, Huang P. Bispectrum Feature Extraction of Gearbox Faults Based on Nonnegative Tucker3 Decomposition with 3D Calculations. Chinese journal of mechanical engineering 2013; 26(6): 1182-1193. http://dx.doi.org/10.3901/CJME.2013.06.1182.
20. Phan A H, Tichavsky P, Cichocki A, IEEE member. Damped Gauss-Newton algorithm for nonnegative Tucker decomposition. IEEE Conference on Statistical Signal Processing Workshop(SSP) 2011; 665-668, http://dx.doi.org/ 10.1109/SSP.2011. 5967789.
21. Gillis N, Glineur F. A multilevel approach for nonnegative matrix factorization. Journal of Computational and Applied Mathematics 2012; 236(7): 1708-1723, http://dx.doi.org/ 10.1016/j.cam.2011.10.002.
22. Bader B W, Kolda T G. Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. ACM Transactions on Mathematical Software(TOMS) 2006; 32(4): 635-653, http://dx.doi.org/ 10.1145/1186785.1186794.
23. Wang H J, Xu F Y, Wang F. Tensor factorization and clustering for the feature extractionbased on Tucker3 with updating core. Advanced Materials Research, Advanced Design Technology, ADME 2011; 308-310, http://dx.doi.org/ 10.4028/www.scientific.net/AMR.308-310.2517.
24. Phan A H, Tichavský P, Cichocki A. Low complexity damped Gauss-Newton algorithms for Candecomp/Parafac. Journal on matrix analysis and applications 2013, 34(1): 126-147, http://dx.doi.org/ 10.1137/100808034.

Haijun WANG
Guoyong DENG
Qinglin LI
Qiang KANG
NVH Institute of vehicle performance integration section
SAIC-GM-Wuling Automobile CO., LTD
18 Hexi Road, Liuzhou 545007, Guangxi, P.R. of China

E-mails: whjun69@sina.com, Guoyong.Deng@sgmw.com.cn, Qinglin.Li@sgmw.com.cn, Kangqiangok@hotmail.com

Article citation info:
ZUBER N, BAJRIĆ R. Application of artificial neural networks and principal component analysis on vibration signals for automated fault classification of roller element bearings. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 299–306, http://dx.doi.org/10.17531/ein.2016.2.19.

Ninoslav ZUBER
Rusmir BAJRIĆ

APPLICATION OF ARTIFICIAL NEURAL NETWORKS AND PRINCIPAL COMPONENT ANALYSIS ON VIBRATION SIGNALS FOR AUTOMATED FAULT CLASSIFICATION OF ROLLER ELEMENT BEARINGS

ZASTOSOWANIE SZTUCZNYCH SIECI NEURONOWYCH ORAZ ANALIZY GŁÓWNYCH SKŁADOWYCH SYGNAŁU DRGAŃ DO AUTOMATYCZNEJ KLASYFIKACJI USZKODZEŃ ŁOŻYSK TOCZNYCH

The article addresses the implementation of feature based artificial neural networks and vibration analysis for automated roller element bearings faults identification purpose. Vibration features used as inputs for supervised artificial neural networks were chosen based on principal component analysis as one of the possible methods of data dimension reduction. Experimental work has been conducted on a specially designed test rig and on a drive of the Ganz port crane in port of Novi Sad, Serbia. Different scalar vibration features derived from time and frequency domain were used as inputs to fault classifiers. Several types of roller elements bearings faults, at different levels of loads were tested: discrete faults on inner and outer race and looseness. It is demonstrated that proposed set of input features enables reliable roller element bearing fault identification and better performance of applied artificial neural networks.

Keywords: *roller elements bearing, vibration, artificial neural network, principal components analysis.*

Artykuł omawia zastosowanie sztucznych sieci neuronowych opartych na cechach oraz analizy drgań do celów automatycznej identyfikacji uszkodzeń łożysk tocznych. Cechy drgań mające posłużyć jako dane wejściowe do nadzorowanych sztucznych sieci neuronowych wybrano na podstawie analizy głównych składowych, która stanowi jedną z metod zmniejszania rozmiaru zbioru danych statystycznych. Badania prowadzono na specjalnie do tego celu zaprojektowanym stanowisku badawczym oraz na układzie napędu żurawia portowego firmy Ganz w porcie Novi Sad w Serbii. Jako wejścia klasyfikatorów uszkodzeń wykorzystano różne skalarne cechy drgań określone w dziedzinie czasu i częstotliwości. Badano kilka typów uszkodzeń łożysk tocznych przy różnych poziomach obciążenia: uszkodzenia dyskretne w obrębie pierścienia wewnętrznego i zewnętrznego łożyska oraz nadmierny luz. Wykazano, że proponowany zbiór cech wejściowych umożliwia niezawodną identyfikację uszkodzeń łożysk tocznych oraz zapewnia lepszą wydajność zastosowanych sztucznych sieci neuronowych.

Słowa kluczowe: *łożysko toczne, drgania, sztuczna sieć neuronowa, analiza głównych składowych.*

1. Introduction

The ultimate goal of every maintenance strategy in a modern plant is to avoid high maintenance costs and productions risks due to the rotating machine's fault. High costs are initiated through the production stops and losses while the production risks are related to the secondary failures of the neighboring machines. Monitoring the machine's health through the implementation of condition based maintenance strategy is based on acquisition and trending the physical parameter that is found to be sensitive to machine degradation. Several methods of non-destructive testing are available nowadays, such as vibration measurement and analysis, infrared thermography, noise measurement, motor current signature analysis, wear particle analysis, ultrasound measurements etc. Mechanical vibration acquired at the bearing's housing (absolute vibration) or directly on a rotating part (relative vibration) is one of the best parameter for early detection of a developing fault inside a machine. If appropriate vibration transducer is engaged and mounted properly and if proper signal processing methods has been used for the suspected fault, then we can say that the vibration signal contains unambiguous information on the existing state of the machine. Methods of vibration signature analysis enable the extraction of type and severity of a fault inside the rotating machine. However the existing guidelines are not universal due to the facts that there may be multiple faults inside the machine and that the content of the acquired vibration signals are dependent on the severity of the fault and on the variation of the rotating speed and load. As a result, derivation of incorrect conclusions and wrong estimation of machine criticality in the plant, is a very common situation. We can avoid this by engagement of highly skilled certified vibration analysts or by the implementation of artificial intelligence (AI) techniques for reliable extraction of an existing fault. In the absence of certified vibration analysts inside the maintenance team the implementation of AI methods through previously developed and validated fault identification algorithm has a promising potential.

For the purpose of automatic machine health determination through automatic fault identification there are several applicable methods of AI such as supervised and unsupervised artificial neural networks (ANN), fuzzy logic, expert systems and hybrid intelligence systems. The most applied are ANN [14, 1] due to their ability to learn i.e. to adopt novelties. This adaptability of ANN results in a possibility for detection of an existence of a new condition (fault) based on the existing data [13, 6]. In addition, ANN are found to be efficient in modeling of highly complex nonlinear phenomena that are present in several types of rotating machinery faults. Several types of ANN are successfully

implemented in automatic fault identification [3, 7, 8, 12, 16, 17, 18, 20]: back propagation feed forward network (BPFF), multiple layer perceptron network (MLP), back propagation multiple layer perceptron (BPMLP), radial basis function network (RBF), self-organized feature map (SOFM). An excellent review of different types of ANN and training algorithms implementation for different types of rotating machinery faults can be found in [10]. The increasing trend of implementation of MLP with back propagation training algorithm, with the number of neurons in hidden layers taken as a variable, is evident.

The success of ANN in identification and classification of machine fault is highly dependent on the definition of the cloud of input variables i.e. on the definition of the most representative vibration features that are sensitive on the fault occurrence and progression over the time. One vibration feature could be appropriate for one type of vibration fault while on the other hand it can be unresponsive for other type of fault.

As a preprocessing tool for selecting the most important vibration features, we used principal component analysis (PCA). PCA is one of the most frequently used multivariate data analysis technique. One of the main goals of PCA implementation is the reduction of dimensionality of the input set of vibration features. Basically it is defined as an orthogonal linear transformation that transforms the cloud of input variables to a new coordinate system in a way that the greatest variance of the input features are aligned on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. As a result, possibly correlated input features are converted into a set of values of linearly uncorrelated variables that we call principal components.

Roller element bearings are present in all types of rotating machines and often they are claimed to be the most critical parts of the machine and the main culprits of the machine failures. If we add to this the fact that roughly just 10% [11] of the bearings run for their complete design life then we can see that development of signal processing techniques and data preprocessing methods combined with the algorithms of ANN is prominent in maximizing the reliability of rotating machinery.

A successful implementation of ANN and PCA for the identification of rotating machine faults can be found in [2, 15, 4, 9]. The authors used different scalar features obtained from vibration data as inputs for neuron classifiers.

In this paper, we used vibration scalar features obtained from frequency and time domains. The initial definition of vibration features is done based on an assumption that these features are sensitive to bearing failures tested in this paper.

2. Vibration analysis techniques for roller element bearings failures identification

Roller element bearing can prematurely fail due to different reasons (Figure 1) and its failure can be initiated through different mechanisms such as: fatique crack, wear, plastic deformation of bearing components, corrosion, brinelling phenomena etc. Often, these mechanisms are overlapping inside the bearing. It is also possible that one mechanism activates the initial damage and that over the time another mechanism runs the bearing to the final failure [19, 16].

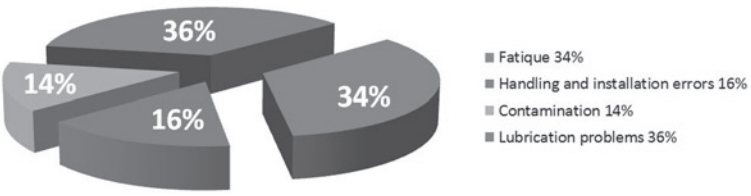


Fig. 1. Most common mechanisms for bearing failure [11]

If we follow the best practices on proper lubrication, handling and installation of bearings then the most expected mechanism of bearing degradation is a material fatigue crack. In that case, roller element bearing which is subjected under the projected dynamic load will fail due to the occurrence of the fatigue crack. Due to the bearing geometry, the most expected place of initial crack occurrence is under the contact surface of internal race and the roller element. If such a bearing is left in operation under the load, the crack is expanding and occurs at the surface on the bearing race (Figure 2). The next stage of bearing degradation is the enlargement of the crack. At that stage other cracks might occur. Flaws from the race damage other components of the bearing and we have a spalling inside a bearing. As a final result we have a bearing with excessive looseness.

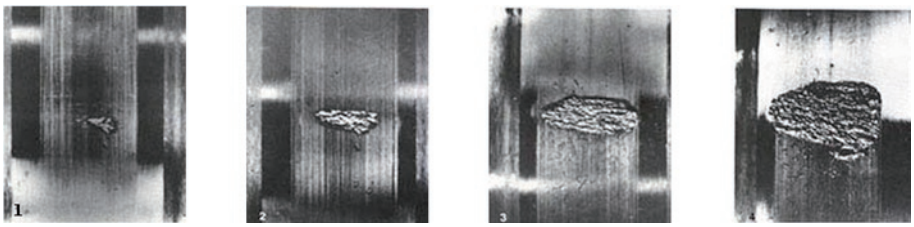


Fig. 2. Development of crack on the internal race of the bearing

Ball bearing has four basic components: inner and outer race, roller elements and a cage. If we have a discrete crack on one of these components, then we have a chance to identify its characteristic forcing frequency in time and frequency domain. Based on the geometry of the bearing we can calculate these frequencies: BPFI (ball pass frequency of inner race), BPFO (ball pass frequency of outer race), BS (ball pass frequency) and CF (cage frequency).

The content of vibration signal from the bearing with a developing damage is highly dependent on the type and stage of degradation [19]. At the initial stage, we can see only minor impacts masked in noise. At later stages the crack develops and impacts are high enough to cause the bearing's natural frequency (f_{res}) excitement. Such a case, with the single crack on the inner race and with BPFI = 4.1X, where X stands for the first harmonic order (shaft speed), is numerically simulated and shown on Figure 3.

Time waveforms from Figure 3 reveal some interesting facts regarding vibration signals from faulty bearing. Signal is a sum of high amplitude and low frequency component from 1X and low amplitude and high frequency components from impacts that are generated every time when the ball hits the crack. Due to the presence of impulse excitation, the system response is in the form of exponentially decayed harmonic component at the bearing's natural frequency. The periodicity of impacts corresponds to the characteristic fault frequency (BPFI in the present case) so in the frequency spectrum we can see sidebands at this fault frequency centered around bearing's natural frequency. In cases when the fault rotates inside the bearing (fault on the inner ring or on the roller) we have an amplitude modulation of the fault frequency component. The carrier frequency is the fault frequency while the modulation frequency is the speed of the fault inside the bearing and is shown on Figure 3. The peak amplitudes of the impulses are not equal during a revolution of the shaft due to the fact that the fault (crack on the inner race in this case) comes in and goes out from the bearing load zone. BPFI component will be amplitude modulated with the 1X component. In case of a crack on a roller the BS component will be modulated with the CF since the cage holds the rollers and determines the speed of the rollers. In a case of a crack on the outer race, we do not expect amplitude modulation around the BPFO component.

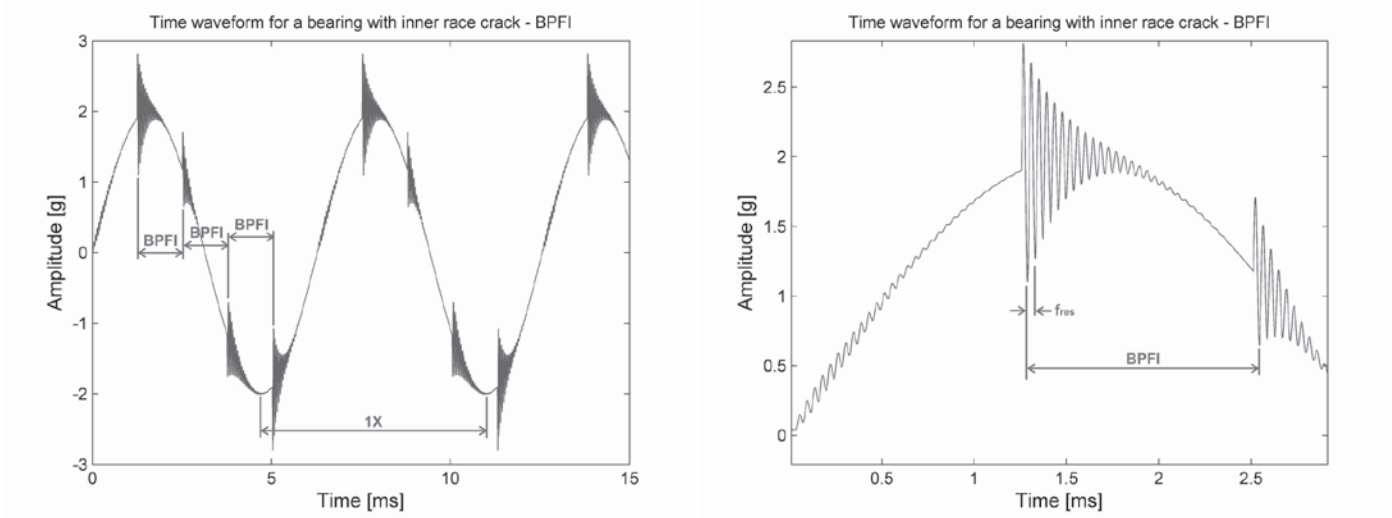


Fig. 3. Bearing with inner race crack – time waveform generated by numerical simulation

As a fault develops, we can expect higher harmonics of bearing fault frequencies, harmonics of the fundamental frequency and a broadband noise level increase due to the excessive looseness. Frequency spectra from a bearing with an outer race damage located in the load zone and with an excessive looseness is shown on Figure 4. Two harmonic families (1X and BPFO) as well as raised broadband noise is easy to see. Frequency spectra recorded on the test rig is shown on Figure 5.

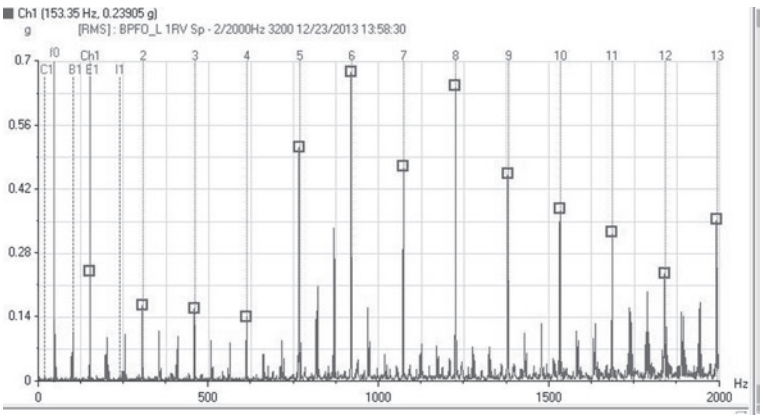


Fig. 4. Frequency spectra from a bearing with an excessive looseness and outer race fault. Harmonic cursor on BPFO family

Vibration signals from faulty bearings can be analysed using well-known methods of signal processing in time, frequency and time – frequency domains. Analysis in time domain can be performed on raw and on filtered signals. Time domain analysis is usefull in later stages of bearing degradation since the impulses from discrete cracks and from excessive looseness are then visible. In addition, it is worth to mention that it is necessary to measure acceleration of vibration due to its high sensitivity to high frequency phenomena. Since faulty bearings generate family of harmonics, which can be treated as periodicity in frequency domain, Cepstrum analysis can be used also. Dominant peaks in cepstrum can indicate the presence of 1X harmonics (looseness in bearings) and amplitude modulations (inner race and roller fault). Analysis in frequency domain is mainly based on analysis of classic Fast Fourier Transform (FFT) and on analysis of acceleration envelope spectra. FFT is an effective tool in analysis of moderate and heavy damages in bearings while the envelope spectra is the most effective universal tool for identification of early faults in bearings. Envelope spectra is calculated on band pass filtered time waveform combined with the methods of signal demodulation. The aim of the band pass filtering is removal of high amplitude low frequency component and enhancement of the high frequency part of the spectra where natural frequency of the bearing amplitude modulated with the bearing fault frequency is located. As a result, we get the envelope spectra with the harmonics of the fault frequency. By measuring the relative height of these bearing fault frequencies from the carpet noise, we could quantify the bearing health state. However, in the later stage of the bearing degradation, we have an increase of the broadband noise, which remains in the resulted envelope spectra. In that case, the component of the bearing fault frequencies are masked in the noise and, despite the bearing's state gets worser, we get a decrease in bearing fault frequencies relative height. For some bearing faults, we get the non-stationarity in vibration signal so methods of analysis, both in frequency and time domain can be used. Some of them are Short Time Fourier Transform (STFT) and wavelet analysis.

For the purpose of vibration trending and implementation of ANN we have to define scalar vibration features that increase with damage development. Due to the impact phenomena, which is present in faulty bearings, acceleration parameters should be used. For later stages of bearing degradation, features based on vibration velocity can be used also.

3. Experimental set up and results

The test rig, designed for the purpose of dataset collection, is shown on the Figure 5. The test rig consists of a 0.37 kW variable frequency drive connected over the flexible coupling to the shaft with two disks for unbalance introduction. Shaft is supported by two single row roller element bearings, type UC201A. The bearing fault frequencies, in term of harmonic orders, are: BPFI = 4.9X, BPFO = 3.1X, BS = 2.1097X and CF = 0.3875X.

Bearing vibrations were measured in radial directions, using industrial type IEPE accelerometers mounted at the roller element bearing housings using mounting studs. Input shaft speed is measured using a non-contacting laser sensor and a reflective mark. Vibration and tacho signals were acquired simultaneously using multichannel vibration analyzers NetdB, OneproD MVX, dbFA and XPR software from 01db-Metravib. All the tests were performed at the 22Hz of input speed.

Before the test impact hammer was used to excite the bearing in order to record its natural frequency. Natural frequency was found at 4.026 kHz.

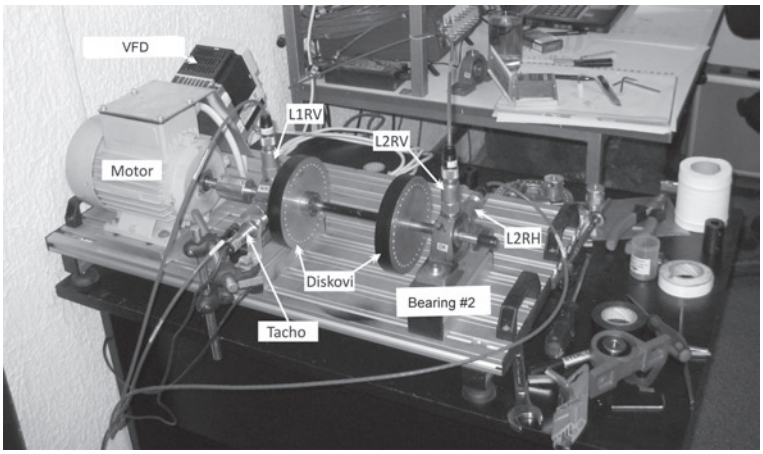


Fig. 5. Test rig used for vibration acquisition on faulty bearings

Table 1. Test rig data: labels for different unbalance levels and bearing fault types

		Bearing fault type						
		OK	I	II	OL	OU	Z1	Z3
Unbalance levels	A	AOK	AI	AII	AOL	AOU	AZ1	AZ3
	B	BOK	BI	BII	BOL	BOU	BZ1	BZ3
	C	COK	CI	CII	COL	COU	CZ1	CZ3
	D	DOK	DI	DII	DOL	DOU	DZ1	DZ3

Table 2. Test rig data: vibration features from the original dataset

Number	*Label*	*Description*	*Unit*
1	RMS	RMS of vibration velocity in range 2 Hz -1 kHz	mm/s
2	S1	Amplitude of the 1X vibration velocity component	mm/s
3	SumS	Sum of first seven harmonics of vibration velocity	mm/s
4	Kurt	Kurtosis parameter	-
5	Acc2-300	RMS of acceleration in range 2 Hz – 300 Hz	g
6	Acc2-2000	RMS of acceleration in range 2 Hz – 20 kHz	g
7	Acc2-20000	RMS of acceleration in range 2 Hz – 20 kHz	g
8	Def	Defect factor[1]	-
9	TSS	Peak-Peak value of acceleration raw time waveform	g
10	BPFO	Amplitude of the first harmonic of BPFO	g
11	SumBPFO	Sum of amplitudes of first four harmonics of BPFO	g
12	BPFI	Amplitude of the first harmonic of BPFI	g
13	SumBPFI	Sum of amplitudes of first four harmonics of BPFI	g
14	BS	Amplitude of the first harmonic of BS	g
15	SumBS	Sum of amplitudes of first four harmonics of BS	g
16	FT	Amplitude of the first harmonic of FT	g
17	SumFT	Sum of amplitudes of first four harmonics of FT	g

Four levels of unbalance were introduced on both disks: 22.5 gmm, 54 gmm, 136.5 gmm and 345 gmm, and we assigned them the following labels A, B, C and D, respectively. With every unbalance level, bearings with the following faults were tested: bearing in health condition (OK), small (I) and moderate (II) crack on the inner race, moderate crack on outer race with a crack located in the loading (OL) and unloaded (OU) zone, moderate (Z1) and excessive (Z3) lossenes. This results in total of 28 states with label definition shown on Table 1.

Cracks on inner and outer race were introduced with small grinder tool. Smaller crack on the inner race is 0.1 mm deep while the moderate crack is 0.2 mm deep. The width of the cracks is 1 mm. The crack on the outer race is 1 mm wide and 0.2 mm deep. Looseness has been introduced by adding a small amount of abrasive material (sand) in the grease and leaving such a bearing in operation (one hour for moderate and three hours for excessive looseness). After that, the contaminated grease has been replaced with a new one.

Despite the main goal of this research was a bearing fault identification we also introduced unbalance as a fault type. The main motivation behind this was the fact that in reality multiple faults are present on the machine and that reliable ANN should identify both unbalance level and the type and severity of the bearing fault. There are many rotating machines that can develop unbalance over time, such as fans or pumps with unbalance growth (impeller in contact with abrasive fluids or dirt accumulated on impeller).

Vibration acquisition included the measurement of: raw time waveforms, narrow band FFT in different frequency ranges with 3200 lines of resolution (2Hz-2kHz, 2Hz-5kHz, 2Hz-20kHz) and envelope spectra. Based on these measurements 17 scalar features were extracted. Their labels and description are shown in Table 2.

The definition of the vibration features has been guided by the type of the faults we are trying to identify. The unbalance presence and levels should be sensitive to the first harmonic of the vibration velocity [16]. Faulty bearings generate raised levels of accelerations in different frequency ranges based on exact type of the bearing fault and on the severity of the fault. In case of discrete cracks on different bearing's components higher vibration levels are generated at characteristic frequency of the bearing fault and on its harmonics. This is taken into account through amplitude extraction from the frequency spectra. First harmonics of the characteristic frequencies (BPFO, BPFI, BS, FT) and sums of their first four harmonics were defined. In case of looseness, high levels of acceleration peak values and harmonics of 1X component are generated. The number of harmonics are dependent on the severity of the looseness. Therefore, peak to peak values and acceleration overall values in different frequency ranges were defined. Labels 1-9 in Table 2, correspond to the time domain while labels 10-17 are extracted from frequency spectra (frequency domain).

For every combination of unbalance level and bearing type (Table 1), 150 recordings were acquired. Vibrations were acquired with a periodicity of 3 minutes between them. This resulted in input matrix with 17 columns (vibration features) and 4200 rows (data).

MLP ANN applied in this paper had a classification task – to detect an exact bearing defect type (Table 1). Several architectures of MLP ANN were tested by the means of choosing the optimal network architecture from the point of the number of neurons in the hidden layer, type of activation functions and type of the learning algorithm. For building, testing and training, Statistica Automatic Neural Networks package has been used. 2940 input vectors (70% of the dataset) were used for training while 1260 input vectors were used for cross verification and testing. The software automatically determined network complexity. 20 networks were tested. As a result the best performance network had 15 neurons in the hidden layer and the accuracy of classification of 85.714%. The confusion matrix for this ANN is shown on table 3.

1 Linear combination of peak and RMS values of acceleration

Table 3. Test rig data: confussion matrix for best performance MLP ANN with 17 input features

	AI	AII	AOK	AOL	AOU	AZ1	AZ3	BI	BII	BOK
Total	150	150	150	150	150	150	150	150	150	150
Correct	150	150	150	150	150	150	150	150	0	150
Incorrect	0	0	0	0	0	0	0	0	150	0
Correct (%)	100	100	100	100	100	100	100	100	0	100
Incorrect (%)	0	0	0	0	0	0	0	0	100	0
	BOL	BOU	BZ1	BZ3	CI	CII	COK	COL	COU	CZ1
Total	150	150	150	150	150	150	150	150	150	150
Correct	150	150	150	0	150	150	150	150	150	150
Incorrect	0	0	0	150	0	0	0	0	0	0
Correct (%)	100	100	100	0	100	100	100	100	100	100
Incorrect (%)	0	0	0	100	0	0	0	0	0	0
	CZ3	DI	DII	DOK	DOL	DOU	DZ1	DZ3	*All cases*	
Total	150	150	150	150	150	150	150	150	4200	
Correct	0	150	150	150	0	150	150	150	3600	
Incorrect	150	0	0	0	150	0	0	0	600	
Correct (%)	0	100	100	100	0	100	100	100	85.714	
Incorrect (%)	100	0	0	0	100	0	0	0	14.286	

Table 4. Test rig data: eignevalues, individual and cumulative variance for PCs

	Eigenvalues	% Total variance	Cumulative %
PC1	10.64789	62.63465	62.63465
PC2	2.27647	13.39103	76.02568
PC3	1.67373	9.84548	85.87116
PC4	1.17158	6.89164	92.76280
PC5	0.80524	4.73673	97.49954
PC6	0.28139	1.65523	99.15477
PC7	0.09024	0.53083	99.68559

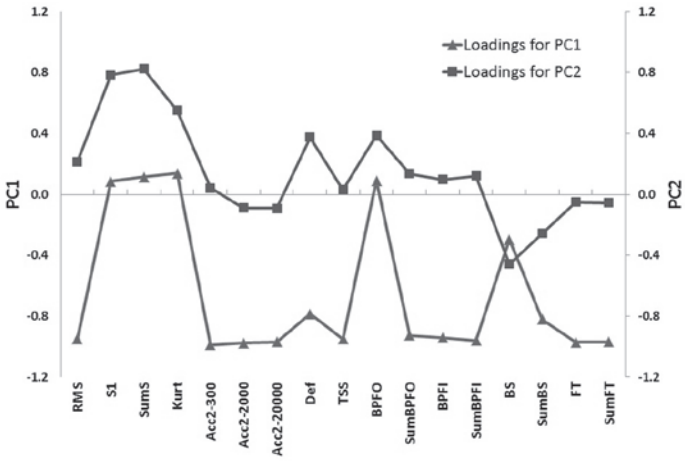


Figure 6. Test rig data: Loadings for PC1 and PC2

Table 3 reveals some interesting facts. The classification rate is either 100% or 0%. Some cases (BII, BZ3, CZ3 and DOL) are completely missed. Therefore, a PCA is conducted on the input dataset to define a dataset with reduced dimension in order to find MLP ANN with better performance. As a result of PCA, 7 principal components (PCs) have been extracted with the eigenvalues shown on table 4.

If we choose the first four PCs with eigenvalues larger than one (criterion proposed by Kaiser in 1960) then we can say that using a 4 dimension PC input space we described the original input dataset with a total of 92.76% of variance. The most influential vibration features for such a space are variables with the highest loadings i.e. highest projection on the coordinate of these dominant PCs. Loadings for each variable for PC1 and PC2 are shown on Figure 6.

The importance of each feature, calculated through modeling power, is based on how well it is represented by the PC model. Features with high modeling power are relevant for the PC model while the variables with low modeling power can be discarded.

Table 5. presents modeling power for each input feature and its importance. As relevant features, first 9 were chosen and this defined the reduced set of input features (Acc2-300, S1, Acc2-2000, SumBPFO, SumBPFI, SumS, RMS, FT and Acc2-20000). Using input matrix with these 9 vibration features several architectures of MLP ANN were tested. As a result the best performance network had 13 neurons in the hidden layer and the accuracy of classification of 99.738%. The confusion matrix for this ANN is shown on table 6. As it can be seen the classification success is much better. There are no completely missing faults as it was the case with the MLP ANN with complete dataset with 17 input features. Only one case for AOU, BZ3, COU, DOL, DZ3 and six cases for CZ3 were incorrectly classified (Table 7).

Table 5. Test rig data: modeling power for each feature and its importance to PC model

Feature	Power	Importance		Feature	Power	Importance
Acc2-300	0.994766	1		SumFT	0.959409	10
S1	0.985738	2		SumBS	0.946349	11
Acc2-2000	0.985107	3		BPFI	0.944689	12
SumBPFO	0.978242	4		TSS	0.911100	13
SumBPFI	0.974435	5		Def	0.899415	14
SumS	0.973051	6		BPFO	0.895808	15
RMS	0.968662	7		BS	0.821585	16
FT	0.963068	8		Kurt	0.606926	17
Acc2-20000	0.961326	9				

4. Case study

PCA and MLP ANN for automatic identification of bearing faults were implemented on the Ganz port crane in port of Novi Sad, where an online system (Figure 7) for the crane surveillance based on strain, stress and vibration measurement has been installed [5].

During the operation of the online monitoring system, several faults were identified. One of them was the drive end (DE) bearing fault on the drive #1 for crane rotation. In March 2014, members of the port's maintenance team reported the occurrence of raised temperatures on the bearing housing and strange noise that is typical for bearings with an excessive looseness. A quick view on the frequency spectra and its comparison with the reference measurement shows the presence of raised carpet noise and higher vibration accelerations,

Table 6. Test rig data: confussion matrix for best performance MLP ANN with 9 input features

	AI	AII	AOK	AOL	AOU	AZ1	AZ3	BI	BII	BOK
Total	150	150	150	150	150	150	150	150	150	150
Correct	150	150	150	150	149	150	150	150	150	150
Incorrect	0	0	0	0	1	0	0	0	0	0
Correct (%)	100	100	100	100	99.333	100	100	100	100	100
Incorrect (%)	0	0	0	0	0.667	0	0	0	0	0
	BOL	BOU	BZ1	BZ3	CI	CII	COK	COL	COU	CZ1
Total	150	150	150	150	150	150	150	150	150	150
Correct	150	150	150	149	150	150	150	150	149	150
Incorrect	0	0	0	1	0	0	0	0	1	0
Correct (%)	100	100	100	99.333	100	100	100	100	99.333	100
Incorrect (%)	0	0	0	0.667	0	0	0	0	0.667	0
	CZ3	DI	DII	DOK	DOL	DOU	DZ1	DZ3	*All cases*	
Total	150	150	150	150	150	150	150	150	4200	
Correct	144	150	150	150	149	150	150	149	4189	
Incorrect	6	0	0	0	1	0	0	1	11	
Correct (%)	96	100	100	100	99.333	100	100	99.333	99.738	
Incorrect (%)	4	0	0	0	0.667	0	0	0.667	0.262	

Table 7. Test rig data: prediction errors for best performance MLP ANN with 9 input features

Phase	Target – input	Prediction - output	Number of occurrence
Training	AOU	AOL	1
Validation	BZ3	CZ3	1
Training	COU	DOL	1
Training	CZ3	DZ3	3
Testing	CZ3	DZ3	1
Validation	CZ3	DZ3	2
Training	DOL	COL	1
Training	DZ3	CZ3	1

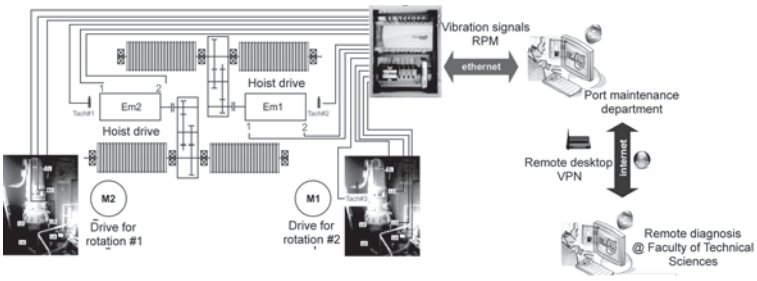


Fig. 7. Scheme of the online monitoring system installed on Ganz port crane [5]

typical for bearings with looseness. The same applies to the time waveform (Figure 8), where the presence of short duration high amplitude impacts can be seen.

Vibration acquisition included measurement of raw time waveforms and narrow band FFT in different frequency ranges with 3200 lines of resolution. Vibration features that were extracted are presented in the table 8.

Spectral extractions at bearing fault frequencies were not defined since the exact geometry of the bearing was unkown. Vibration trends for RMS and Acc features, as examples, are shown on Figure 9. Vibration data are shown in terms of index and they cover the complete development of the bearing fault development. Two important facts are worth to mention when observing Figure 9. First, it can be seen that Acc parameter has an increasing trend with the fault development while this can not be seen in case of RMS feature. Second, the machine under surveillance works under different loads and speeds. The data presented are not filtered to the specific load and speed levels. This is one of the reason for an absence of positive trend with RMS feature.

When assigning output labels to the input vectors a constant rate of bearing fault development is assumed. The original dataset consists of 1100 individual vectors and covers the data before the high temperatures occurrence up to the date when the machine was stopped for bearing replacement. The output labels are: Z1 (healthy bearing), Z2 (moderate looseness) and Z3 (excessive looseness). Using input matrix

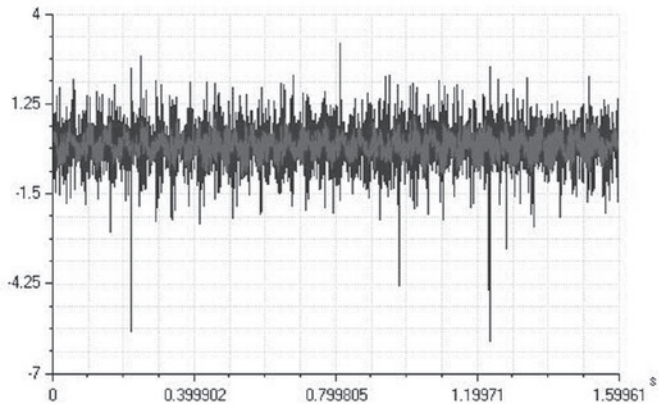


Fig. 8. Motor DE bearing: time waveforms for bearing in good and faulty state

with 9 features and 1100 labeled vectors several architectures of MLP ANN were tested. The best performance network had 13 neurons in the hidden layer and the accuracy of classification of 87.091%.

PCA on the input dataset resulted in extraction of 3 PCs with eigenvalues (Table 10). First two PCs have eigenvalues larger than 1 and projecting input features on these two PCs we are describing the input features dataset with 79.06% of variance.

The importance of each feature is evaluated through modeling power and the results are shown on Table 11. Kurtosis parameters and RMS of vibration velocity were excluded from the input feature matrix for MLP ANN. Using six input features several MLP ANN were tested. The best performance network had 11 neurons in the hidden layer and the accuracy of classification of 92.337%. As in the case of the test rig data, the input dataset with the reduced set of input features resulted in the MLP ANN with better classification rate (Table 12).

Table 8. Case study data: vibration features from the original dataset

Number	Label	Description	Unit
1	RMS	RMS of vibration velocity in range 2 Hz -1 kHz	mm/s
2	Acc	RMS of acceleration in range 2 Hz – 2 kHz	g
3	Acc_2_500	RMS of acceleration in range 2 Hz – 500 Hz	g
4	Acc_500_1000	RMS of acceleration in range 500 Hz – 1 kHz	g
5	Acc_1000_2000	RMS of acceleration in range 1 kHz – 2 kHz	g
6	PeakPeak	Peak-Peak value of acceleration raw time waveform	g
7	Def	Defect factor	g
8	Kurtosis MVX	Kurtosis parameter of the raw time waveform	-
9	Kurtosis Postprocess	Kurtosis parameter of the high pass filtered (500 Hz – 2kHz) time waveform	-

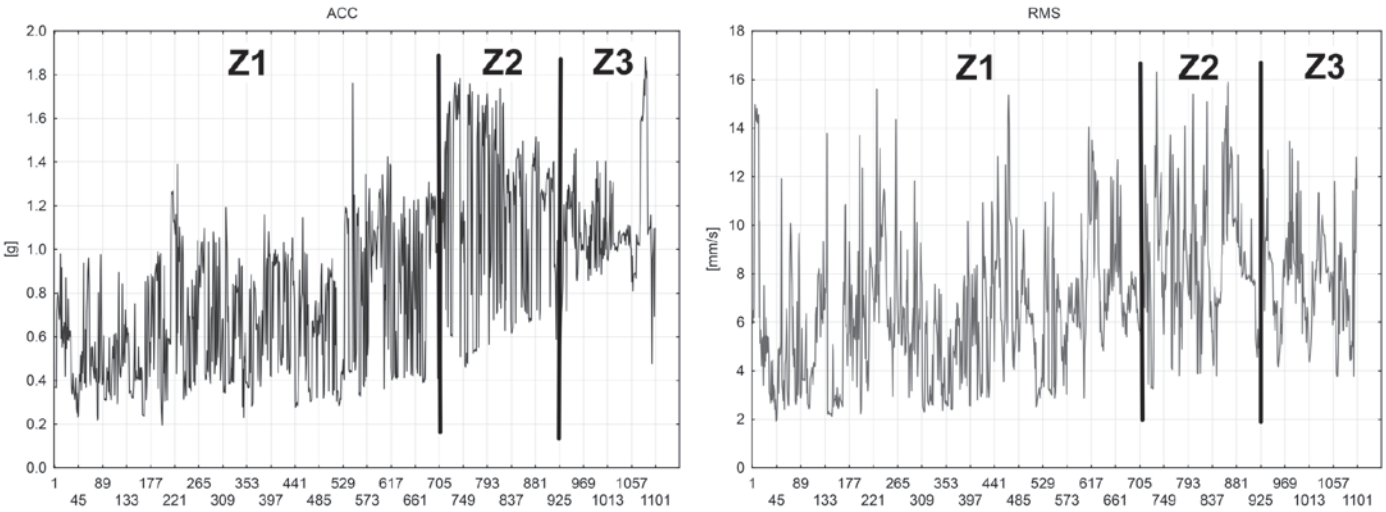


Fig. 9. Case study data: vibration trends for Acc and RMS features

Table 9. Case study data: confussion matrix for best performance MLP ANN with 9 input features

	Z1	Z2	Z3	All cases
Total	699	238	163	1100
Correct	689	116	153	958
Incorrect	10	122	10	142
Correct (%)	98.5694	48.7395	93.8650	87.091
Incorrect (%)	1.4306	51.2605	6.1350	12.909

Table 11. Case study data: modeling power for each feature and its importance to PC model

	Feature	Power	Importance
Acc	2	0.961748	1
PeakPeak	6	0.955579	2
Acc_500_1000	4	0.950476	3
Acc_1000_2000	5	0.939676	4
Acc_2_500	3	0.924648	5
Def	8	0.779565	6
Kurtosis MVX	9	0.672782	7
Kurtosis Postprocess	7	0.642616	8
RMS	1	0.288415	9

Table 10. Case study data: eignevalues, individual and cumulative variance for PCs

	Eigenvalues	% Total variance	Cumulative %
PC1	5.440144	60.44605	60.44605
PC2	1.675361	18.61512	79.06117
PC3	0.828145	9.20161	88.26278

Table 12. Case study data: confussion matrix for best performance MLP ANN with 6 input features

	Z1	Z2	Z3	All cases
Total	495	168	107	770
Correct	478	132	101	711
Incorrect	17	36	6	59
Correct (%)	96.5657	78.5714	94.3925	92.3377
Incorrect (%)	3.4343	21.4286	5.6075	7.6623

5. Conclusion

Implementation of ANN techniques in predictive maintenance of rotating machines based on vibration measurement and analysis can help in solving complex problems especially in presence of lack of highly skilled vibration analysts. However, for the reliable ANN an optimal set of vibration features must be defined. In this paper we demonstrated PCA as one of the possible technique for input features

space dimension reduction and optimal selection of input features based on vibration measurement.

In case of bearing faults combined with four levels of imbalance introduced on test rig, vibration features based on overall acceleration and vibration velocity as well as spectral extractions at bearing fault frequencies were the best choice of input features. The ANN with better performance is generated and classification rate is raised by 14%. In case of real life application on the port crane, the classification rate is raised by 5% through exclusion of overall vibration velocity and kurtosis parameters. In later case, RMS value of vibration velocity is irrelevant as an input feature, since all bearing fault cases correspond to the same level of imbalance. However in both cases kurtosis parameters are irrelevant. This is an interesting conclusion since it is believed that kurtosis parameter is the preferred vibration feature for bearing fault identification in the presence of variable load and speed of the machine.

Acknowledgement: *The research work financed with the means of the State Ministry of science and technological development (Serbia) in the years 2011-2015 as a research project "The application of information technologies in the ports of Serbia from the monitoring of machines to the networked system with the EU environment"*

References

1. Baillie D, Mathew J. Diagnosing Rolling Element Bearing Faults with Artificial Neural Networks. Acoustics Australia 1994; 22(3): 79-84.
2. Bartkowiak A , Zimroz R. Sparse PCA for gearbox diagnostics. Computer Science and Information Systems (FedCSIS) 2011: 25-31.
3. Bishop C. Neural Networks for Pattern Recognition. Oxford University Press, 1995.
4. Czech P. Classification of fault diagnosis in a gear wheel by used probabilistic neural network, fast Fourier transform and principal component analysis. Transport Problems: an International Scientific Journal 2007: 99-106.
5. Georgijevic M et all. The application of information technologies in the ports of Serbia from the monitoring of machines to the networked system with the EU environment. Project TR35036 financed by the Serbian Ministry of Science and Technological Development of Republic of Serbia, 2011-2015.
6. Hoon S, Worden K, Farrar C. Novelty Detection under Changing Environmental Conditions. SPIE's 8th Annual International Symposium on Smart Structures and Materials, Newport Beach, CA 2001.
7. Jyh-Shing R J, Chuen-Tsai S, Mizutani E. Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. MATLAB curriculum series. Prentice Hall, 1997.
8. Lazarz B, Wojnar G, Czech P. Early fault detection of toothed gear in exploitation conditions. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2011; 1(49): 68-77.
9. Malhi A, Gao R X. PCA-based feature selection scheme for machine defect classification. Instrumentation and Measurement, IEEE Transactions 2004; 53 (6): 1517 - 1525.
10. Sick B. On-Line And Indirect Tool Wear Monitoring In Turning With Artificial Neural Networks: A Review Of More Than A Decade Of Research. Mechanical Systems and Signal Processing 2002; 16(4): 487-546, http://dx.doi.org/10.1006/mssp.2001.1460.
11. Tranter J. Mobius Institute Vibration training manual - category III. Mobius Institute, 2009.
12. Veelenturf L. Analysis and Applications of Artificial Neural Networks. Prentice Hall, 1995.
13. Worden K, Sohn H, Farrar C. R. Novelty Detection in a Changing Environment: Regression and Interpolation Approaches. Journal of Sound and Vibration 2002; 258 (4): 741-761, http://dx.doi.org/10.1006/jsvi.2002.5148.
14. Zhong B. Developments in intelligent condition monitoring and diagnostics. System Integrity and Maintenance, 2nd Asia-Pacific Conference (ACSIM2000) Brisbane Australia 2000; 1-6.
15. Zimroz R, Bartkowiak A. Investigation on Spectral Structure of Gearbox Vibration Signals by Principal Component Analysis for Condition Monitoring Purposes. Journal of Physics 2011; Conference Series 305: article number 012075.
16. Zuber N. Automation of rotating machinery failures by the means of vibration analysis. PhD Thesis, Faculty of Technical Sciences - University of Novi Sad, 2010.
17. Zuber N, Ličen H, Bajrić R. An innovative approach to the condition monitoring of excavators in open pits mines, Technics Technologies Education Management - TTEM 2010; 5(3): 841-847.
18. Zuber N, Cvetkovic D, Bajric R. Multiple fault identification using vibration signal analysis and artificial intelligence methods. Acoustics & Vibration of Mechanical Structures 2013; 430: 63-69, http://dx.doi.org/10.4028/www.scientific.net/amm.430.63.
19. Zuber N, Cvetkovic D. Rolling element bearings fault identification in rotating machines, existing methods of vibration signal processing techniques and practical considerations. Acoustics & Vibration of Mechanical Structures 2013; 430: 70-77, http://dx.doi.org/10.4028/www.scientific.net/amm.430.70.
20. Zuber N, Bajric R, Sostakov R. Gearbox faults identification using vibration signal analysis and artificial intelligence methods. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2014; 16 (1): 61-65.

Ninoslav ZUBER
Faculty of Technical Sciences
University of Novi Sad
Trg Dositeja Obradovica 6, Serbia

Rusmir BAJRIĆ
EPC Elektroprivreda BiH
Coal Mines Kreka
Mije Keroševica 1,
Tuzla, Bosnia and Herzegovina

E-mails: zuber@uns.ac.rs, rusmir.bajric@gmail.com

Article citation info:
PAN Z, FENG J, SUN Q. Lifetime distribution and associated inference of systems with multiple degradation measurements based on gamma processes. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2016; 18 (2): 307–313, http://dx.doi.org/10.17531/ein.2016.2.20.

Zhengqiang PAN
Jing FENG
Quan SUN

LIFETIME DISTRIBUTION AND ASSOCIATED INFERENCE OF SYSTEMS WITH MULTIPLE DEGRADATION MEASUREMENTS BASED ON GAMMA PROCESSES

WYZNACZANIE ROZKŁADU CZASÓW ŻYCIA ORAZ WNIOSKOWANIE DLA SYSTEMÓW WYMAGAJĄCYCH POMIARÓW WSPÓŁISTNIEJĄCYCH DEGRADACJI W OPARCIU O PROCESY GAMMA

With development of science and technology, many engineering systems take on high reliable characteristic and usually have complex structure and failure mechanisms, with their reliability being evaluated by multiple degradation measurements. In certain physical situations, the degradation of these performance characteristics would be always positive and strictly increasing. Therefore, the gamma process is usually considered as a degradation process due to its independent and non-negative increments properties. In this paper, we suppose that a system has multiple dependent performance characteristics and that their degradation can be modeled by gamma processes. For such a multivariate degradation involving three or more performance characteristics, we propose to use a multivariate Birnbaum-Saunders distribution and its marginal distributions to approximate the reliability function and give the corresponding lifetime distribution. And then, the inferential method for the model parameters is developed. Finally, for an illustration of the proposed model and method, a simulated example is discussed and some computational results are presented.

***Keywords**: multiple degradation, lifetime distribution, gamma process, Birnbaum-Saunders distribution, MCMC method.*

Wraz z rozwojem nauki i techniki, powstaje coraz więcej systemów inżynieryjnych o wysokich parametrach niezawodnościowych, które zwykle charakteryzują się złożoną strukturą i złożonymi mechanizmami uszkodzeń. Ocena niezawodności w przypadku takich systemów wymaga pomiarów współwystępujących procesów degradacji . W pewnych sytuacjach fizycznych, degradacja właściwości użytkowych systemu będzie zawsze dodatnia oraz ściśle rosnąca. Proces degradacji jest zwykle procesem gamma, który charakteryzują niezależne i nieujemne przyrosty. W niniejszej pracy, założono, że system ma wiele zależnych charakterystyk pracy oraz że ich degradację można modelować procesem gamma. W przypadkach takiej wielowymiarowej degradacji obejmującej trzy lub więcej charakterystyk pracy zaproponowano zastosowanie rozkładu Birnbauma-Saundersa (uwzględniającego wiele zmiennych) oraz jego rozkładów brzegowych do aproksymacji funkcji niezawodności oraz określania odpowiadającego jej rozkładu czasu pracy. Opracowano metodę wnioskowania dla parametrów modelu. Wreszcie, dla zilustrowania proponowanego modelu oraz metody, omówiono przykład symulacyjny oraz przedstawiono niektóre wyniki obliczeniowe.

***Słowa kluczowe**: degradacja współistniejąca, rozkład czasów życia, proces gamma, rozkład Birnbauma-Saundersa, metoda MCMC.*

1. Introduction

In Astronautics, Aeronautics, Communications and other fields, there are many high reliable systems. Therefore, how to estimate the reliability of system has become a very significant effort to assess the reliability of critical systems to mitigate the probability of system failure and improve safety environment. Particularly, some systems may experience multiple degradation paths and they are either independent or dependent. When they are dependent, predicting system reliability becomes a challenging problem.

In the available literature, extensive research has been devoted to estimate reliability of systems/products experiencing bivariate or multivariate degradation data. Crk [3] assumed the system failure is governed by several independent mechanisms and presented an effective way to estimate the system's reliability by monitoring each performance characteristic. Wang and Coit [14] described a general modeling and analysis approach for reliability prediction based on degradation modeling, considering multiple degradation measures. Xu and Zhao [19] also considered this problem and introduced two methods to model and analyze systems with multiple degradation measures. First, they considered the correlation between the degradation measure and the failure event by introducing a probabilistic measure, and then proposed a state-space model to describe the evolution of the degradation process by incorporating both the degradation dynamics and random stress effects. Sari et al. [11] proposed a two-stage reliability model for bivariate degradation data. With the proposed model, not only the marginal reliability but also the system reliability can be assessed. The flexibility of the model to accommodate serial correlation of marginal degradation data, different marginal degradation distribution functions, and dependency between performance characteristics increases the probability to model and analyze the data more accurately compared with models with stricter assumptions. They analyzed the actual experiment data of the LED tube light system and gave some insights into the failure behavior

of the LED lamps system. Barker and Newby [1] used a multivariate Wiener process to describe the degradation of a complex multi-component system, and then provide an optimal non-periodic inspection policy for it. Son and Savage [13] proposed a design stage method for assessing performance reliability of systems with multiple time-variant responses due to component degradation. They assume that the system component degradation profiles over time is known and the degradation of the system is related to component degradation using mechanistic models. The cumulative distribution function of time to soft failure has been determined incrementally by summing probabilities of unions of failure sets established from shifted limit-state surfaces over time. They also present a set-theory method for assessing systems reliability where failure events may be described by time-variant parallel and/or series systems [12]. Li et al. [5] discussed a reliability model of a series system with dependent component degradation processes. Mercier et al. [6] discussed a track geometry model based on the observation of two dependent randomly increasing deterioration indicators through a bivariate gamma process constructed by trivariate reduction, and then give the intervention scheduling of a railway track. Pan and Balakrishnan [8] proposed a bivariate constant-stress accelerated degradation test model based on Wiener processes and Copulas. And the corresponding copula parameter is a function of the stress level that can be described by logistic function. Particularly, it is worth mentioning that Zhou et al. [20] proposed a bivariate degradation model based on copulas when a product has two performance characteristics and they can be governed by gamma processes. Furthermore, Pan and Balakrishnan [9] introduced the reliability model for the degradation of products with two performance characteristics by assuming that the performance characteristics are governed by gamma processes. In that case, they used a bivariate Birnbaum-Saunders distribution and its marginal distributions to approximate the reliability of the product. Wang et al. [17] gave the reliability equations when typical degradation and shocks are involved. The failure modes included catastrophic failure, degradation and failure due to shocks. Furthermore, they constructed a system reliability model on competitive failure processes under fuzzy degradation data and evaluated the proposed model by multi-state system reliability theory [18]. Wang and Pham [15] proposed a dependent competing risk model for a deteriorating system subject to shock processes, and a maintenance model involving imperfect maintenance actions. They also develop a dependent competing risk model for systems subject to multiple degradation processes and random shocks using time-varying copulas [16]. Peng et al. [10] presented a comprehensive Bayesian framework for the integration of multilevel heterogeneous data sets, including the pass-fail data, lifetime data and degradation data at different system levels, which gave a more practical tool for real engineering applications.

This paper extends the work of Pan and Balakrishnan [9] and assumes that a system has multiple dependent performance characteristics and that their degradation can be modeled by gamma processes. For such a multivariate degradation involving three or more performance characteristics, we propose to use a multivariate Birnbaum-Saunders distribution and its marginal distributions to approximate the reliability function and give the corresponding lifetime distribution.

The rest of the paper is organized as follows. In Section 2, the formulation of systems with multiple degradation measurements is given. In Section 3, the lifetime distribution and associated inference for such systems are presented. Section 4 discusses the estimation of model parameters. A simulated example is given to illustrate the model and method proposed in Section 5. Finally, some concluding remarks are made in Section 6.

2. Formulation of Systems with Multiple Degradation Measurements

2.1. Assumptions

(1) All the degradation paths of the systems are governed by gamma processes. Let $G_k(t)$ denotes the kth degradation path with shape parameter v_k and scale parameter u_k, where $t \geq 0$, and $G_k(0)$. Here, $k=1,\cdots,K$, and K is the number of the degradation paths. For a given t and Δt,

$$\Delta G_k(t) \sim Ga(v_k \Delta t, u_k)$$

where $\Delta G_k(t) = G_k(t+\Delta t) - G_k(t)$, and $Ga(v_k \Delta t, u_k)$ is a gamma distribution with shape parameter $v_k \Delta t$ and scale parameter u_k. The probability density function (PDF) of a random variable X having a gamma distribution with shape parameter v and scale parameter u can be given by:

$$g(x) = \frac{1}{\Gamma(v) u^v} x^{v-1} e^{-\frac{x}{u}} I_{(0,\infty)}(x),$$

where:

$$I_{(0,\infty)}(x) = \begin{cases} 1 & if\ x \in (0,\infty) \\ 0 & if\ x \notin (0,\infty) \end{cases},$$

and $\Gamma(v) = \int_0^\infty x^{v-1} e^{-x} dx$ is the complete gamma function.

(2) The degradation paths are not pairwise independent and the dependence of all the degradation paths can be described by variance-covariance matrix.

(3) All observations of the degradation paths are made at the same predetermined times (case of balanced data), and the measurement frequency is a constant denoted as Δt.

(4) In the case of different predetermined measurement times of the degradation paths, the dependency can be ignored, that is, we can consider $\Delta G_{k_1}(t_j)$ and $\Delta G_{k_2}(t_{j'})$ to be independent, when , $j \neq j'$ where $k_1=1,\cdots,K$, $k_2=1,\cdots,K$, $j=1,\cdots,M$, $j'=1,\cdots,M$.

2.2. Formulation of the Problem

For degradation paths involving independent nonnegative increments, gamma processes are more suitable for describing the deterioration of the system. In the present work, we suppose that a system has K degradation paths which are dependent each other and all the degradation paths are governed by gamma processes. For such a system, m measurements for all the paths are observed up to the termination time T, which results in degradation measurements $\boldsymbol{G}(t_j) = (G_1(t_j),\cdots,G_K(t_j))'$ corresponding to time t_j, $j=1,\cdots,M$. In general, the multiple degradation data for this model can be presented in the form:

$$\boldsymbol{G}_{K\times M} = \begin{pmatrix} G_1(t_1) & \cdots & G_1(t_M) \\ \vdots & \ddots & \vdots \\ G_K(t_1) & \cdots & G_K(t_M) \end{pmatrix}$$

For $k=1,\cdots,K$, let:

$$\Delta G_k(t_j) = G_k(t_j) - G_k(t_{j-1}), t_0 = 0$$

For each degradation path, by the independent increment property of the gamma process, we have independent but non-identical random variables:

$$\Delta G_k\left(t_j\right) \sim Ga\left(v_k \Delta t, u_k\right), \quad \Delta t = t_j - t_{j-1}. \tag{1}$$

So, the PDF of $\Delta G_k\left(t_j\right)$ is:

$$g_k\left(\Delta G_k\left(t_j\right)\right) = \frac{1}{\Gamma\left(v_k \Delta t\right)\left(u_k\right)^{v_k \Delta t}}\left(\Delta G_k\left(t_j\right)\right)^{v_k \Delta t - 1} \cdot \exp\left(-\frac{\Delta G_k\left(t_j\right)}{u_k}\right), \tag{2}$$

with the corresponding mean and variance given by:

$$E\left[\Delta G_k\left(t_j\right)\right] = u_k v_k \Delta t, \mathrm{Var}\left[\Delta G_k\left(t_j\right)\right] = u_k^2 v_k \Delta t.$$

The degradation increments of the paths, $\Delta G_k\left(t_j\right)$ can be normalized as:

$$Y_{kj} = \frac{\Delta G_k\left(t_j\right) - u_k v_k \Delta t}{\sqrt{v_k \Delta t} u_k}.$$

For $\boldsymbol{Y}_j = \left(Y_{1j}, \cdots, Y_{Kj}\right)$, it is known that they are i.i.d. vectors satisfying:

$$E\left(Y_{kj}\right) = 0, \quad Var\left(Y_{kj}\right) = 1, \tag{4}$$

with correlation coefficient being:

$$corr\left(Y_{k_1 j}, Y_{k_2 j}\right) = corr(\Delta G_{k_1}\left(t_j\right), \Delta G_{k_2}\left(t_j\right)) = \rho_{k_1 k_2}. \tag{5}$$

For each path, we know that the distribution of the first passage time to its threshold value can be approximated by Birnbaum-Saunders distribution [2]. Pan and Balakrishnan [9] gave the lifetime distribution of product with two dependent performance characteristics using bivariate Birnbaum-Saunders distribution. In the following section, we will discuss the lifetime distribution and associated inference of systems with any K degradation paths, where $K \geq 3$.

3. Lifetime Distribution and Associated Inference

Let T_k be the first passage times of kth degradation paths with the threshold values ω_k, $k=1,\cdots,K$. Any path of the K degradation pathes exceeds its threshold value, the system fails. Therefore, the system's reliability at time t can be expressed as:

$$\begin{aligned}
R(t) &= P\left(T_1 > t, \cdots, T_K > t\right) \\
&= P\left(G_1(t) < \omega_1, \cdots, G_K(t) < \omega_K\right) \\
&= P\left(\sum_{j=1}^{M} \Delta G_1\left(t_j\right) < \omega_1, \cdots, \sum_{j=1}^{M} \Delta G_K\left(t_j\right) < \omega_K\right) \\
&= P\left(\sum_{j=1}^{M} \frac{\Delta G_1\left(t_j\right) - v_1 u_1 \Delta t}{\sqrt{v_1 \Delta t} u_1} < \omega_1, \cdots, \sum_{j=1}^{M} \frac{\Delta G_K\left(t_j\right) - v_K u_K \Delta t}{\sqrt{v_K \Delta t} u_K} < \omega_K\right) \\
&= P\left(\frac{1}{\sqrt{M}} \sum_{j=1}^{M} Y_{1j} < \frac{\omega_1 - v_1 u_1 t}{\sqrt{v_1 t} u_1}, \cdots, \frac{1}{\sqrt{M}} \sum_{j=1}^{M} Y_{Kj} < \frac{\omega_K - v_K u_K t}{\sqrt{v_K t} u_K}\right).
\end{aligned}$$

According to multidimensional central limit theorem and the property of covariance, (6) can be approximated by:

$$R(t) = \Phi_K\left(\frac{\omega_1 - v_1 u_1 t}{\sqrt{v_1 t} u_1}, \cdots, \frac{\omega_K - v_K u_K t}{\sqrt{v_K t} u_K}; 0, \Sigma\right), \tag{7}$$

where:

$$\Sigma = \begin{pmatrix} Var(Y_{11}) & Cov(Y_{11}, Y_{21}) & \cdots & Cov(Y_{11}, Y_{K1}) \\ Cov(Y_{21}, Y_{11}) & Var(Y_{21}) & \cdots & Cov(Y_{21}, Y_{K1}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_{K1}, Y_{11}) & Cov(Y_{K1}, Y_{21}) & \cdots & Var(Y_{K1}) \end{pmatrix}, \tag{8}$$

and Φ_K is the multivariate normal distribution of a K-dimensional random vector. In terms of (4) and (5), $\boldsymbol{\Sigma}$ can be rewritten as:

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1K} \\ \rho_{21} & 1 & \cdots & \rho_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{K1} & \rho_{K2} & \cdots & 1 \end{pmatrix}.$$

Let:

$$\Sigma_{(k_1, \cdots, k_l)} = \begin{pmatrix} 1 & \rho_{k_1 k_2} & \cdots & \rho_{k_1 k_l} \\ \rho_{k_2 k_1} & 1 & \cdots & \rho_{k_2 k_l} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k_l k_1} & \rho_{k_l k_2} & \cdots & 1 \end{pmatrix}_{l \times l},$$

$$\Sigma_{(k_1, \cdots, k_l)}^{-k_m} = \Sigma_{(k_1, \cdots, k_{m-1}, k_{m+1}, \cdots, k_l)}.$$

(7) can, therefore, be represented as:

$$
\begin{aligned}
R(t) &\cong 1-\sum_{k=1}^{K}\Phi_1\left(\frac{\omega_K - v_K u_K t}{\sqrt{v_K t u_K}}\right) \\
&+\sum_{k_1=1}^{K-1}\sum_{k_2=k_1+1}^{K}\Phi_2\left(\frac{v_{k_1}\boldsymbol{u}_{k_1}t-\omega_{k_1}}{\sqrt{v_{k_1}tu_{k_1}}},\frac{v_{k_2}u_{k_2}t-\omega_{k_2}}{\sqrt{v_{k_2}tu_{k_2}}};0,\Sigma_{(k_1,k_2)}\right) \\
&+\cdots+(-1)^K\Phi_K\left(\frac{v_1u_1t-\omega_1}{\sqrt{v_1tu_1}},\cdots,\frac{v_Ku_Kt-\omega_K}{\sqrt{v_Ktu_K}};0,\Sigma_{(1,\cdots,K)}\right) \\
&=1-\sum_{\boldsymbol{k}=1}^{K}\Phi_1\left(\frac{1}{\alpha_k}\left(\sqrt{\frac{t}{\beta_k}}-\sqrt{\frac{\beta_k}{t}}\right)\right) \\
&+\sum_{k_1=1}^{K-1}\sum_{k_2=k_1+1}^{K}\Phi_2\left(\frac{1}{\alpha_{k_1}}\left(\sqrt{\frac{t}{\beta_{k_1}}}-\sqrt{\frac{\beta_{k_1}}{t}}\right),\frac{1}{\alpha_{k_2}}\left(\sqrt{\frac{t}{\beta_{k_2}}}-\sqrt{\frac{\beta_{k_2}}{t}}\right);0,\Sigma_{(k_1,k_2)}\right) \\
&+\cdots+(-1)^K\Phi_K\left(\frac{1}{\alpha_1}\left(\sqrt{\frac{t}{\beta_1}}-\sqrt{\frac{\beta_1}{t}}\right),\cdots,\frac{1}{\alpha_K}\left(\sqrt{\frac{t}{\beta_K}}-\sqrt{\frac{\beta_K}{t}}\right);0,\Sigma_{(1,\cdots,K)}\right)
\end{aligned}
\tag{9}
$$

where: $\alpha_k=\dfrac{1}{\sqrt{\omega_k/u_k}}$, and $\beta_k=\dfrac{\omega_k}{v_k u_k}$, $k=1,\cdots,K$. Kundu et al. [4] proposed a bivariate Birnbaum-Saunders distribution in their work. Here, we can extend their result to any -dimensional random vector $(T_1,T_2,\cdots,T_k)$, which have the multivariate Birnbaum-Saunders distribution, if the joint cumulative distribution function of $(T_1,T_2,\cdots,T_k)$ can be expressed as:

$$
\begin{aligned}
&\mathrm{P}\left(T_1<t_1,\cdots,T_K<t_K\right) \\
&=\Phi_K\left(\frac{1}{\alpha_1}\left(\sqrt{\frac{t}{\beta_1}}-\sqrt{\frac{\beta_1}{t}}\right),\cdots,\frac{1}{\alpha_K}\left(\sqrt{\frac{t}{\beta_K}}-\sqrt{\frac{\beta_K}{t}}\right);0,\Sigma_{(1,\cdots,K)}\right).
\end{aligned}
$$

According to (9), it is easy to obtain the lifetime distribution of $T, F(t)=1-R(t)$, and we find that $F(t)$ can be approximated by multivariate Birnbaum-Saunders distribution of a K-dimensional random vector and all its marginal distributions.

Let:

$$
U_k(t)=\frac{1}{\alpha_k}\left(\sqrt{\frac{t}{\beta_k}}-\sqrt{\frac{\beta_k}{t}}\right),
$$

and:

$$
\begin{aligned}
&\boldsymbol{U}_{(k_1,\cdots,k_l)}(t)=\left(U_{k_1}(t),\cdots,U_{k_l}(t)\right) \\
&\boldsymbol{U}^{-k_m}_{(k_1,\cdots,k_l)}(t)=\boldsymbol{U}_{(k_1,\cdots,k_{m-1},k_{m+1},\cdots,k_l)}(t), \\
&\boldsymbol{\Omega}_{(k_1,\cdots,k_l)}=\left\{\left(-\infty,U_{k_1}(t)\right),\cdots,\left(-\infty,U_{k_l}(t)\right)\right\}, \\
&\boldsymbol{\Omega}^{-k_m}_{(k_1,\cdots,k_l)}=\boldsymbol{\Omega}_{(k_1,\cdots,k_{m-1},k_{m+1},\cdots,k_l)}, \\
&\boldsymbol{x}_{(k_1,\cdots,k_l)}=\left(x_{k_1},\cdots,x_{k_l}\right), \\
&\boldsymbol{x}^{-k_m}_{(k_1,\cdots,k_l)}=\boldsymbol{x}_{(k_1,\cdots,k_{m-1},k_{m+1},\cdots,k_l)}, \\
&\boldsymbol{\rho}_{(k;k_1,\cdots,k_l)}=\left(\rho_{kk_1},\cdots,\rho_{kk_l}\right), \\
&\boldsymbol{\rho}^{-k_m}_{(k;k_1,\cdots,k_l)}=\boldsymbol{\rho}_{(\boldsymbol{k};k_1,\cdots,k_{m-1},k_{m+1},\cdots,k_l)}.
\end{aligned}
$$

According to (9), it can be expressed as:

$$
\begin{aligned}
R(t)&=1-\sum_{k=1}^{K}\Phi_1\left(U_k(t)\right)+\sum_{k_1=1}^{K-1}\sum_{k_2=k_1+1}^{K}\Phi_2\left(\boldsymbol{U}_{(k_1,k_2)}(t);0,\Sigma_{(k_1,k_2)}\right) \\
&+\cdots+(-1)^K\Phi_K(U_{(1,\cdots,K)}(t);0,\Sigma_{(1,\cdots,K)}).
\end{aligned}
\tag{10}
$$

Therefore, we can obtain the PDF of the lifetime of the system as follows:

$$
\begin{aligned}
f(t)&=-\frac{dR(t)}{dt} \\
&=\sum_{k=1}^{K}f_k(t;\alpha_k,\beta_k)-\sum_{k_1=1}^{K-1}\sum_{k_2=k_1+1}^{K}\frac{d\Phi_2\left(\boldsymbol{U}_{(k_1,k_2)}(t);0,\Sigma_{(k_1,k_2)}\right)}{dt} \\
&+\cdots+(-1)^{(K+1)}\frac{d\Phi_K\left(U_{(1,\cdots,K)}(t);0,\Sigma_{(1,\cdots,K)}\right)}{dt},
\end{aligned}
\tag{11}
$$

where: $f_k(t;\alpha_k,\beta_k)$, $k=1,\cdots,K$, is the PDF of Birnbaum-Saunders distribution given by:

$$
f_k(t;\alpha_k,\beta_k)=\frac{1}{2\sqrt{2\pi}\alpha_k\beta_k}\left[\left(\frac{\beta_k}{t}\right)^{\frac{1}{2}}+\left(\frac{\beta_k}{t}\right)^{\frac{3}{2}}\right]\exp\left[-\frac{1}{2\alpha_k^2}\left(\frac{t}{\beta_k}-2+\frac{\beta_k}{t}\right)\right], t>0
$$

We know that:

$$
\Phi_K\left(\boldsymbol{U}_{(1,\cdots,K)}(t);0,\Sigma_{(1,\cdots,K)}\right)=\int_{\Omega_{(k_1,\cdots,k_l)}}(2\pi)^{-\frac{K}{2}}|\Sigma_{(1,\cdots,K)}|^{-\frac{1}{2}}exp\left\{-\frac{1}{2}\boldsymbol{x}'(\Sigma_{(1,\cdots,K)})^{-1}\boldsymbol{x}\right\}d\boldsymbol{x}_{(1,\cdots,K)}.
$$

Define

$$
p(x_k)=\int_{\Omega^{-k}_{(k_1,\cdots,k_l)}}(2\pi)^{-\frac{K}{2}}\left|\Sigma_{(1,\cdots,K)}\right|^{-\frac{1}{2}}exp\left\{-\frac{1}{2}\boldsymbol{x}'\left(\Sigma_{(1,\cdots,K)}\right)^{-1}\boldsymbol{x}\right\}d\boldsymbol{x}^{-k}_{(1,\cdots,K)},
$$

then, we have:

$$
\frac{d\Phi_K\left(\boldsymbol{U}_{(1,\cdots,K)}(t);0,\Sigma_{(1,\cdots,K)}\right)}{dt}=\sum_{k=1}^{K}\frac{dU_k(t)}{dt}p\left(U_k(t)\right).
$$

Furthermore, we can obtain:

$$
p\left(U_k(t)\right)=P\left(\boldsymbol{x}^{-k}_{(1,\cdots,K)}\mid x_k=U_k(t)\right)\phi_1\left(U_k(t)\right).
$$

In terms of the property of multivariate normal distribution, the conditional distribution $P\left(\boldsymbol{x}^{-k}_{(1,\cdots,K)}\mid x_k=U_k(t)\right)$ is multivariate normal distribution with mean and covariance matrix:

$$
\begin{aligned}
&\tilde{\mu}^{-k}_{(1,\cdots,K)},=\boldsymbol{\rho}^{-k}_{(k;1,\cdots,K)}U_k(t), \\
&\tilde{\Sigma}^{-k}_{(1,\cdots,K)}=\Sigma^{-k}_{(1,\cdots,K)}-\left(\boldsymbol{\rho}^{-k}_{(k;1,\cdots,K)}\right)'\boldsymbol{\rho}^{-k}_{(k;1,\cdots,K)}.
\end{aligned}
$$

Thus,

$$
p\left(U_k(t)\right)=\phi_1\left(U_k(t)\right)\Phi_{K-1}\left(\boldsymbol{U}^{-k}_{(1,\cdots,K)}(t);\tilde{\mu}^{-k}_{(1,\cdots,K)},\tilde{\Sigma}^{-\boldsymbol{k}}_{(1,\cdots,K)}\right),
$$

and so:

$$\frac{d\Phi_K\left(\boldsymbol{U}_{(1,\cdots,K)}(t);0,\Sigma_{(1,\cdots,K)}\right)}{dt}$$
$$=\sum_{k=1}^{K}\frac{dU_k(t)}{dt}\phi_1\left(U_k(t)\right)\Phi_{K-1}\left(\boldsymbol{U}_{(1,\cdots,K)}^{-k}(t);\tilde{\boldsymbol{\mu}}_{(1,\cdots,K)}^{-k},\tilde{\Sigma}_{(1,\cdots,K)}^{-k}\right)$$
$$=\sum_{k=1}^{K}f_k\left(t;\alpha_k,\beta_k\right)\Phi_{K-1}\left(\boldsymbol{U}_{(1,\cdots,K)}^{-k}(t);\tilde{\boldsymbol{\mu}}_{(1,\cdots,K)}^{-k},\tilde{\Sigma}_{(1,\cdots,K)}^{-k}\right). \quad (12)$$

Similarly, for , $\forall l, 1<l<K$, we obtain:

$$\frac{d\Phi_l\left(\boldsymbol{U}_{(k_1,\cdots,k_l)}(t);0,\Sigma_{(k_1,\cdots,k_l)}\right)}{dt}$$
$$=\sum_{m-1}^{l}f_{k_m}\left(t;\alpha_{k_m},\beta_{k_m}\right)\Phi_{l-1}\left(\boldsymbol{U}_{(k_1,\cdots,k_l)}^{-k_m}(t);\tilde{\boldsymbol{\mu}}_{(k_1,\cdots,k_l)}^{-k_m},\tilde{\Sigma}_{(k_1,\cdots,k_l)}^{-k_m}\right) \quad (13)$$

According to (12) and (13), (11) can be rewritten as:

$$f(t)=-\frac{dR(t)}{dt}$$
$$=\sum_{k=1}^{K}f_k\left(t;\alpha_k,\beta_k\right)$$
$$-\sum_{k_1=1}^{K-1}\sum_{k_2=k_1+1}^{K}\sum_{m=1}^{2}f_{k_m}\left(t;\alpha_{k_m},\beta_{k_m}\right)\Phi_1\left(\boldsymbol{U}_{(k_1,k_2)}^{-k_m}(t);\tilde{\boldsymbol{\mu}}_{(k_1,k_2)}^{-k_m},\tilde{\Sigma}_{(k_1,k_2)}^{-k_m}\right)$$
$$+\cdots+(-1)^{(K+1)}\sum_{k=1}^{K}f_k(t;\alpha_k,\beta_k)\Phi_{K-1}\left(\boldsymbol{U}_{(1,\cdots,K)}^{-k}(t);\tilde{\boldsymbol{\mu}}_{(1,\cdots,K)}^{-k},\tilde{\Sigma}_{(1,\cdots,K)}^{-k}\right). \quad (14)$$

4. Estimation of Model Parameters

In this section, we discuss the estimation of the model parameters. The procedure consists of two steps. Firstly, we can estimate the covariance matrix Σ. Let, $\Delta\boldsymbol{G}\left(t_j\right)=\left(\Delta G_1\left(t_j\right),\cdots,\Delta G_K\left(t_j\right)\right)'$, $j=1,\cdots,M$ be the multivariate degradation increment vector. According to (5) and (8), the estimator of $\Sigma,\tilde{\Sigma}$ can be obtained by $\Delta\boldsymbol{G}\left(t_j\right)$. Its MLE for a sample of observations is:

$$\hat{\Sigma}\frac{1}{M}\sum_{j=1}^{M}\left(\Delta\boldsymbol{G}\left(t_j\right)-\Delta\bar{\boldsymbol{G}}(t)\right)\left(\Delta\boldsymbol{G}\left(t_j\right)-\Delta\bar{\boldsymbol{G}}(t)\right)', \quad (15)$$

where $\Delta\bar{\boldsymbol{G}}(t)=\left(\Delta\bar{G}_1(t),\cdots,\Delta\bar{G}_K(t)\right)'$, and $\Delta\bar{G}_k(t)=\frac{1}{M}\sum_{j=1}^{M}\Delta G_k\left(t_j\right)$.

As described in Section 3, for $\left(\frac{1}{\sqrt{M}}\sum_{j=1}^{M}Y_{1j},\cdots,\frac{1}{\sqrt{M}}\sum_{j=1}^{M}Y_{Kj}\right)$, its joint distribution can be approximated by multivariate normal distribution according to (6) and (7), and follows gamma distribution with shape parameter $v_k\Delta t$ and scale parameter . Thus, the log-likelihood function based on measurements on the K degradation paths is given by:

$$\ln L=\log\phi_K\left[\frac{1}{\sqrt{M}}\sum_{j=1}^{M}Y_{1j},\cdots,\frac{1}{\sqrt{M}}\sum_{j=1}^{M}Y_{Kj};0,\hat{\Sigma}\right]+\sum_{k=1}^{K}\sum_{j=1}^{M}\log g_k\left(\Delta G_k\left(t_j\right)\right). \quad (16)$$

From (16), we observe that the model is quite complicated from a computational viewpoint. For this reason, we make use of the Bayesian MCMC method for estimating the model parameters. In most practical applications in which Bayesian approach is used, it is difficult to compute analytically the posterior distribution. The MCMC method can be used to generate a sample from the posterior distribution large enough based on a Markov Chain so that any desired feature of the posterior distribution can be accurately estimated. Here, we assume that the prior distributions of all the unknown parameters are non-informative and we then utilize Matlab MCMC toolbox to implement the Metropolis-Hastings sampling after which we can estimate the model parameters of interest.

5. Simulated Example

To illustrate the model and inference method proposed in the preceding sections, we give a simulated example in this section. Firstly, the algorithm to simulate the data for multiple degradation measurements based on gamma processes is presented. And then, the estimation of the model parameters will be obtained according to the procedure in section 4.

5.1. Simulation of Data

Recently, copulas have become popular in simulation models. Copulas are functions that describe dependencies among variables, and provide a way to create distributions to model correlated multivariate data. About the details of copula function, please see the book written by Nelsen [7]. Using a copula, a data analyst can construct a multivariate distribution by specifying marginal univariate distributions, and choosing a particular copula to provide a correlation structure between variables. Bivariate distributions, as well as distributions in higher dimensions, are possible. To simulate dependent multivariate data using a copula, we have seen that we need to specify the copula family (we use Gaussian copula here), the rank correlations among variables, and the marginal distributions for each variable.

In this simulated example, we assume that a system with three degradation paths and the corresponding simulation parameters are denoted as , where and are the shape and scale parameters of gamma distributions, respectively, and is the kendall's tau which describes the rank correlations among the three degradation paths. There is a simple 1-1 mapping between Kendall's tau and the linear correlation coefficient:

$$\tau=(2/\pi)\times arcsin(\Sigma)\ \ or\ \ \Sigma=sin(\tau\times\pi/2). \quad (17)$$

Suppose that (t_j),$1\leq j\leq 500$ time increments are random chosen with similar magnitude as the data of system degradation. Let v=(0.04,0.05,0.06), u=(0.05,0.04,0.025) and τ=(1,0.6,0.4; 0.6,1,0.3; 0.4,0.3,1). According to (17), it is easy to know that Σ =(1,0.8090,0.5878; 0.8090,1,0.4540; 0.5878,0.4540,1), and let ρ=(0.8090,0.5878,0.4540). Then, 500 values of degradation measurements can be simulated according to the following procedure.

(1) specify the copula family as Gaussian copula with parameters (0,Σ);
(2) the rank correlations among the triple degradation paths are τ;
(3) the marginal distribution of the increments for i^{th} degradation path is gamma distribution with parameters (v_i, u_i), i=1,2,3.

Therefore, the correlated multivariate data $\Delta\mathbf{G}$ that we need can be generated by Matlab as follows. And then, the triple simulated degradation paths are obtained easily.

```
Z=mvnrnd([0,0,0],Σ,500);
U=normcdf(Z,0,1);
ΔG=[gaminv(U(:,1),v(1),u(1) ),gaminv(U(:,2),v(2),u(2) ),gaminv(U(:,3),v(3),u(3))]
```

5.2. Estimation of Parameters

According to the model parameters settings above, we generate 500 data sets to validate the model and method proposed. The parameters are estimated by using Matlab MCMC toolbox. For each group data, 5000 realizations of the parameters are from the posterior, and the last 4000 are used for estimation of the parameters. Therefore, we can obtain 500 groups estimated parameters. Figure 1 gives the hist graphs of the estimated parameters. And then, the mean, standard deviation, root mean squared error (RMSE) of the parameters can be obtained and listed in Table 1.

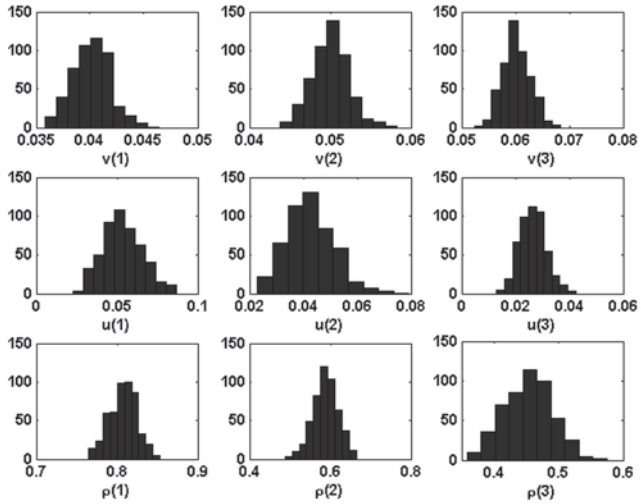


Fig. 1. The hist graphs of the estimated parameters

Table 1. Model parameter estimation results

Parameters	Results			True value
	mean	std	RMSE	
$v(1)$	0.0401	1.7345e−3	1.7381e−3	0.0400
$v(2)$	0.0500	2.2582e−3	2.2582e−3	0.0500
$v(3)$	0.0601	2.5248e−3	2.5277e−3	0.0600
$u(1)$	0.0536	0.0121	0.0126	0.0500
$u(2)$	0.0426	8.8039e−3	9.1847e−3	0.0400
$u(3)$	0.0265	4.8775e−3	5.0939e−3	0.0250
$\rho(1)$	0.8076	0.0168	0.0169	0.8090
$\rho(2)$	0.5879	0.0309	0.0309	0.5878
$\rho(3)$	0.4530	0.0367	0.0367	0.4540

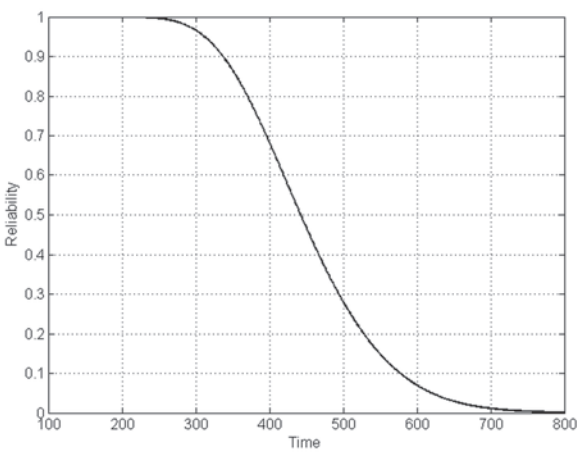


Fig. 2. The reliability of the system

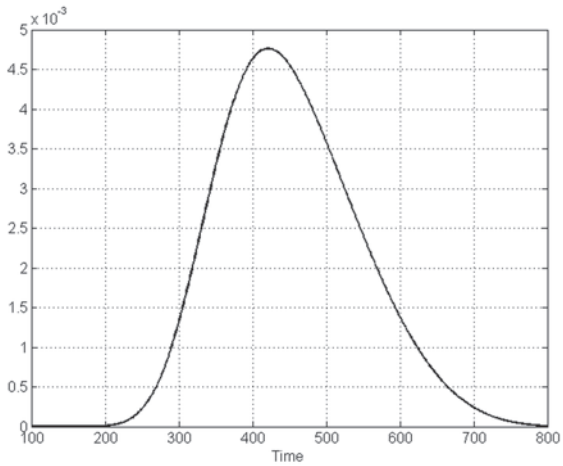


Fig. 3. The PDF of lifetime of the system

From Table 1, it is seen that the estimated parameters are very close to their true values. And it also reveals that RMSE of all the parameters are quite small. We, therefore, feel that the proposed model as well as the corresponding inferential method are performing very well, in this case.

Based on the estimates in Table 1, the reliability function and the corresponding PDF can be computed from (10) and (14) when $K=3$. The corresponding plots are illustrated in Figures 2 and 3, respectively.

6. Concluding Remark

The work of this paper is the extension of the results of Pan and Balakrishnan [9] that discussed the reliability model of bivariate degradation of products. We have introduced the lifetime distribution and associated inferential method of systems with multiple degradation measurements by assuming that all the degradation paths are governed by gamma processes. In this situation, we extend the work of Kundu et al. [4] and use a multivariate Birnbaum-Saunders distribution and its marginal distributions to approximate the reliability of the system. The inference of the model parameters is quite involved due to the complex form of the model and for this reason we used the Bayesian MCMC method for making inference. From the example in Section 5, we find that the proposed model as well as the inferential method for the model parameters work well.

Acknowledgements

This research was supported by the National Science Foundation of China under agreements 61273041 and 61304221 and supported by the Science Foundation of Hunan Province under agreements 14JJ3009.

1. Barker C T, Newby M J. Optimal non-periodic inspection for a multivariate degradation model. Reliability Engineering and System Safety 2009; 94: 33-43, http://dx.doi.org/10.1016/j.ress.2007.03.015.
2. Birnbaum Z W, Saunders S C. A new family of life distributions. Journal of Applied Probability 1969; 6: 319-327, http://dx.doi.org/10.2307/3212003.
3. Crk V. Reliability assessment from degradation data. In: The annual reliability and maintainability symposium-product quality & integrity 2004; Los Angeles: RAMS; 155-161.
4. Kundu D, Balakrishnan N, Jamalizadeh A. Bivariate birnbaum-saunders distribution and associated inference. Journal of Multivariate Analysis 2010; 101: 113-125, http://dx.doi.org/10.1016/j.jmva.2009.05.005.
5. Li J R, Coit D W, Elsayed A E. Reliability modeling of a series system with correlated or dependent component degradation processes. ICQR2MSE 2011; 388-393.
6. Mercier S, Meier-Hirmer C, Roussignol M. Bivariate gamma wear processes for track geometry modeling with application to intervention scheduling. Structure and Infrastructure Engineering 2012; 8:357-366, http://dx.doi.org/10.1080/15732479.2011.563090.
7. Nelsen P B. An Introduction to Copulas(2nd edn), New York, 2006.
8. Pan Z Q, Balakrishnan N. Constant-stress accelerated degradation model and inference. Communications in Statistics-Simulation and Computation 2011; 40: 259-269, http://dx.doi.org/10.1080/03610918.2010.534227.
9. Pan Z Q, Balakrishnan N. Reliability modeling of degradation of products with multiple performance characteristics based on gamma processes. Reliability Engineering and System Safety 2011; 96: 949-957, http://dx.doi.org/10.1016/j.ress.2011.03.014.
10. Peng W W, Huang H Z, Xie M, Yang Y J, Liu Y, A Bayesian Approach for System Reliability Analysis With Multilevel Pass-Fail, Lifetime and Degradation Data Sets. IEEE Transactions on Reliability 2013; 62: 689-699, http://dx.doi.org/10.1109/TR.2013.2270424.
11. Sari J K, Newby M J, Brombacher A C, Tang L C, Bivariate constant stress degradation model: LED lighting system reliability estimation with two-stage modeling. Quality and Reliability Engineering International 2009; 25: 1067-1084, http://dx.doi.org/10.1002/qre.1022.
12. Savage G J, Son Y K. The set-theory method for systems reliability of structures with degrading components. Reliability Engineering and System Safety 2011; 96: 108-116, http://dx.doi.org/10.1016/j.ress.2010.07.009.
13. Son Y K, Savage G J. Set theoretic formulation of performance reliability of multiple response time-variant systems due to degradations in system components. Quality and Reliability Engineering International 2007; 23: 171-188, http://dx.doi.org/10.1002/qre.783.
14. Wang P, Coit D W. Reliability prediction based on degradation modeling for systems with multiple degradation measures. In: The annual reliability and maintainability symposium - product quality & integrity 2004; Los Angeles: RAMS; 302-307, http://dx.doi.org/10.1109/rams.2004.1285465.
15. Wang Y P, Pham H. A multi-objective optimization of imperfect preventive maintenance policy for dependent competing risk systems with hidden failure. IEEE Transactions on Reliability 2011; 60: 770-781, http://dx.doi.org/10.1109/TR.2011.2167779.
16. Wang Y P, Pham H. Modeling the dependent competing risks with multiple degradation processes and random shock using time-varying copulas. IEEE Transactions on Reliability 2012; 61:13-22, http://dx.doi.org/10.1109/TR.2011.2170253.
17. Wang Z L, Huang H Z, Li Y F, Xiao N C, An Approach to Reliability Assessment Under Degradation and Shock Process. IEEE Transactions on Reliability 2011; 60: 852-863, http://dx.doi.org/10.1109/TR.2011.2170254.
18. Wang Z L, Huang H Z, Du L, Reliability Analysis on Competitive Failure Processes Under Fuzzy Degradation Data. Applied Soft Computing 2011; 11:2964-2973, http://dx.doi.org/10.1016/j.asoc.2010.11.018.
19. Xu D, Zhao W B. Reliability prediction using multivariate degradation data. In: The annual reliability and maintainability symposium-product quality & integrity 2005; Alexandria: RAMS; 337-341.
20. Zhou J L, Pan Z Q, Sun Q. Bivariate degradation modeling based on gamma process. WCE 2010; London; III: 1783-1788.

Zhengqiang PAN
Jing FENG
Quan SUN
College of Information System and Management
National University of Defense Technology
Sanyi Avenue, Changsha, China

E-mail: panzhq.yy@hotmail.com, panzhengqiang@nudt.edu.cn

INFORMATION FOR AUTHORS

Eksploatacja i Niezawodnosc – Maintenance and Reliability – the journal of the Polish Maintenance Society, under the scientific supervision of the Polish Academy of Sciences (Branch in Lublin), published four times a year.

The scope of the Quarterly

The quarterly *Eksploatacja i Niezawodnosc – Maintenance and Reliability* publishes articles containing original results of experimental research on the durability and reliability of technical objects. We also accept papers presenting theoretical analyses supported by physical interpretation of causes or ones that have been verified empirically. *Eksploatacja i Niezawodność – Maintenance and Reliability* also publishes articles on innovative modeling approaches and research methods regarding the durability and reliability of objects.

The following research areas are particularly relevant to the journal:

1. degradation processes of mechanical and biomechanical systems,
2. diagnosis and prognosis of operational malfunctions and failures.
3. analysis of failure risk/wear,
4. reliability-and-environmental-safety engineering in the design, manufacturing and maintenance of objects,
5. management and rationalization of object maintenance,
6. risk management in the processes of operation and maintenance,
7. the human factor and human reliability in operation and maintenance systems.

Terms and Conditions of Publication

The quarterly *Eksploatacja i Niezawodnosc – Maintenance and Reliability* publishes only original papers written in English or in Polish with an English translation. Translation into English is done by the Authors after they have received information from the Editorial Office about the outcome of the review process and have introduced the necessary modifications in accordance with the suggestions of the referees! Acceptance of papers for publication is based on two independent reviews commissioned by the Editor.

The quarterly *Eksploatacja i Niezawodnosc – Maintenance and Reliability* proceeds entirely online at submission.ein.org.pl

Technical requirements

- After receiving positive reviews and after acceptance of the paper for publication, the text must be submitted in a Microsoft Word document format.
- Drawings and photos should be additionally submitted in the form of high resolution separate graphical files in the TIFF, SVG, AI or JPG formats.
- A manuscript should include: names of authors, title, abstract, and key words that should complement the title and abstract (in Polish and in English), the text in Polish and in English with a clear division into sections (please, do not divide words in the text); tables, drawings, graphs, and photos included in the text should have descriptive two-language captions, if this can be avoided, no formulae and symbols should be inserted into text paragraphs by means of a formula editor; references (written in accordance with the required reference format); author data – first names and surnames along with scientific titles, affiliation, address, phone number, fax, and e-mail address.

The Editor reserves the right to abridge and adjust the manuscripts. All submissions should be accompanied by a submission form.

Detailed instructions to Authors, including evaluation criteria can be found on the journal's website: www.ein.org.pl

Editor contact info

Editorial Office of „Eksploatacja i Niezawodnosc - Maintenance and Reliability"
Nadbystrzycka 36, 20-618 Lublin, Poland
e-mail: office@ein.org.pl

INFORMATION FOR SUBSCRIBERS

Fees

Yearly subscription fee (four issues) is 100 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.

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