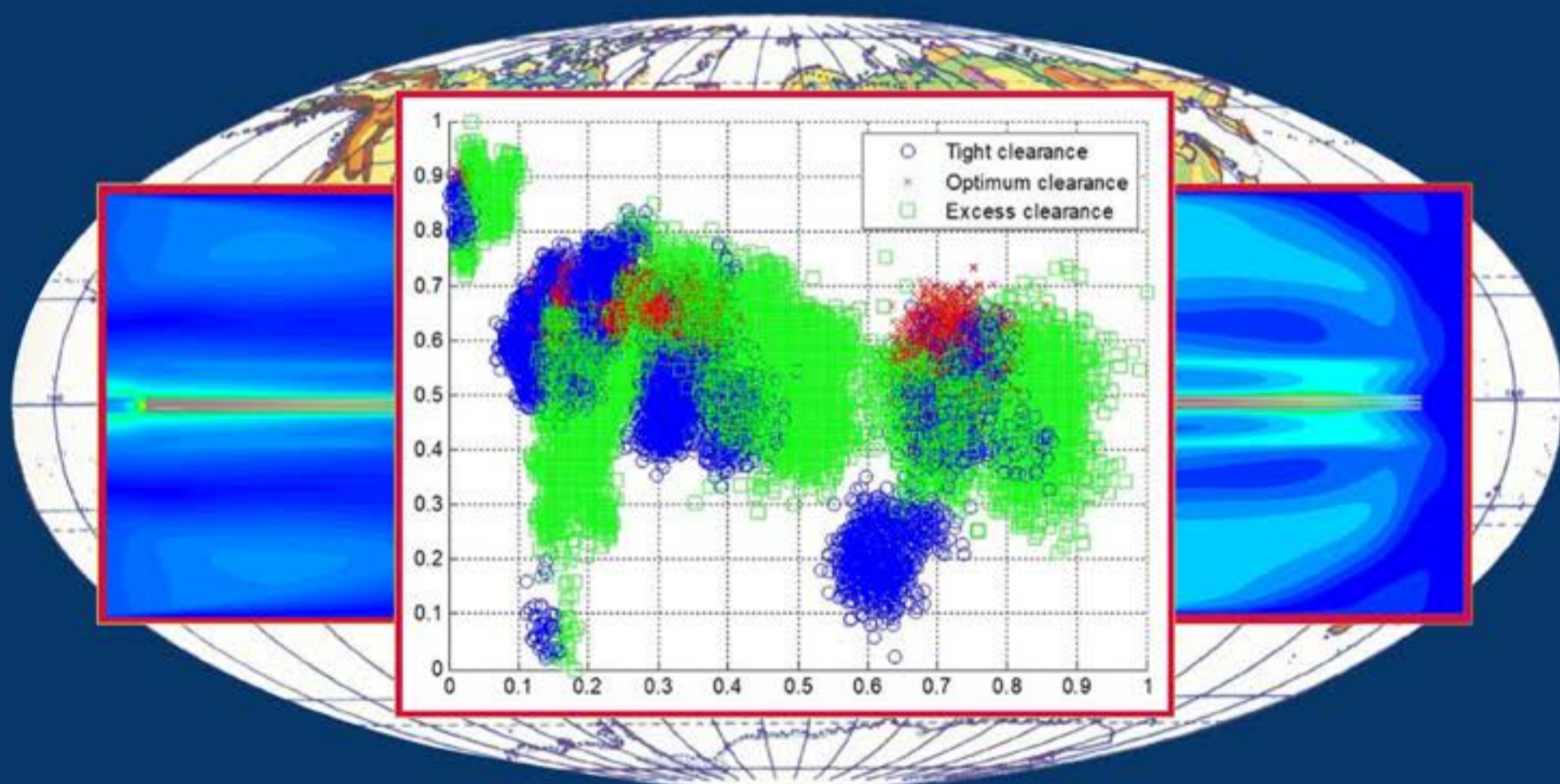


# EKSPLOATACJA I NIEZAWODNOŚĆ

## MAINTENANCE AND RELIABILITY

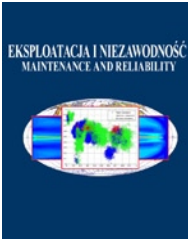


Polskie Naukowo Techniczne Towarzystwo Eksploatacyjne  
Warszawa

Polish Maintenance Society  
Warsaw

## TABLE OF CONTENTS

Łukasz Knypiński	<b>Constrained optimization of line-start PM motor based on the gray wolf optimizer</b> .....	1
Xiaofei Cui, Xiaoxia Liang, Ujjwal Bharadwaj	<b>A framework for corrosion assessment in metallic structures, from data analysis to risk based inspection</b> .....	11
Robert Konowrocki, Dariusz Kalinowski, Tomasz Szolc, Artur Marczewski	<b>Identification of safety hazards and operating conditions of the low-floor tram with independently rotating wheels with various drive control algorithms</b> .....	21
Zhixin Xu, Dingqing Guo, Jinkai Wang, Xueli Li, Daochuan Ge	<b>A numerical simulation method for a repairable dynamic fault tree</b> .....	34
Bartłomiej Ambrożkiewicz, Arkadiusz Syta, Nicolas Meier, Grzegorz Litak, Anthimos Georgiadis	<b>Radial internal clearance analysis in ball bearings</b> .....	42
Norbert Kępczak, Grzegorz Bechciński, Radosław Rosik	<b>Experimental verification of the deep hole boring bar model</b> .....	55
Jiri Hlinka, Rostislav Kostial, Michaela Horpatzka	<b>Application of enhanced methods for safety assessment of FADEC</b> .....	63
Liming Fan, Kunsheng Wang, Dongming Fan	<b>A combined universal generating function and physics of failure Reliability Prediction Method for an LED driver</b> .....	74
Przemysław Podulka	<b>Fast Fourier Transform detection and reduction of high-frequency errors from the results of surface topography profile measurements of honed textures</b> .....	84
Piotr Kulinowski, Piotr Kasza, Jacek Zarzycki	<b>Identification of the operating parameters of the friction drum drive in industrial conditions</b> .....	94
Andrius Ružinskas, Martin Giessler, Frank Gauterin, Klaus Wiese, Marijonas Bogdevičius	<b>Experimental investigation of tire performance on slush</b> .....	103
Isaac Segovia Ramirez, Behnam Mohammadi-Ivatloo, Fausto Pedro García Márquez	<b>Alarms management by supervisory control and data acquisition system for wind turbines</b> .....	110
Grzegorz Bocewicz, Eryk Szwarc, Jarosław Wikarek, Peter Nielsen, Zbigniew Banaszak	<b>A competency-driven staff assignment approach to improving employee scheduling robustness</b> .....	117
Murat Basaran, Mehmet Fidan	<b>Induction motor fault classification via entropy and column correlation features of 2D represented vibration data</b> .....	132
Mirosław Kowalski, Mariusz Izdebski, Jolanta Żak, Paweł Gołda, Jerzy Manerowski	<b>Planning and management of aircraft maintenance using a genetic algorithm</b> .....	143
Gao Zhiyong, Li Jiwu, Wang Rongxi	<b>Prognostics uncertainty reduction by right-time prediction of remaining useful life based on hidden Markov model and proportional hazard model</b> .....	154
Marek Rośkowicz, Jan Godzimirski, Michał Jasztal, Jarosław Gąsior	<b>Improvement of fatigue life of riveted joints in helicopter airframes</b> .....	165
Tadeusz Dziubak, Tadeusz Wysocki, Sebastian Dziubak	<b>Selection of vehicles for fleet of transport company on the basis of observation of their operational reliability</b> .....	184
Varun Kumar, Girish Kumar, Rajesh Kumar Singh, Umang Soni	<b>Degrading systems availability analysis: analytical semi-Markov approach</b> .....	195



Article citation info:

Knypiński Ł. Constrained optimization of line-start PM motor based on the gray wolf optimizer. *Eksploatacja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 1–10, <http://dx.doi.org/10.17531/ein.2021.1.1>.

## Constrained optimization of line-start PM motor based on the gray wolf optimizer

Indexed by:



Łukasz Knypiński<sup>a</sup>

<sup>a</sup>Institute of Electrical Engineering and Electronics, Faculty of Control, Robotics and Electrical Engineering, Poznan University of Technology, ul. Piotrowo 3A, 60-965 Poznań, Poland

### Highlights

- Constrained optimization procedure for the line-start permanent magnet motor.
- The mathematical model of designed motor was developed on the basis of Finite Element Method (FEM).
- The objective function contains product of multiplication of three maintenance parameters of the motor.
- The performance reliability of three heuristic algorithm was compared, i.e. (a) GWO, (b) PSO, (c) BA and (d) GA.

### Abstract

This paper presents the algorithm and computer software for constrained optimization based on the gray wolf algorithm. The gray wolf algorithm was combined with the external penalty function approach. The optimization procedure was developed using Borland Delphi 7.0. The developed procedure was then applied to design of a line-start PM synchronous motor. The motor was described by three design variables which determine the rotor structure. The multiplicative compromise function consisted of three maintenance parameters of designed motor and one non-linear constraint function was proposed. Next, the result obtained for the developed procedure (together with the gray wolf algorithm) was compared with results obtained using: (a) the particle swarm optimization algorithm, (b) the bat algorithm and (c) the genetic algorithm. The developed optimization algorithm is characterized by good convergence, robustness and reliability. Selected results of the computer simulation are presented and discussed.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

heuristic algorithms, gray wolf algorithm, constrained optimization, external penalty function, line-start PM synchronous motor.

### Nomenclature:

$A_k, C_k$  – factors of the gray wolf algorithm,  
 $\mathbf{X}_{k-1}^p$  – vector of the location of the prey,  
 $r, r_1, r_2, \kappa$  – random numbers from range (0, 1),  
 $b_k$  – coefficient describing the ability of migration of wolves,  
 $\mathbf{X}_{k-1}^\alpha, \mathbf{X}_{k-1}^\beta, \mathbf{X}_{k-1}^\delta$  – vectors of positions of the  $\alpha, \beta$  and  $\delta$  wolves, respectively,  
 $k$  – number of iteration of optimization algorithm,  
 $\mathbf{x}_i, \mathbf{v}_i$  – vectors of velocity and positions the  $i$ -th particle for the PSO algorithm,  
 $\mathbf{x}_B$  – vector of leader position,  
 $\mathbf{x}_L$  – vector of best self position in previous iterations,  
 $w_1$  – weight of inertia,  
 $c_1, c_2$  – learning factors,  
 $F^i, A^i, P^i$  – frequency, loudness and pulse emission of  $i$ -th bat,  
 $\zeta, \gamma$  – bat algorithm constants,  
 $P_N, V_N, I_N, n_N, T_N$  – rated power, voltage, current, velocity and torque,  
 $\cos\varphi, \eta$  – power factor and efficiency,  
 $T_{\max}$  – peak torque,  
 $l_m, g_m$  – length and thickness of the permanent magnet,  
 $r_m$  – distance between poles,

$T_{80}$  – electromagnetic torque for a speed of about 0,8 of synchronous speed,  
 $B_r, H_c$  – residual induction and coercive force of the permanent magnet,  
 $D_o, D_i$  – outer and inner diameter of the stator,  
 $L_s$  – stack length,  
 $N_s, N_r$  – number of the stator and the rotor slots,  
 $d_o, d_i$  – outer diameter of the rotor and diameter of the shaft,  
 $q_1, q_2, q_3$  – weighting coefficient of the objective function,  
 $\mathbf{s}$  – the vector made up of design parameters,  
 $f(\mathbf{s}), h(\mathbf{s})$  – objective and modified objective functions,  
 $g(\mathbf{s})$  – non-linear constraint function,  
 $p(\mathbf{s})$  – penalty term,  
 $n$  – number of external penalty function iteration,  
 $\sigma$  – penalty factor,  
 $m_{pm}(\mathbf{s}), m_{pm}^i$  – permanent magnet mass and imposed mass of the permanent magnet,  
 $N$  – number of individuals in optimization procedure,  
 $N_{of}$  – number of calls function.

E-mail addresses: [lukasz.knypinski@put.poznan.pl](mailto:lukasz.knypinski@put.poznan.pl)

# 1. Introduction

At present, complex models of the various phenomena in the designed device are used in the design process. These models consist of: (a) an equation describing of the electromagnetic field, (b) supply circuit equations, (c) an equation describing rotational equilibrium and also (d) an equation describing the thermal phenomena [2, 3, 9, 11, 33]. All of these phenomena are usually taken into account when the finite element method (FEM) is used. FEM models are very complex, therefore, the optimization processes which incorporate them are very time-consuming.

Constrained optimization (CO) is the most important tool in the process of modern design of electromagnetic devices, such as: electric motors, transformers and electromagnetic actuators. The design problem must often necessarily to take into account constraints related to the dimensions of the devices and the imposed functional parameters. Solutions of these constrained problems require new, more and more effective methods of optimization.

In the optimization process the objective functions have usually economic features and are closely connected with the minimization of the production costs and power losses [19, 31]. They may also include components related to the protection of natural environment.

In international literature, intensive development of new optimization algorithms has been observed. Currently heuristic (non-deterministic) algorithms [1, 6] are being most dynamically developed. These types of optimization algorithms are especially effective in solving the design challenge connected with electromagnetic converters [8, 10]. Classification of the optimization algorithms proposed by the author is presented in Figure 1.

In the last two decades, an increased number of papers devoted to the application of different probabilistic algorithms elaborated on the basis of the natural environment (nature-inspired algorithm [36]) has been observed in relation to the design of PM motors. Among PM motors, brushless direct-current motors (BLDC) and permanent magnet synchronous motors (PMSM) are currently developing most dynamically. The Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) algorithms are often used to optimize these two types of motors [7, 11, 14] and other technical problem [32]. In order to achieve a better convergence of optimization processes, scientists are continuously applying new optimization algorithms. New optimization algorithms are being used more and more often. This group of algorithms consists of: the Ant Colony Optimization (ACO) algorithm [26], the Cuckoo Search (CS) algorithm [5, 37], the Bat Algorithm (BA) [4] and the Gray Wolf Optimizer (GWO) algorithm [14]. Optimization calculations are often performed on simplified models (analytical or lumped parameters) of the PM motors [14, 26, 27]. There are not many articles on the subject of optimization algorithms using 2D FEA models and gray wolf optimization.

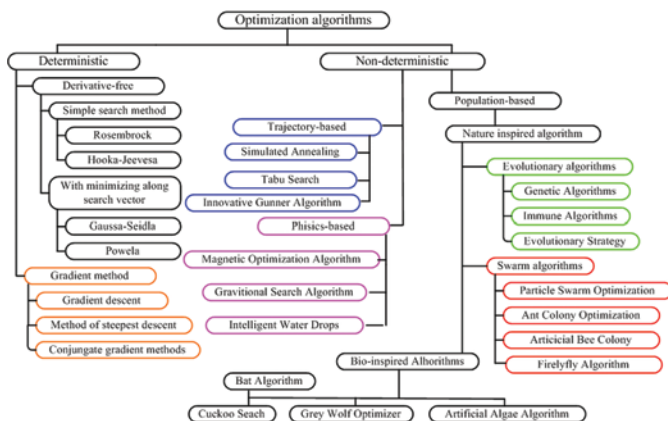


Fig. 1. The classification schema of the optimization algorithms

The aim of this paper is to develop a constrained optimization procedure based on the GWO algorithm. The developed algorithm was

employed for the constrained optimization of a line-start permanent magnet motor (LSPMSM). Additionally, the convergence of the developed procedure in comparison to other nature-inspired algorithms was investigated.

The organization and social rules in a wolf-pack and the mathematical rules of the gray wolf method are presented in details and are described in section 2. Also, the main equations for the PSO and the BA are presented in section 2. Next, the algorithm and computer software for the constrained optimization of the LSPMSM are presented in section 3. In section 4, a comparison of the performance of the GWO, the PSO, the BA and the GA has been performed. The summarizing conclusions are discussed in section 5.

## 2. Nature inspired optimization algorithms

In general, the PSO and GA algorithms are most often used for those problems related to electromagnetic design optimization. Recently, a rapid development of metaheuristic algorithms has been observed, including: CS, BA, GWO and others. These algorithms are increasingly used to optimize and design technical devices. In the international literature, there are no papers which confirm the advantages of these algorithms over PSO and GA.

### 2.1. Mathematical description of the organization and social rungs in wolf pack

Wolves represent the Carnivora order and are members of the Canidae family. Wolves are social animals organized in groups called packs. Wolf demonstrates an expanded system of social rungs (hierarchy), which decide the position of each member within the wolf pack.

Each pack occupies a specific area where it lives and hunts and which it protects from other wolves. Just like any other social group, the pack needs a leader – an individual to keep the order in the group. The leader of pack is the **alpha individual** ( $\alpha$ ). The alpha (the most well adapted individual) always leads the wandering group, initiates attacks on other wolves infringing on the pack's territory, initiates hunting as well as all other activities of the wolf pack [20]. The wolf pack hierarchy is linear, the group leader is the wolf who won direct confrontations with other pack members. Family relations can also determine the pack's hierarchy [24]. A very important role in the pack is the **beta wolf** ( $\beta$ ). This individual submits only to the alpha, while being stronger than other members of the group. For groups living in the wild, the beta individual is usually the strongest individual, but less ingenious and less intelligent than the alpha. The alpha and beta individuals are a complementary pair. The beta male is strong, brave and confident but submits to the alpha male. The beta assumes the pack leadership when the alpha leaves the group grows old or dies.

The third level of the pack hierarchy are the  **$\delta$  individuals**, which are weaker than  $\alpha$  and  $\beta$  individuals but stronger than **omega individuals** ( $\omega$ ). The omega individuals form the lowest rung of the pack hierarchy [22]. They submit to all other members of the pack; they are the last to be allowed to feed and are often used as "scapegoats". Most often, the omega individuals are the oldest or frailest.

The mathematical model of the GWO algorithm was based on the wolves hunting behavior. Depending on the size of their prey, wolves can use different hunting tactics.

The wolves recognize the herds of potential victims before starting the hunt. This is the **searching for prey stage** [25]. At the beginning, the wolves choose their victim. Then the predator get very close to it. If the prey does not get scared by the predator or even starts to approach it, the wolf retreats. If the animal starts to run away, the wolves immediately start the chase. This is the **chase stage**. During the chase, the wolves often change which of them, pursues the prey. The main purpose of this tactic is to disorientate the victim. The wolves may also force the animal to change its direction of escape. After stopping the prey, wolves immediately attack prey.

In the numerical implementation of the GWO algorithm it was assumed that the global extreme point is situated between the three leaders ( $\alpha$ ,  $\beta$  and  $\delta$ ). Thus, the vector of position for each  $i$ -th individual in  $k$ -th iteration ( $k$ -th time step) is determined as [17]:

$$\mathbf{X}_k^i = \mathbf{X}_{k-1}^p - A_k \left| C_k \mathbf{X}_{k-1}^p - \mathbf{X}_{k-1}^i \right|, \quad (1)$$

where  $A_k$  and  $C_k$  are the factors of the gray wolf algorithm,  $\mathbf{X}_{k-1}^p$  is the location vector of the prey (the global extreme).

The factor  $A_k$  and  $C_k$  are different for the  $\alpha$ ,  $\beta$  and  $\delta$  individuals in the developed algorithm and are calculated as follows:

$$A_k^\alpha = 2b_k r_1 \quad A_k^\beta = 2b_k r_2 \quad A_k^\delta = 2b_k r_3, \quad (2)$$

$$C_k^\alpha = 2r_4 \quad C_k^\beta = 2r_5 \quad C_k^\delta = 2r_6, \quad (3)$$

in which:  $r_1$  to  $r_6$  are the numbers randomly selected from range (0, 1),  $b_k$  is the factors which describes the ability of wolves to migrate in the permissible area of the solved task. In the case of a large value of this parameter the individuals can move freely in the area of the considered task. The algorithm then has the properties of a global search algorithm. A low value of this parameter means the algorithm has the properties of a local search method. The value of the  $b$  parameter is usually decreased from 2 to 0 [17, 25].

In the developed algorithm the author proposed a decrease in the value of the coefficient during the process of optimization. The value of  $b$  in subsequent  $k$ -th iterations is calculated according to the following formula [15]:

$$b_k = 0.97 + 5e^{-k}. \quad (4)$$

According to the developed procedure the three best-adapted wolves are represented as individuals  $\alpha$ ,  $\beta$  and  $\delta$ . Thus, there is one  $\alpha$ , one  $\beta$  and one  $\delta$  in each iteration. In order to determine the new position of  $i$ -th  $\omega$  individual, it is necessary to calculate the distance of this individual from the best wolves in the pack [38]:

$$D^\alpha = \left| C_k^\alpha \mathbf{X}_{k-1}^\alpha - \mathbf{X}_{k-1}^i \right|, \quad D^\beta = \left| C_k^\beta \mathbf{X}_{k-1}^\beta - \mathbf{X}_{k-1}^i \right|, \quad D^\delta = \left| C_k^\delta \mathbf{X}_{k-1}^\delta - \mathbf{X}_{k-1}^i \right|, \quad (5)$$

where:  $\mathbf{X}_{k-1}^\alpha$ ,  $\mathbf{X}_{k-1}^\beta$  and  $\mathbf{X}_{k-1}^\delta$  denote the positions of the three best individuals ( $\alpha$ ,  $\beta$  and  $\delta$ ) in the previous iteration.

Finally, the new position for each  $i$ -th  $\omega$  individual in the  $k$ -th iteration ( $k$ -th time step) is determined using the following equation:

$$(\mathbf{X}^\omega)_k^i = \frac{\mathbf{X}_1^\alpha + \mathbf{X}_2^\beta + \mathbf{X}_3^\delta}{3}, \quad (6)$$

where:  $\mathbf{X}_1^\alpha = \mathbf{X}_{k-1}^\alpha - A_k^\alpha D^\alpha$ ,  $\mathbf{X}_2^\beta = \mathbf{X}_{k-1}^\beta - A_k^\beta D^\beta$ ,  $\mathbf{X}_3^\delta = \mathbf{X}_{k-1}^\delta - A_k^\delta D^\delta$ .

The flowchart of the GWO method is illustrated in Figure 2.

## 2.2. The classical Particle Swarm Optimization

The classical PSO algorithm was developed in 1995 [30]. This algorithm mimics the foraging behavior of flocks of birds and fish shoals. The optimization process uses the interaction between the leader (the best particle in the group) and the rest of the particles. All of the particles form the swarm system. The best-adapted particle is

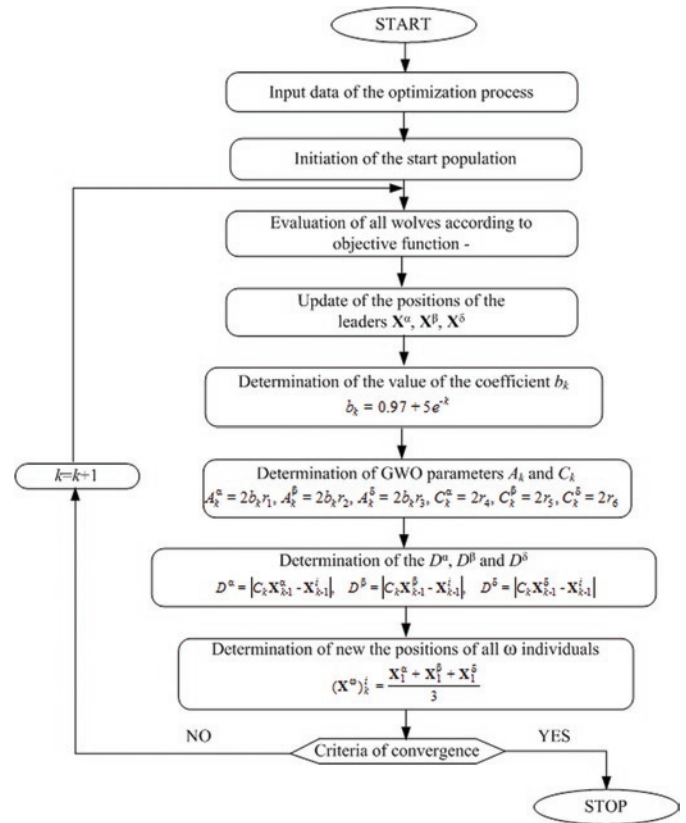


Fig. 2. The block diagram of the elaborated GWO algorithm

the leader of the swarm. Each particle is described by two vectors: (a) the position vector ( $\mathbf{x}^i$ ) and (b) the velocity vector ( $\mathbf{v}^i$ ). In order to determine the new position of the  $i$ -th particle, the position vector of the leader ( $\mathbf{x}_B$ ), the local best known position of the  $i$ -th particle and the vector of velocity from the previous iteration are taken into account. The velocity vector of the  $i$ -th particle in the  $k$ -th iteration is calculated according to the following formula:

$$\mathbf{v}_k^i = w_1 \mathbf{v}_{k-1}^i + c_1 r_1 (\mathbf{x}_L^i - \mathbf{x}_{k-1}^i) + c_2 r_2 (\mathbf{x}_B - \mathbf{x}_{k-1}^i), \quad (7)$$

where:  $w_1$  is the weight of inertia,  $c_1$  and  $c_2$  are the learning factors,  $r_1$ ,  $r_2$  are the random numbers from range (0, 1),  $\mathbf{x}_B$  is the position vector of the swarm leader and  $\mathbf{x}_L$  is the vector of the best location of the  $i$ -th particle in the previous iteration.

Finally, the position of the  $i$ -th particle is determined as follows:

$$\mathbf{x}_k^i = \mathbf{x}_{k-1}^i + \mathbf{v}_k^i. \quad (8)$$

## 2.3. The Bat Algorithm

The Bat Algorithm (BA) was developed in 2010 [34]. The mathematical model of the BA is based on the echolocation navigation of small species of bats. Each bat is characterized by a velocity vector ( $\mathbf{v}^i$ ), a position vector ( $\mathbf{x}^i$ ), varying frequency ( $F^i$ ), loudness ( $A^i$ ) and pulse emission ( $r^i$ ). A group of bats constitutes a bat colony. The search of the global extreme is carried out by randomly searching the permissible area. The individual with the best objective function value is the colony leader. The position of the leader is updated in every  $k$ -th iteration of the algorithm. The position vector of  $i$ -th bat is calculated using: the position of the best bat in the colony, the velocity vector of  $i$ -th bat from the previous iteration and the random value of frequency. The position of each bat is determined as follows:

$$\mathbf{x}_k^i = \mathbf{x}_{k-1}^i + \left[ \mathbf{v}_{k-1}^i + F^i (\mathbf{x}_{k-1}^i - \mathbf{x}_B) \right], \quad (9)$$

$$F^i = F_{\max} + r(F_{\max} - F_{\min}), \quad (10)$$

where:  $r$  is a randomly selected number from the range (0, 1);  $F_{\max}$  and  $F_{\min}$  are the maximum and minimum frequency values.

In each step of the algorithm, a test modification of the leader/random individual is performed, this represents the local search capability. The test position  $\mathbf{x}^*$  near the leader/random bat is determined as follows:

$$(\mathbf{x}^*)_k^i = \mathbf{x}_k^i + \kappa A_{av} \quad (12)$$

where:  $\kappa$  is a randomly selected number from the range (0, 1),  $A_{av}$  is the value of the average loudness of a bat colony in the  $j$ -th time step.

If the test position  $(\mathbf{x}^*)_k^i$  has a better objective function value than  $\mathbf{x}_k^i$ , then  $\mathbf{x}_k^i$  is updated by  $\mathbf{x}^*$ . The loudness  $A^i$  and rate of pulse emission  $r^i$  of a bat in the next iteration are calculated as follows:

$$A_{k+1}^i = \zeta A_k^i, \quad r_{k+1}^i = r_0[1 - \exp(-\gamma j)] \quad (13)$$

in which:  $\zeta$  and  $\gamma$  are the bat algorithm constants,  $r_0$  is the initial value of emission rate.

### 3. Constrained optimization of PM synchronous motor

The GWO algorithm is very often applied for the purpose of solving global optimization problem. However, there are not too many articles in international literature regarding the application of this method to solve constrained optimization tasks [13, 17]. Moreover, in the case of optimization of PM motors the simplified model of phenomena is usually applied.

In order to validate the effectiveness and performance of the GWO algorithm in combination with the external penalty function, constrained optimization of the LSPMSM has been performed. The in-house optimization software consists of two independent modules: (a) the optimization module and (b) the numerical model of PM motor. The optimization module comprises the gray wolf algorithm and was developed using Delphi 7.0. The non-linear constraint was taken into account through the external penalty function [35]. The numerical model of the LSPMSM was developed in the ANSYS Maxwell environment using 2D finite element method (FEM). This model consist two independent 2-D FEM transient models: (a) a steady-state operation model at synchronous speed and (b) a model describing start-up of motor. The efficiency and power factor under the rated load condition was calculated in the steady-state operation model. The model which describes the start-up allows for the calculation of the value of electromagnetic torque during the start-up of the motor. A stator from a serially manufactured six-poles induction motor, type

Table 1. Rated parameters of the MS2 132M-6 type motor

Parameter	Symbol	Value	Unit
Rated output	$P_N$	5,5	kW
Rated voltage	$V_N$	400	V
Rated current	$I_N$	12	A
Rated speed	$n_N$	960	rpm
Efficiency	$\eta$	86	%
Power factor	$\cos\varphi$	0,78	-
Rated torque	$T_N$	54,7	Nm
Peak torque	$T_{\max}$	147,1	Nm

MS2 132M-6 was applied. The rated parameters of the MS2 132M-6 motor are listed in Table 1.

The stator winding has a three-phase, a double-layer which overlaps the whole-coiled winding. The winding is wye-connected. The main dimensions of the stator and the rotor of the induction motor are listed in Table 2.

Table 2. Main stator and rotor dimensions of the MS2 132M-6 type motor

Parameter	Symbol	Value	Unit
Outer stator diameter	$D_o$	240	mm
Inner stator diameter	$D_i$	165	mm
Length of stator core	$L_s$	115	mm
Number of slot in stator	$N_s$	54	-
Outer rotor diameter	$d_o$	164,2	mm
Shaft diameter	$d_i$	45	mm
Number of slots in rotor	$N_r$	51	-

The optimization task consists of such a selection of the structure of the rotor, which include cage winding made of aluminium and permanent magnets. Only the dimension of the permanent magnet and its location in the rotor structure were taken into account. Three design parameters were adopted:  $s_1 = l_m$  – magnet length,  $s_2 = g_m$  – magnet thickness and  $s_3 = r_m$  – distance between poles (see Figure 3). The ranges of the design variables are presented in Table 3. During study, the all stator dimensions and winding parameters represented constants.

Table 3. The range of the design variables

Variable	Structure parameter	Down	Up
$s_1$	$l_m$	1 mm	16 mm
$s_2$	$g_m$	2 mm	12 mm
$s_3$	$r_m$	5 mm	67 mm

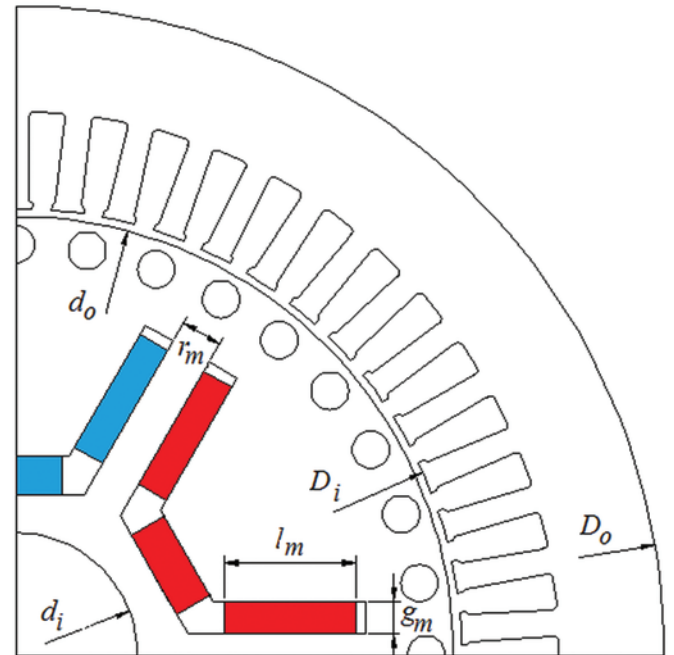


Fig. 3. The structure of discussed LSPMSM

The air gap in the designed LSPMSM motor was increased to 0,8 mm in comparison to the MS2 132M-6 induction motor. The air gap

extension was obtained by reducing the outer diameter of the rotor  $d_o$ . The permanent material NdFeB N33 with nominal magnetic properties:  $B_r = 1,15$  T and  $H_c = 836$  kA/m was applied.

An optimal motor should be characterized by high values of three maintenance (functional) parameters: (a) efficiency [18], (b) power factor and (c) synchronization capability. The synchronization capability is determined by value of the electromagnetic torque for a speed equal 0,8 of the synchronous speed. In the case of induction motors, a speed of about 0,8 of synchronous speed is the highest value of electromagnetic torque, i.e. the peak torque. During the LSPMSM start-up process we can observe two types of torque [39]. The first one is the asynchronous torque generated by squirrel cage winding, where slips differ from 0. The second one is the synchronous torque, which includes (a) the opposite braking torque generated by the permanent magnet and (b) the reluctance torque [23]. The braking torque depends on the dimensions of the permanent magnets. The reluctance torque is generated by the diversity of the rotor structure in which gaps for permanent magnets are present. The braking torque degrades the line-start performance. The total electromagnetic torque generated by a machine during the synchronization period is the sum of the synchronous and the asynchronous torques. Figure 4 illustrates the waveform of the total torque, the sum of the asynchronous and the reluctance torque and the opposite torque generated by the PM during the start-up of the LSPMSM. The presented torque-slip curves were calculated for a machine obtained in the start population. It can be observed, that magnets with excessively large dimensions were used in the studied LSPMSM. The motor was characterized by a high value of opposite torque generated by the PM. The high value of the opposite torque significantly weakens the value of the asynchronous torque. As a result of this, the motor can have problems when going into the synchronous state. The authors of papers concerning the optimization of the LSPMSM very frequently use the  $T_{80}$  value to guarantee a proper synchronization process [11, 12].

In the case of multi-criteria optimization problems, two types of compromise objective functions can be applied: (a) multiplicative and (b) additive [28]. After performing many computational simulations for the three iterations of the GWO algorithm, it was observed that the multiplicative function is more sensitive to changes in functional parameters than the additive function. During the optimization of the LSPMSM the most important parameter is the synchronization capability; in the case of multiplicative function the obtained values of this parameter were better. Thus, in the developed algorithm, the objective function for  $j$ -th wolf has been defined as follows:

$$f^j(\mathbf{s}) = \left( \frac{\eta^j(\mathbf{s})}{\eta_0} \right)^{q_1} \left( \frac{\cos\varphi^j(\mathbf{s})}{\cos\varphi_0} \right)^{q_2} \left( \frac{T_{80}^j(\mathbf{s})}{T_0} \right)^{q_3}, \quad (13)$$

where  $\mathbf{s}=[s_1, s_2, s_3]^T$  is the vector made up of design parameters,  $\cos\varphi^j(\mathbf{s})$ ,  $\eta^j(\mathbf{s})$  and  $T_{80}^j(\mathbf{s})$  represent the power factor, the efficiency and the out put torque at a speed equal to 0,8 of the synchronous speed,  $\eta_0$ ,  $\cos\varphi_0$  and  $T_0$  are the values of these 3 parameters calculated as a mean value of such parameters from 15 runs of optimization procedure for random distribution of the start wolf pack before the start of the optimization process,  $q^1$ ,  $q^2$  and  $q^3$  are the weighting coefficients.

Moreover, an economic factor has been taken into account during the optimization process and the consumption of the permanent magnet material was included into the design process. The non-linear constraint concerning the total mass of permanent magnet material was defined as  $m_{pm}^j(\mathbf{s}) \leq m_{pm}^i$ , where  $m_{pm}^i$  is the imposed total mass of permanent magnet material in the designed motor. The non-linear constraint function has been normalized and calculated as:

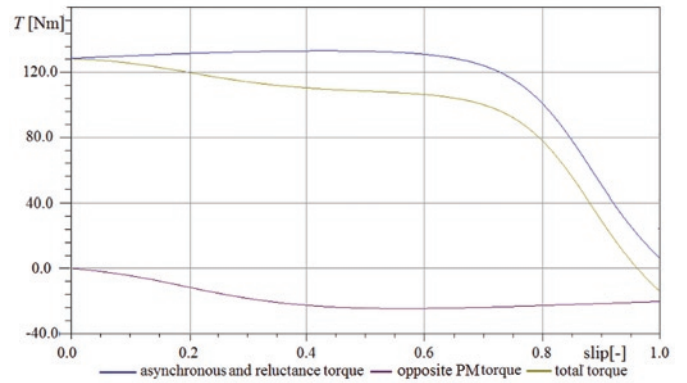


Fig. 4. The torque components during start-up process of LSPMSM

$$g^j(\mathbf{s}) = \frac{m_{pm}^j(\mathbf{s}) - m_{pm}^i}{m_{pm}^i}. \quad (14)$$

The imposed constraint was included in the optimization process by using the external penalty function [35]. In the penalty function approach, the modified objective function  $h(\mathbf{s})$  is created. The  $h(\mathbf{s})$  function is composed of: (a) objective function  $f(\mathbf{s})$  and (b) penalty term  $p(\mathbf{s})$ . The penalty term  $p(\mathbf{s})$  for the  $j$ -th wolf is calculated as follows:

$$p^j(\mathbf{s}) = \sigma^n g^j(\mathbf{s}). \quad (15)$$

In the developed algorithm, the iterations related to a change in external penalty ( $n$ ) are intertwined with iterations of the GWO algorithm ( $k$ ). The value of the penalty coefficient is modified after performing three iterations of the GWO algorithm. In each subsequent iteration of the penalty ( $n$ ), the penalty coefficient value is increased. After performing the maximum number of penalty iterations ( $n_{max}$ ), the optimization process is finished.

In the developed algorithm the objective function  $f(\mathbf{s})$  is maximized. Therefore, the modified objective function  $h(\mathbf{s})$  is defined as follows:

$$h^j(\mathbf{s}) = \begin{cases} f^j(\mathbf{s}) & \text{for } m_{pm}(\mathbf{s}) \leq m_{pm}^i \\ f^j(\mathbf{s}) - p^j(\mathbf{s}) & \text{for } m_{pm}(\mathbf{s}) > m_{pm}^i \end{cases}. \quad (16)$$

#### 4. Simulation results for different optimization algorithms

In order to evaluate the convergence and reliability of the developed optimization procedure (containing the gray wolf algorithm), test calculations which consisted of the constrained optimization of a LSPMSM motor were performed. All the studied optimization procedures were developed by the author in the Borland Delphi 7.0 environment. The optimization procedures containing each method (GWO, PSO, BA and GA) were repeated 12 times. The best results obtained from all runs of the optimization software for the GWO algorithm were compared with those of the PSO, the BA and the GA in Tables 4, 5, 6, 7 and 8. Next, statistical analysis of the results was performed for all of the optimization algorithms used.

##### 4.1. Calculations for the GWO method

The optimization calculation was performed for a wolf pack size equal to 32 individuals. Due to the application of the FEM model of the LSPMSM, such a number of individuals provided a compromise between good convergence of the optimization procedure and the calculation time. The following parameters of the optimization procedure

Table 4. The course of the constrained optimization of LSPMSM for  $\alpha$  wolf

k	$r_m$	$g_m$	$l_m$	$\eta(s)$	$\cos\varphi(s)$	$T_{80}(s)$	$m_{pm}(s)$	$h(s)$	$N_{of}$
	mm	mm	mm	%	-	Nm	kg	-	-
1	7,16	6,85	43,38	91,01	0,808	96,845	1,518	2,786569	61
2	6,98	3,28	56,57	90,86	0,807	120,661	0,948	3,358127	88
3	8,09	4,90	64,42	91,81	0,940	102,498	1,612	3,370711	119
5	6,97	4,61	65,10	91,76	0,933	105,709	1,529	3,488402	177
7	6,05	4,45	64,98	91,73	0,927	106,567	1,477	3,505084	235
9	8,08	4,54	64,89	91,76	0,931	106,115	1,508	3,507330	293
10	8,49	4,53	64,81	91,76	0,931	106,163	1,503	3,508139	322
12	9,58	4,50	64,79	91,75	0,930	106,219	1,498	3,509258	380
15	9,49	4,49	64,79	91,75	0,930	106,500	1,499	3,516157	467
20	9,49	4,49	64,79	91,75	0,930	106,500	1,499	3,516157	612
30	9,49	4,49	64,79	91,75	0,930	106,500	1,499	3,516157	902

Table 5. The course of the constrained optimization of LSPMSM for  $\beta$  wolf

k	$r_m$	$g_m$	$l_m$	$\eta(s)$	$\cos\varphi(s)$	$T_{80}(s)$	$m_{pm}(s)$	$h(s)$
	mm	mm	mm	%	-	Nm	kg	-
1	9,18	6,32	36,08	90,214	0,739	107,329	1,165	2,767361
2	11,38	5,36	52,66	91,405	0,860	103,287	1,442	3,158777
3	11,98	4,84	54,38	91,382	0,860	107,259	1,344	3,257946
5	7,86	4,82	64,62	91,812	0,941	103,193	1,598	3,403105
7	11,29	4,52	65,03	91,764	0,932	105,577	1,500	3,497235
9	9,58	4,50	64,97	91,752	0,930	106,219	1,495	3,502258
10	10,13	4,51	65,84	91,759	0,931	105,960	1,498	3,504899
13	8,01	4,53	64,97	91,757	0,931	106,246	1,503	3,509144
15	9,63	4,52	64,92	91,758	0,931	106,138	1,499	3,509673
30	9,37	4,50	64,81	91,753	0,930	106,327	1,499	3,511512

cedure were assumed:  $q_1=1, q_2=4/5, q_3=1/3, m_{pm}^i=1,5$  kg and  $\sigma=1,2$ . The penalty coefficient was increased in every third internal iterations of the GWO method. As a stop criteria, a maximum number of external iterations  $n_{max}=10$  was assumed, i.e. 30 iterations of the GWO method. The following values of reference parameters were adopted:  $\eta_0=81,325$  %,  $\cos\varphi_0=0,753$  and  $T_0=50,124$  Nm. Such values was the same for all investigated optimization procedures.

The results of the optimization calculation for the selected GWO iterations are presented in Tables 4 and 5. The results for the  $\alpha$  individual are reported in Table 4, while the results for the  $\beta$  individuals are listed in Table 5. In the successive columns, the design parameters values ( $r_m, g_m, l_m$ ), the functional parameters ( $\eta, \cos\varphi$  and  $T_{80}$ ) of the designed motor, the modified objective function value ( $h$ ), and the number of calls function ( $N_{of}$ ), i.e. number of calculations of the objective function are listed.

The calculations were made on a computer with the following parameters: processor: Ryzen 5 six-core, 3,40 GHz and 16,0 GB RAM. The calculation time for one individual is equal to 8 minutes and 6 seconds. The calculation time depends on the saturation, especially at the initiation of the wolf pack. The approximate calculation time for the optimization process was about 116 hours for a single computer. The developed optimization procedure was tested and a good convergence was achieved. It should be noted that the developed algorithm determined the optimal solution after about 12 iterations, i.e. after four iterations related to the increasing penalty, and the imposed total mass of permanent material was attained. In the suc-

cessive iteration of optimization process, the maximized functional parameters were improved.

#### 4.2. Calculations for the PSO algorithm

Next, the calculations for the classical PSO algorithm were performed. The number of particles was  $N=32$ . The following values of the PSO coefficients were assumed:  $w=0,2, c_1=0,35$  and  $c_2=0,45$ . The values of the PSO coefficients were assumed on the basis of many computer simulations to provide good convergence of the optimization procedure for the first three iterations. The PSO algorithm was combined with an external penalty. The parameters of the optimization procedure are the same as the parameters of the GWO algorithm. The course of the optimization process for the selected iterations is presented in Table 6.

The total approximate calculation time in case of PSO optimization procedure is 132 hours. The computation time for the PSO method is slightly longer than for the GWO method. Moreover, the obtained result is lower quality in comparison to the GWO algorithm.

#### 4.3. Calculations for the BA algorithm

Thus, the constrained optimization process was executed for the BA algorithm. The number of bats in a colony was  $N=32$  and the maximum external penalty iteration is equal to  $n_{max}=10$ . The adopted values of the BA parameters were based on the author's previous experience. The following values have were adopted: range of frequency  $F_{min}=0, F_{max}=1,0$ , initial pulse emission value  $r_0=0$ , and initial



Table 6. The course of the constrained optimization of LSPMSM for PSO algorithm

k	$r_m$	$g_m$	$l_m$	$\eta(s)$	$\cos\varphi(s)$	$T_{80}(s)$	$m_{pm}(s)$	$h(s)$	$N_{of}$
	mm	mm	mm	%	-	Nm	kg	-	-
1	9,68	8,20	40,69	90,081	0,692	107,672	1,7042	2,594397	64
2	9,68	7,07	42,21	90,950	0,801	95,928	1,5242	2,750185	96
3	10,86	6,93	61,41	92,007	0,990	74,609	2,3010	2,806432	128
5	8,33	4,56	63,87	91,764	0,932	105,855	1,5142	3,507002	192
10	9,16	4,52	64,08	91,757	0,931	106,151	1,4991	3,509267	352
13	9,16	4,52	64,08	91,757	0,931	106,151	1,4991	3,509267	448
14	8,90	4,52	64,65	91,757	0,931	106,176	1,5001	3,510162	480
20	9,04	4,52	64,65	91,756	0,930	106,193	1,4996	3,510163	672
23	9,35	4,50	64,66	91,754	0,930	106,283	1,4994	3,510606	768
25	9,44	4,51	64,72	91,755	0,930	106,255	1,4998	3,510748	832
30	9,50	4,49	64,82	91,752	0,930	106,333	1,4998	3,510911	992

Table 7. The course of the constrained optimization of LSPMSM for BA algorithm

k	$r_m$	$g_m$	$l_m$	$\eta(s)$	$\cos\varphi(s)$	$T_{80}(s)$	$m_{pm}(s)$	$h(s)$	$N_{of}$
	mm	mm	mm	%	-	Nm	kg	-	-
1	7,79	8,90	47,53	91,824	0,906	71,773	2,160	2,218321	64
3	9,78	8,27	41,15	90,141	0,697	105,722	1,738	2,573938	128
5	9,78	8,27	41,15	90,141	0,697	105,722	1,738	2,573938	192
7	6,20	6,79	53,69	91,685	0,899	89,486	1,861	2,942079	256
10	6,20	6,79	53,69	91,685	0,899	89,486	1,861	2,942079	352
14	4,97	5,68	62,06	91,861	0,945	95,951	1,801	3,281955	480
18	9,43	4,47	60,65	91,588	0,896	108,529	1,385	3,439104	576
20	7,68	4,59	62,71	91,692	0,916	106,836	1,468	3,471697	672
22	6,34	4,50	64,13	91,746	0,929	106,929	1,495	3,506710	736
24	3,49	4,50	64,52	91,737	0,928	107,548	1,495	3,514491	800
28	3,31	4,52	64,78	91,741	0,928	107,425	1,500	3,516855	928
30	3,45	4,52	64,78	91,741	0,929	107,386	1,500	3,517128	992

loudness value  $A_0=1$ ,  $\zeta=0,7$  and  $\gamma=0,6$ . The results of the optimization process for the selected time steps are presented in Table 7. In the successive columns, the number of time steps, the coordinates of the best bat (individual), the modified objective function for the leader of a bat colony and the number of call function are listed.

In the case of the BA algorithm, the obtained approximate operating time of the optimization procedure is similar to the PSO and equal 132 hours.

#### 4.4. Calculations for the GA algorithm

Also, the calculations for GA procedure for parameters: population size equal  $N=32$ , probability of mutation  $p_m=0,007$  were executed. The optimization procedure consists of three genetic operators: reproduction, crossover and mutation [27]. Additionally, the simple elitism procedure has been applied to save the best individual during genetic operations, especially mutation procedure. The roulette wheel reproduction and one cut-point in chromosome crossover methods have been applied. Also, improved crossover procedure was applied. The crossover procedure consists of two phases. First, all individuals are crossover. The new generation was created from the best half of parents and the best half of children [16]. Such a crossover operation significantly improves the convergence of an elaborated optimization procedure based on genetic algorithm. The GA optimization procedure is adapted to perform optimization calculation by the use of the FEM models. In each genetic generation, the values of objective func-

tion are calculated only for new individuals created in crossover and mutation procedures. The course of the optimization process for the selected genetic iterations is presented in Table 8.

The total calculation time for GA was the longest from all the tested algorithms and was equal 195 h. However, the obtained result was the best.

#### 4.5. Statistical analysis of all the tested methods

On the basis of the obtained results it can be noted, that the investigated optimization algorithms (GWO, PSO, BA and GA) correctly found the global maximum. Similar values of design variables were determined for the GWO and the PSO algorithms. After about 12 iterations, i.e. 380 calls of objective function, the gray wolf algorithm determined a solution which was close to optimal. The value of the modified objective function for the  $\alpha$  individual improved in subsequent iterations (see Table 4). In comparison with the PSO algorithm, the  $\beta$  individual has much more accurate results than the PSO leader. The PSO method provided the least accurate solution from among all the tested methods. It can be observed that for the PSO algorithm, the modified objective function value was improved until the last iteration of the algorithm. The highest value of the objective function was obtained for GA.

The convergence process for all the studied algorithms is presented in Figure 5. The fastest convergence at the beginning of the optimiza-

Table 8. The course of the constrained optimization of LSPMSM for GA

k	$r_m$	$g_m$	$l_m$	$\eta(s)$	$\cos\varphi(s)$	$T_{80}(s)$	$m_{pm}(s)$	$h(s)$	$N_{of}$
	mm	Mm	mm	%	-	Nm	kg	-	-
1	11,99	8,68	40,53	91,283	0,830	75,827	1,795	2,353208	127
2	11,97	7,44	65,00	92,022	0,981	58,508	2,469	2,365551	173
3	6,87	5,12	58,24	91,629	0,899	103,787	1,524	3,321283	217
5	4,98	4,18	57,03	91,299	0,854	113,598	1,217	3,392834	296
7	3,71	4,49	63,38	91,681	0,916	108,336	1,454	3,509650	384
9	9,03	4,52	65,00	91,756	0,931	106,181	1,499	3,509902	501
11	8,07	4,52	64,88	91,754	0,930	106,395	1,500	3,514408	568
13	7,59	4,52	64,88	91,752	0,930	106,512	1,499	3,516853	659
16	7,69	4,51	64,92	91,750	0,930	106,600	1,499	3,517968	792
20	7,25	4,52	64,87	91,751	0,930	106,605	1,499	3,518801	973
30	7,25	4,52	64,88	91,751	0,930	106,605	1,499	3,518801	1463

tion process was provided by the GWO algorithm. The BA algorithm was characterized by the worst convergence.

The process of searching for the global maximum in the case of the BA algorithm is very interesting. On the basis of the data presented in Figure 5, it can be observed that every few iterations, there is a change in the position of the best adapted bat. In the BA method the individuals move randomly around the best bat. The result, which is close to the optimal one is obtained after 24 time steps of the BA algorithms, i.e. after 800 calls of the objective function. This is double the number of the calls function than in the case of the GWO.

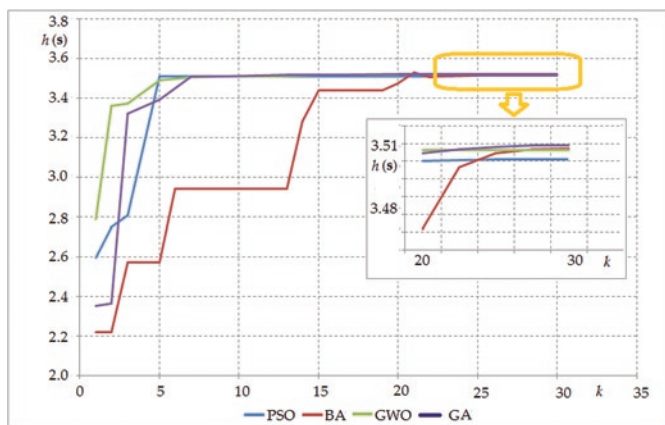


Fig. 5. Convergence curves for investigated methods

Subsequently, the changes of the design variables during the optimization process were analyzed. Figure 6 shows the variation of all design variables in successive iterations of the optimization process.

It can be observed, that in the case of the distance between the poles ( $r_m$ ), the design variables changed during the optimization process. As a result of the optimization process for the PSO and the GWO algorithms, similar “optimal” values were received (see Figure 6a). However, for the BA and GA a different “optimal” value of the ( $r_m$ ) variable than for the other tested algorithms were obtained. In the case of the magnet length ( $l_m$ ), as a result of the optimization process, a similar PM magnet length value was obtained for all the examined optimization algorithms (see Figure 6). It was also observed, that the BA convergence was the worst when compared with the other algorithms. For the magnet thickness ( $g_m$ ) for all the optimization algorithms, similar of the optimal value of this design parameter were obtained.

The results were obtained during 12 runs of the optimization software for different optimization procedures. For all the optimization algorithms, the best, and the worst solution, as well as the average and

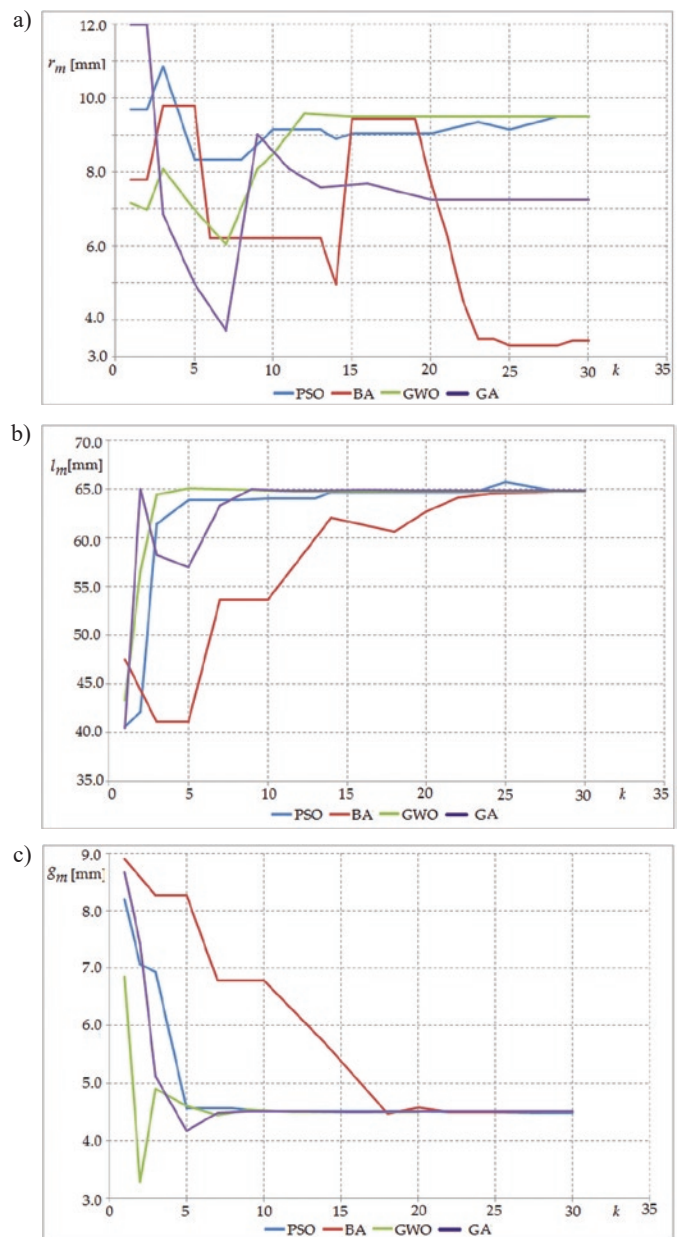


Fig. 6. Variations of design variables in successive iteration of the optimization process: (a)  $r_m$ , (b)  $l_m$  and (c)  $g_m$ .

Table 9. Statistical data for 12 runs of optimization software

Algorithm	Best	Worst	Mean	Standard deviation	Approximate calculation time
GWO	3,516157	3,490543	3,501622	0,009157	116 h
PSO	3,510914	3,456814	3,490405	0,017013	132 h
BA	3,517136	3,426121	3,488883	0,025126	132 h
GA	3,518801	3,476884	3,501936	0,015807	195 h

standard deviation were determined. The obtained results are shown in Table 9.

The best value of the objective function was obtained for the BA. However, the other best results (mean value, worst objective function value and standard deviation) were attained in the GWO algorithm. The PSO algorithm achieved the worst optimal value of the modified objective function and the worst mean value among the investigated algorithms. While, the worst value of the standard deviation for 12 runs of the optimization procedure was obtained for the BA.

## 5. Conclusions

In the article the optimization procedure for the constrained optimization problem was developed. The optimization procedure is a combination of the gray wolf algorithm and the external penalty function. The procedure was applied to optimization of the LSPMSM. The designed motor was described by the FEM model. The multiplicative compromise objective function composed of three parameters of the motor has been used. Additionally, the non-linear constraint function concerning the mass of the permanent magnet material was taken into consideration.

The performance reliability of the developed optimization procedure containing the GWO algorithm was compared with other nature-inspired optimization algorithms (PSO and BA). All the investigated optimization procedures were developed in Delhi 7.0 environment. In the GWO, the results close to the optimal one can be reached almost two times faster in comparison to the BA. Of course the BA enables to obtain more accurate solution. In the computational experiment, the worst value of objective function was obtained for the PSO method. Furthermore, the GWO algorithm allows obtaining the best standard

deviation and average values of objective function in the constrained optimization of the LSPMSM.

A comparison of the quality of results between alpha individual and alternative leader ( $\beta$  individual) was presented. A much precise optimal solution was obtained for the  $\alpha$  wolf than for the  $\beta$  wolf. The beta individual has a higher value of the objective function than the leader for the PSO algorithm, which underlines the advantage of the GWO algorithm over the PSO.

In the GWO algorithm, the new positions of the omega individuals in subsequent iterations are determined on the basis of the position of the leader and two alternative leaders ( $\beta$  and  $\delta$  individuals). This feature causes high-efficiency in comparison to other investigated methods. For the majority of nature-inspired method, in order to determine the new position of individuals only the position of the leader is taken into account. Including only the leader position may lead to a movement of the swarm/colony towards the local extreme, which causes a disturbance in the optimization process. Therefore, taking into account several leaders with different rungs can improve the performance of the optimization algorithm. The presented results are very encouraging and show clearly that the GWO method is a very interesting optimization tool.

The developed optimization procedure can be successfully applied to an optimization of different electromagnetic devices described by models of varying complexity.

In the future research, the author will build a prototype of the PM motor and will perform an experimental verification.

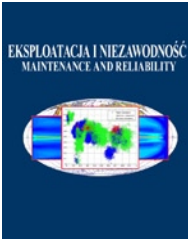
## Acknowledgements

The presented research is supported by the project under agreement number 0212/SBAD/0521.

## References

1. Amoiralis E I, Tsili M A, Kladas A G. Global transformer design optimization using deterministic and non-deterministic algorithms, Proceedings of the International Conference on Electrical Machines, Marssille, France, 2-5 September 2012, <https://doi.org/10.1109/ICEIMach.2012.6350207>.
2. Bacchus A, Tounzi A, Arguad J. P, Bouriquet B, Biet L, Macaire L, Le Menach Y. Estimation of FEM model parameters using data assimilation and its application to an electrical machine. IEEE Transactions on Magnetics 2016; 52(3), <https://doi.org/10.1109/TMAG.2015.2495265>.
3. Barański M, Szelał W, Lyskawiński W. An analysis of a start-up process in LSPMSMs with aluminum and copper rotor bars considering the coupling of electromagnetic and thermal phenomena, Archives of Electrical Engineering 2019; 68(4): 933 - 946.
4. Bora T. C, Coelho L, Lebensztajn L. Bat-inspired optimization approach for brushless DC motor problem. IEEE Transactions on Magnetics 2012; 48(2): 947 - 950, <https://doi.org/10.1109/TMAG.2011.2176108>.
5. Cvetkovski G, Petkovska L, Lefly P. Cuckoo search as a tool for optimal design of PM brushless DC motor. COMPEL 2018; 37(5): 1732 - 1743, <https://doi.org/10.1108/COMPEL-01-2018-0025>
6. Deng S, Brisset S, Clenet S. Comparative study of method for optimization of electromagnetic devices with uncertainty. COMPEL 2018; 37(2): 704 - 717, <https://doi.org/10.1108/COMPEL-11-2016-0502>.
7. Gao J, Liu J, Huang S, Huang K. Optimum Design of Permanent Magnet Synchronous Motor Based on Gene Handling Genetic Algorithms, Proceedings of the International Conference on Electrical and Control Engineering, Wuhan, China, 25-27 June 2010, <https://doi.org/10.1109/iCECE.2010.358>.
8. Hossain Mohammadi M, Silva R, Lowther D. A. Finding optimal performance indices of synchronous AC motors. IEEE Transactions on Magnetics 2017; 53(6), <https://doi.org/10.1109/TMAG.2017.2662705>.
9. Idziak P, Kowalski K, Nowak L, Knypiński Ł. FE transient analysis of the magnetostrictive actuator. International Journal of Applied Electromagnetics and Mechanics 2016; 51 (S1): 81 - 87, <https://doi.org/10.3233/JAE-2011>.
10. Ilea D, Radulescu M. M, Gillon F, Brochet P. Particle-swarm optimized design of switched reluctance motors for light electric traction applications. Electromotion Journal 2010; 17(3): 23 - 29, <https://doi.org/10.1109/VPPC.2010.5729005>.
11. Jedryczka C, Knypiński Ł, Demenko A, Sykulski J. Methodology for cage shape optimization of a permanent magnet synchronous motor under line start condition. IEEE Transactions on Magnetics 2018; 54 (3), <https://doi.org/10.1109/TMAG.2017.2764680>.

12. Jedryczka C, Wojciechowski R, Demenko, A. Influence of squirrel cage geometry on the synchronization of the line start permanent magnet synchronous motor. *IET Science, Measurement & Technology* 2015; 9( 2): 197 - 203, <https://doi.org/10.1049/iet-smt.2014.0198>.
13. Joshi H, Arora S. Enhanced grey wolf optimisation algorithm for constrained optimisation problems. *International Journal of Swarm Intelligence* 2017; 3(2/3): 126 - 152, <https://doi.org/10.1504/IJSI.2017.087871>.
14. Karnavas Y L, Chasiotis I D, Peponakis E L. Permanent magnet synchronous motor design using grey wolf optimizer algorithm. *International Journal of Electrical and Computer Engineering* 2016; 6(3): 1353 - 1362, <https://doi.org/10.11591/ijece.v6i3.9771>.
15. Knypiński Ł. Adaptation of the penalty function method to genetic algorithm in electromagnetic devices designing. *COMPEL* 2019; 38 (4): 1285-1294, <https://doi.org/10.1108/COMPEL-01-2019-0010>.
16. Knypiński Ł, Nowak L. The algorithm of multi-objective optimization of PM synchronous motors. *Przegląd Elektrotechniczny* 2019; 95 (4): 242-245, <https://doi.org/10.15199/48.2019.04.46>.
17. Kohli M, Arora S. Chaotic grey wolf optimization algorithm for constrained optimization problems. *Journal of Computational Design and Engineering* 2018; 5: 458 - 472, <https://doi.org/10.1016/j.jcde.2017.02.005>.
18. Kosucki A, Stawiński Ł, Malenta P, Zaczyński J, Skowrońska J. Energy consumption and energy efficiency improvements of overhead crane's mechanism. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2020; 22(2): 232 - 330, <https://doi.org/10.17531/ein.2020.2.15>.
19. Lei G, Zhu J, Gou Y, Liu C, Ma B. A review of design optimization method for electrical machines. *Energies* 2017; 1962(10), <https://doi.org/10.3390/en10121962>.
20. Lewis M, Murray J. D. Modeling territoriality and wolf-deer interactions. *Nature* 1993; 366: 738 -740, <https://doi.org/10.1038/366738a0>.
21. Li Y, Yang R, Zhao X. Reactive power convex optimization of active distribution network based on improved grey wolf optimizer. *Archives of Electrical Engineering* 2020; 61,(1): 117 - 131.
22. Madadi A, Motlagh Mosheni A. Optimal control of DC motor using Grey Wolf optimizer algorithm. *Technical Journal of Electrical Engineering and Applied Science* 2014; 4(4): 373 - 379.
23. Marcic T. A short review of energy-efficient line-start motor design. *Przegląd Elektrotechniczny* 2011; 87 (3): 119 - 122.
24. Meach D L. Alpha statue dominance, and divisions of labor in wolf packs. *Canadian Journal of Zoology* 1999; 77: 1196 -1203, <https://doi.org/10.1139/z99-099>.
25. Mirjalili S, Mirjalili S. M, Lewis A. Grey wolf optimizer. *Advances in Engineering Software* 2014; 69: 46 - 61, <https://doi.org/10.1016/j.advengsoft.2013.12.007>.
26. Moussouni F, Brisset S, Brochet P. Comparison of two multi-agents ACO and PSO for optimization of a brushless DC wheel motor. *Intelligent Computer Techniques in Applied Electromagnetic Systems* 2008: 3 - 10, [https://doi.org/10.1007/978-3-540-78490-6\\_1](https://doi.org/10.1007/978-3-540-78490-6_1).
27. Mutluer M, Sahman A, Cunkas M. Heuristic optimization based on penalty approach for surface permanent magnet synchronous machines. *Arabian Journal for Science and Engineering* 2020; 45: 6751-6767, <https://doi.org/10.1007/s13369-020-04689-y>.
28. Nowak L, Knypiński Ł, Jedryczka C, Kowalski K. Decomposition of the compromise objective function in the permanent magnet synchronous motor optimization. *COMPEL* 2015; 34(2): 496 - 504, <https://doi.org/10.1108/COMPEL-07-2014-0173>.
29. Sarac V. Performance optimization of permanent magnet synchronous motor by cogging torque reduction, *Journal of Electrical Engineering* 2019; 70(3): 218 - 226, <https://doi.org/10.2478/jee-2019-0030>.
30. Shaari G, Tekbiyik-Ersoy N, Dagbasi M. The state of art in particle swarm optimization based unit commitment: a review. *Processes* 2019; 7(733), <https://doi.org/10.3390/pr7100733>.
31. Sorgdrager A J, Wang R, Grobler A J. Multiobjective design of line-start PM motors using the taguchi method. *IEEE Transactions on Industrial Applications* 2018; 54(5), <https://doi.org/10.1109/TIA.2018.2834306>.
32. Sun K, Wang G, Lu Y. Optimization method of bevel gear reliability based on genetic algorithm and discrete element. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2019; 21(3): 186 - 196, <https://doi.org/10.17531/ein.2019.2.2>.
33. Wang Y, Filippini M, Bacco G, Bianchi N. Parametric design and optimization of magnetic gears with differential evolution method. *IEEE Transactions on Industrial Applications* 2019; 55(4): 3445-3452, <https://doi.org/10.1109/TIA.2019.2901774>.
34. Yang X S, Gandomi A H. Bat algorithm: a novel approach for global engineering optimization. *Engineering Computations* 2012; 29(5): 464 - 483, <https://doi.org/10.1108/02644401211235834>.
35. Yeniay Ö. Function methods for constrained optimization with genetic algorithms. *Mathematical and Computational Application* 2005; 10(1): 45 - 56, <https://doi.org/10.3390/mca10010045>.
36. Zang H, Zang S, Hapeshi K. A review of nature-inspired algorithm. *Journal of Bionic Engineering* 2010; 7: 232 - 237, [https://doi.org/10.1016/S1672-6529\(09\)60240-7](https://doi.org/10.1016/S1672-6529(09)60240-7).
37. Zio E, Mellal M. System reliability-redundancy optimization with cold-standby strategy by an enhanced nest cuckoo optimization algorithm. *Reliability Engineering & System Safety* 2020; 201, <https://doi.org/10.1016/j.ress.2020.106973>.
38. Zhao M, Wang X, Yu J, Bi J, Xiao Y, Zhan J.. Optimization of construction duration and schedule robustness based on hybrid grey wolf optimizer with sine cosine algorithm. *Energies* 2020; 13(215), <https://doi.org/10.3390/en13010215>.
39. Zöhra B, Akar M, Eker M. Design of novel line-start synchronous motor rotor. *Electronics* 2019; 8(25), <https://doi.org/10.3390/electronics8010025>.



Article citation info:

Cui X, Liang X, Bharadwaj U. A framework for corrosion assessment in metallic structures, from data analysis to risk based inspection. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 11–20, <http://dx.doi.org/10.17531/ein.2021.1.2>.

## A framework for corrosion assessment in metallic structures, from data analysis to risk based inspection

Indexed by:



Xiaofei Cui<sup>a,\*</sup>, Xiaoxia Liang<sup>a,\*</sup>, Ujjwal Bharadwaj<sup>a</sup>,

<sup>a</sup>TWI Ltd, Granta Park, Great Abington, Cambridge, CB21 6AL, United Kingdom

### Highlights

- A stepwise guideline for fully automation quantitative corrosion estimation.
- Enhanced corrosion analysis method by the statistical approach.
- Software tools developed by TWI Ltd implementing the proposed framework.
- A model widely adaptive to various metallic structure suffering from corrosion damage.

### Abstract

Metallic corrosion is a big challenge affecting many sectors in a nation's economy. Necessary corrosion prevention actions have to be taken in order to maintain the integrity of engineering assets susceptible to corrosion. This paper proposes a holistic framework to support the management of corrosion in metallic structures. It is a fully automation corrosion assessment process, with risk updated by Bayesian theory. Through analyzing the thickness data measured by non-destructive testing (NDT) techniques, the influence of corrosion on the component can be estimated using statistical methods, which will enable users to make decisions on maintenance based on quantitative information. A case study using corrosion data from a steel bridge is included to demonstrate the proposed framework. It improved the conventional corrosion analysis method by the proposed statistical approach using representative thickness data, which aims to take full use of the remaining life. This model can be adapted to a wide range of metallic structure suffering from corrosion damage.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

corrosion estimation; steel bridge; metallic structure; risk based inspection; software.

### Nomenclature

A	Surface area of the object	2D ACF	Two-dimensional auto-correlation function
Age	Service time, year	CDF	Cumulated density function
CR	Corrosion rate	COF	Consequence of Failure
CI	Confidence interval, %	EVD	Extreme value distributions
p	Probability of the quantile p of a statistical distribution	EVT	Extreme value theory
S	Small specimens that are sampled randomly from this object	FFS	Fitness for Service
time	Failure time, independent variables	GEVD	Generalized Extreme Value distribution
$t_{mm}$	Minimum measured thickness	iid	Independent and identically distributed
$t_{nom}$	Nominal thickness	MLE	Maximum likelihood estimation
T	Return period	MRL	Mean residual life
x	Random variable from the statistical distribution	MRR	Median rank regression
$Xp$	GEVD return level	$\xi$	Shape parameter of a EVD
$\mu$	Mean of a statistical distribution / population	MCMC	Markov chain Monte Carlo
$\bar{\mu}$	Location parameter of an EVD	NACE	National Association of Corrosion Engineers
$\sigma$	Standard deviation of a statistical distribution	POT	Peak-over-Threshold
$\bar{\sigma}$	Scale parameter of a EVD	POF	Probability of Failure
RUL	Remaining useful life	RBI	Risk based inspection

(\*) Corresponding author.

E-mail addresses: X. Liang - [xiaoxia.liang@twi.co.uk](mailto:xiaoxia.liang@twi.co.uk), X. Cui - [xiaofei.cui@twi.co.uk](mailto:xiaofei.cui@twi.co.uk), U. Bharadwaj - [ujjwal.bharadwaj@twi.co.uk](mailto:ujjwal.bharadwaj@twi.co.uk)

## 1. Introduction

Steel bridges play an important role in the transportation network and support the nation's economy and traffic [31]. Corrosion is one of the main causes of deterioration of steel bridges [32]. It may cause metal loss and fatigue cracks in the steel components, which would lead to the collapse of steel bridges [32]. Therefore, periodical inspection of steel bridge is essential for the long-term safety of public infrastructures.

A traditional method of inspecting a steel bridge is to send inspectors up to the bridge using scaffolding or ropes, and inspect it manually [31]. The disadvantages of this method are the inconvenient and dangerous for the inspectors, as well as the quality of the inspection, which can be subjective as it mainly depends on the inspectors' experience. Another traditional way is to lift inspectors and some measuring equipment to the bridge using a bucket truck or platform snooper [31]. Although is better than using scaffolding or ropes, there are still disadvantages exist on the safety of inspectors, and sometimes, the road may be too narrow or even be blocked, so that the inspection truck may not be able to reach to bridge.

To overcome these disadvantages, inspection robots are built, assisting inspectors to reach bridges, which are too dangerous to be accessed by a human due to high altitudes, radiation and other hazardous environments. Generally, there are two types of inspection robots, namely climbing robots [20] and aerial robots [30]. Compared with climbing robot, an aerial robot can achieve a high level of complexity for professional applications and speed with a low cost [31].

In terms of steel bridge corrosion detection, many state-of-the-art techniques were introduced, such as infrared thermography [9], giant magnetoresistance [32], computer-vision based method [16], Ultrasonic testing method [19], etc. Among these methods, the Ultrasonic testing method attracted our attention, because after processing, the thickness data, which is a widely recognized way for corrosion analysis in industry [4], can be obtained from the Ultrasonic testing (UT) data.

The paper is based on an Innovate UK (IUK) funded project – AS-SAI, using thickness data from the UT sensor installed on an aerial robot (namely, an unmanned aerial system) to carry out corrosion estimation for steel bridges. It is a fully automation corrosion assessment process, sending aerial robot to the target steel bridge, data acquisition from UT sensor, converting UT data to thickness data, and corrosion assessment. The paper focused on using the thickness data to automatically carry out quantitative corrosion estimation. The contribution includes: 1) The development of a stepwise guideline for quantitative corrosion estimation, using statistical methods for use in a framework developed for integrity assessments of structures that are susceptible to corrosion damage. 2) The conventional corrosion analysis, which estimated corrosion rate through minimum thickness data, was improved by our proposed statistical approach. 3) A case study is included to demonstrate the application of this model to support the integrity assessment of a metallic component using the corrosion data obtained from a steel bridge. A suite of software tools has been developed by TWI Ltd to perform the corrosion assessment, as an outcome from this research. In addition, the proposed model is envisaged to be adaptive to a wide range of metallic structure suffering from corrosion damage.

## 2. Background

NACE (National Association of Corrosion Engineers) International recommend that corrosion assessment can be conducted following the below procedures (derived from NACE Internal Corrosion Direct Assessment Methodology [18]), which is the traditional procedure that is followed by corrosion engineers:

- Step 1: Pre-assessment – conducting a detailed right-of-way (ROW) inspection and collecting all operating data
- Step 2: Indirect inspection – identifying factors affecting corrosion distribution

- Step 3: Direct examination - performing excavations and conducting detailed examinations to determine whether corrosion has occurred
- Step 4: Post Assessment - analyzing the data collected from the three previous steps to provide test results and recommendations necessary for corrosion protection

Nowadays, the abovementioned traditional procedure has been further developed to incorporate models from different disciplines (such as statistics) to improve its flexibility and capability. Application of statistical methods, as proposed in this paper, increases the efficiency of the Step 1 and Step 2 in identifying corrosion distribution as well as data collection, processing and management. It can also provide recommendations on corrosion management strategy, such as future inspection intervals, hotspots identification, corrosion rate estimation, etc., as part of the Step 4 outputs. This section reviews various existing models [22, 23, 25, 28, 29] that use statistical methods for corrosion assessment.

Studies on the statistical nature of corrosion and its relationship to inspection have been carried out since 1950s. The initial work using extreme values to simulate corrosion was carried out by Gumbel to estimate the condition of the pipeline with external corrosion [29]. Especially in Japan, the use of these methods for analyzing corrosion data has been referred to consistently since 1980s. A useful overall text to the statistical method of analyzing corrosion data was given by Kowaka [14]. However, such work does not comprehensively include strategies of data collection by NDT (non-destructive testing) methods, and does not generally validate the results obtained by comparing a sample with the whole population.

To address this, a guideline was prepared by TWI Ltd for the Health and Safety Executive (HSE) in 2002 to advice plant engineers and inspection personnel on methods for analyzing and extrapolating inspections for large plant items [29]. It introduced methods of statistical analysis of corrosion inspection data, including general comments on data collection with suggestions of data analysis methods using normal distribution and extreme value distribution families. At the time when [29] was published, complete data population sets rarely occurred in practice, and full scanning of a component was considered impractical. Therefore, the target of statistical treatment and procedures was to make a prediction on the basis of a limited sample information to infer the greater population behavior. The supporting theory is the belief that a statistical sample may follow the same distribution of its parent population and this may be a fundamental distribution. This can be further demonstrated using an example given by Kowaka [14] – in which laboratory SCC acceleration tests were used to predict the condition in field service. However, the difference of coupon size and duration between laboratory test and the field examination was extremely large, making any direct extrapolation of laboratory data for life prediction unreliable.

Nowadays, with the development of non-destructive testing technology, high resolution, and complete inspection coverage of the component surface for the corrosion inspection is becoming more common. This means the corrosion measurement from inspection will show the entire condition of the component surface in a large size sample. Therefore, the conventional way of data analysis needed updating to keep up with the development of inspection technology. A method for bridging such technology gap was developed by Shibata [25] based on statistical theory of extreme values. To apply this method, Shibata [25] proposed a concept of “size factor (T)”, which is the return period of the sample data following a generalized extreme value distribution (GEVD). This method forms the basis of the framework proposed in this paper. In addition, considering the GEVD may not always be the best-fitting probability distribution model for measured thickness data, in this paper, we add a procedure to select a best-fitting probability distribution model among some candidate distributions mentioned in [24].

As a holistic model, in addition to data processing, the proposed framework includes steps to conduct integrity assessment following corrosion assessment. This includes the estimation of minimum thickness for FFS assessment, corrosion rate for remaining life prediction and the plot of probability of failure (POF) for RBI. Reference [27] provided POF calculation method using statistical data. In this paper, same method is applied to calculate the probability of failure caused by corrosion using thickness data. Reference [10] gave equations for corrosion propagation for rebar in concrete structure. In this paper, to better estimate the corrosion propagation for the target bridges, regression method is applied when user have historical data. According to API 579/ASME FFS-1 [4], if the remaining thickness is less than the design minimum required thickness, the component is not fit for continuous service. Since corrosion is a time dependent damage mechanism, it is important to predict when the component will fail as well as its failure probability even if it was serviceable at the time of inspection. A limit state equation was proposed in [23], and applied in analyzing corrosion of a pipeline after 20 years in operation. In which case, failure was defined as that leakage would happen when the remaining thickness becomes zero. These models will also be included in this paper as part of the assessment framework. As part of the holistic model, the Bayesian approach, which has a powerful capability of probability reasoning, dynamic behavior modeling and multi-model synthesis [17], is applied for POF updates after inspection carried out. The model is developed upon the assumption that higher inspection effectiveness levels will reduce the uncertainty of the damage state of the component hence improving the accuracy of the assessment.

Although various analysis methods have been developed and guidelines have been proposed to apply these methods routinely, this paper seeks to provide step-wise analysis procedures in a single document. The software developed implementing this framework will help the non-specialists in statistics to perform data collection, analysis and extrapolation.

### 3. Framework flowchart

Figure 1 is the flow chart of the overall approach from data collection and processing to risk prediction and inspection scheduling, including function to update the results after receiving additional data. The following paragraphs will explain each step in detail. In the flowchart, the text box filled with green color are the input parameters, which mainly include:

- Thickness data (including both historical and new thickness readings);
- Minimum required thickness (or design thickness and corrosion allowance);
- Inspection specification;
- Service age;
- Consequence of failure (COF).

The text box filled with pink color is the expected output from this assessment, including:

- The best fit of the distribution that the thickness measurement follows;
- Representative thickness of the current data;
- Corrosion rate;
- Predicted remaining useful life (RUL);
- Risk value and a risk matrix.

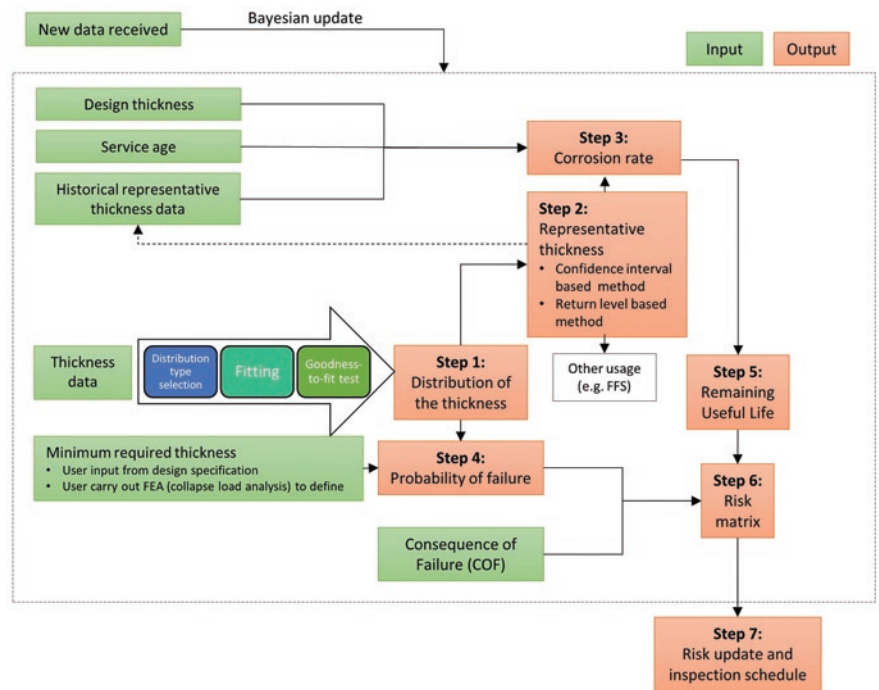


Fig. 1. Flow chart for the overall approach

#### 3.1. Step 1: Fit a distribution

Statistical corrosion assessment is supported by the belief that a statistical sample may follow the same distribution of its parent population and this may be a fundamental distribution [14, 29]. It means analysing the thickness measurements from a local area to extrapolate the condition of the entire component. Accordingly, the first step after receiving the thickness data is to characterise it by a distribution.

This process starts from identifying the candidate distributions. However, the choices of the statistical distributions seem endless, but the options can be narrowed by identifying the following properties:

- Is the data discrete or continuous?
- What is the Skewness and Kurtosis of the data set?
- How is the likelihood of observing extreme values in the distribution?

By identifying the above properties, it becomes easier to choose the statistical distribution for a given dataset. A simplified procedure for checking appropriate types of distribution taking into account of these criteria is shown in Figure 2 adapting information from [8]. According to [25], the distributions commonly used for corrosion data includes normal, lognormal, Poisson, Spatial and three types of extreme value distributions (Type I, II, III as shown in Figure 3).

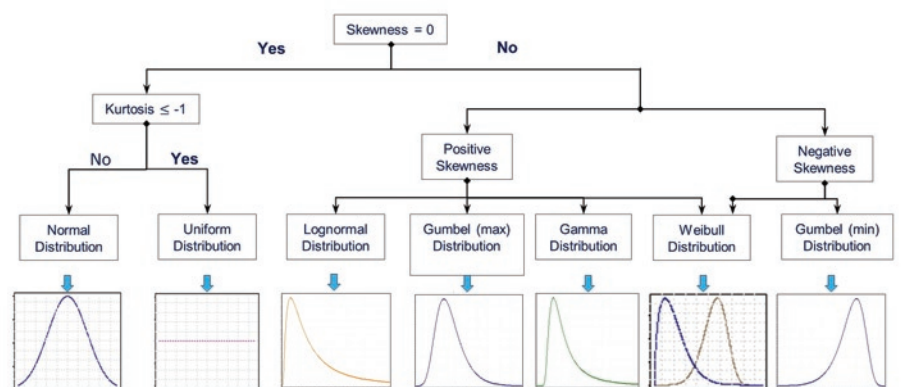


Fig. 2. Simplified distributional choices

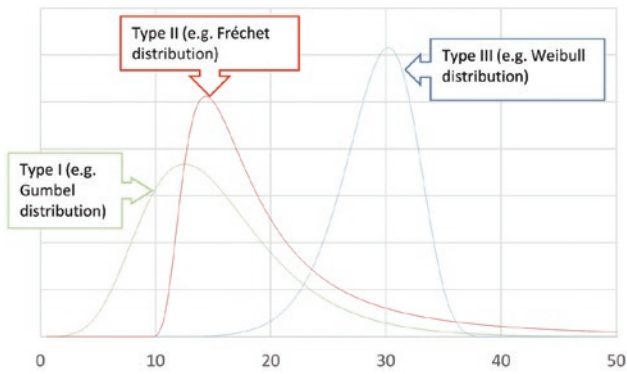


Fig. 3. Typical example PDF of Type I, II, III distributions of GEVD

The parameters of the distributions can be estimated using various methods such maximum likelihood estimation (MLE), method of moments, maximum spacing estimation or median rank regression, etc., depending on the target distribution. Finally, the goodness-of-fit test should be performed to quantify how the selected distribution matches the original thickness data. There are numerous statistics measurements and statistical fitting tests, which are commonly used for evaluating the goodness of the distribution fitting. Some of the popular statistics measurements are R squared and root mean square error, and statistical fitting tests include Kolmogorov-Smirnov test, Anderson-Darling test, Cramer-von Mises test, Chi-Squared test, etc.

### 3.2. Step 2: Calculation of representative thickness

Traditionally, the minimum measured thickness from the inspection is used to estimate the corrosion rate, perform FFS assessment and predict RUL of the component. However, using the lowest value as minimum measured thickness is overly conservative for many applications. Instead, it is proposed in this model to choose a representative value of given thickness data for the assessment, called “representative thickness ( $t_r$ )”.

Depending on the nature of the data, there are three commonly employed analysis methods:

**Method A:** Identify a fit to the underlying distribution (parent distribution) of the raw data.

**Method B:** Partition the surface into rectangular “blocks”, and fit an extreme value distribution to the minimum thickness/ maximum wall loss of these blocks.

**Method C:** The peak-over-threshold method by fitting a Generalised Pareto distribution (GPD) to the exceedances.

Method A is usually the first choice and the classical method to evaluate corrosion by extrapolation based on confidence index of statistical distribution that is symmetric such as Normal, Logistic, Cauchy, Uniform distributions, etc. However, in fact, the true parent distribution from inspection sample data is hard to identify. Moreover, the confidence index approach may not applicable to the asymmetric distributions that may depict reality more closely. Method B, using extreme value theory, is a good alternative, especially for localised corrosion, to model the extremely corroded area of equipment and components. In cases when the corrosion does not happen uniformly or block minima/maxima cannot be extracted, such as tank bottom wall loss (pit depth), Method C based on GPD is more appropriate. Method C is outside the remit of this paper as the underlying data is not of the type to which it is suited.

The flow chart of calculating the representative thickness is shown in Figure 4. For Method A, the confidence interval (CI) is a user input based on engineering judgement. The procedure of applying Method B is as follows:

Taking the example of minimum generalised extreme value (GEV (min)) distribution, its CDF is written as:

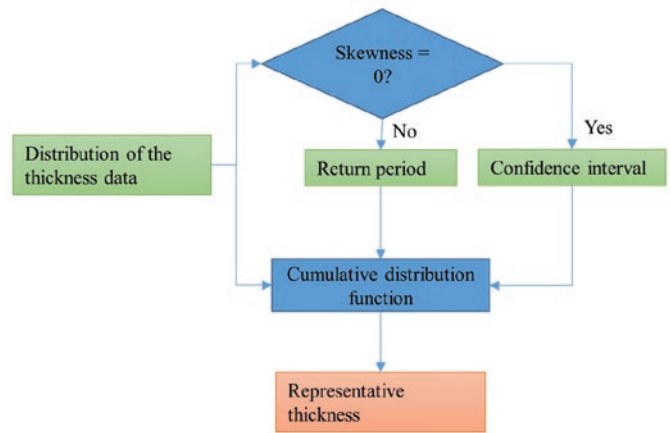


Fig. 4. Flow chart of representative thickness calculation

$$F(x; \mu, \sigma, \xi) = \begin{cases} 1 - \exp \left\{ - \left[ 1 + \xi \left( \frac{-x + \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}, \xi \neq 0 \\ 1 - \exp \left[ - \exp \left( \frac{x - \mu}{\sigma} \right) \right], \xi = 0 \end{cases} \quad (1)$$

where the  $\xi$ ,  $\mu$ , and  $\sigma$  represent a shape, location, and scale of the distribution function, respectively.  $\sigma \geq 0$ , and  $1 + \xi \left( \frac{x - \mu}{\sigma} \right) \geq 0$ . In extreme value terminology, quantities such as  $t_r$  is normally named as return level. In GEVD, the return level is defined as a level that is expected to be equal or exceeded on average once every certain observations/ interval (T) with a probability of p [7], T is called return period and:

$$p = \frac{1}{T} \quad (2)$$

The return level  $X_p$  hence can be calculated using quantile function as:

$$X_p = \begin{cases} \bar{\mu} + \frac{\bar{\sigma}}{\xi} \left\{ 1 - [-\ln(1-p)]^{-\xi} \right\}, \xi \neq 0 \\ \bar{\mu} + \bar{\sigma} \cdot \ln[-\ln(1-p)], \xi = 0 \end{cases} \quad (3)$$

To apply this method, analyser needs to identify the return period first. To do so, Shibata [6]. proposed a method called “size factor”, where the return period T is calculated as:

$$T = \frac{A}{S} \quad (4)$$

As illustrated in Figure 5, A is the surface area of the object, i.e. the total area of inspection and S is the small specimen that is sampled randomly from this object. Apply size factor to Equation (1) to (3), the representative thickness can be calculated. In this process, the calculated return level  $t_r$  represents the maximum corrosion at a given return period, T. In the other word, it estimates the maximum corrosion for the larger surface area (A), which is T times larger than the small sample area (S). Choice of block size S is scenario based and critical. If it is too small, the limiting arguments supporting the GEV might not be valid, and the extrema will be too close to assume iid (independent and identically distributed), while if it is too large, there will be insuf-



efficient data for analysis and result in sampling errors. One possible approach in this scenario is to define the block size by looking at the correlation among data. It is to be noted that Extreme Value Theory (EVT) is based on the assumption that individual thickness measurements are iid variables that any correlation between the data is negligible. Stronger correlation between adjacent data points is described in [21] as “clustering” and such high degree of correlation would overthrow the basic assumption of EVT and make it difficult to extrapolate the underlying distribution of raw data, even if Method A had revealed an adequate fit [23]. Reducing the sample to that of the block extrema can help to mitigate such clustering effect. Therefore, in principle, the chosen dimension (or area) of the block should enable the pairs of data points separated by such dimension (or area) to be weakly correlated. In other words, to reduce clustering effect, the chosen dimension of the block should enable sampled data points to be weakly correlated. For instance, the approach used by [23] to gauge the strength of correlation is through computing the “two-dimensional auto-correlation function (2D ACF)” by applying a pair of Fast Fourier transforms.

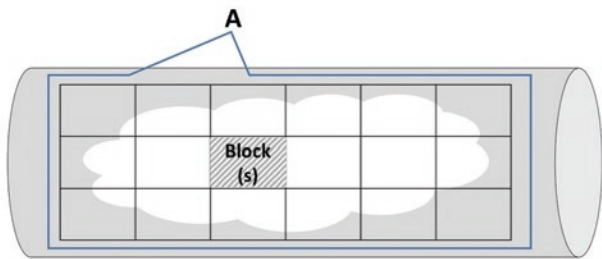


Fig. 5. Example of inspection surface

### 3.3. Step 3: Corrosion rate calculation

Corrosion rate is the speed at which any metal in a specific environment deteriorates. It can also be defined as the loss per year in thickness of a metal component due to corrosion. The corrosion rate depends on the environmental conditions and the properties of the material [15]. The corrosion rate can be obtained by carrying out laboratory tests, historical data analysis or calculation using the formulas from codes and guidelines. In terms of data driven approach, depending on the availability of the data, there are two methods:

Method 1: When the historical data is not available, the corrosion rate can be calculated by Equation (5).

$$CR = \frac{t_r - t_{nom}}{Age} \quad (5)$$

where  $t_r$  is the current representative thickness,  $t_{nom}$  is the design nominal thickness of the component. Age is the time in service so far.

Method 2: If the historical thickness data are available, the corrosion rate can be calculated by fitting all the representative thickness data to a regression equation, such as linear regression and polynomial regression. According to the linear regression method, the corrosion rate CR is an absolute value of the slope of the fitted line. An example of corrosion rate calculation using linear regression model is shown in Figure 6.

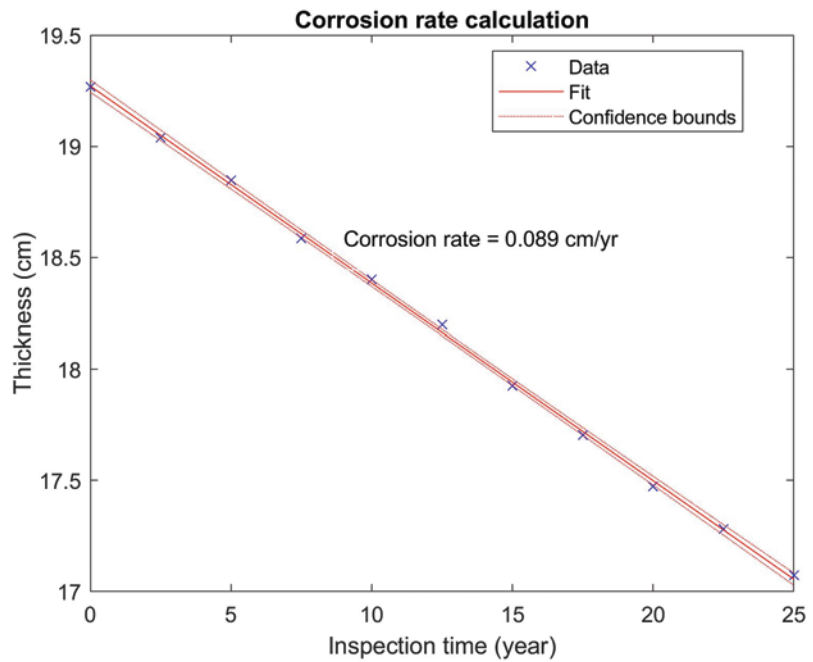


Fig. 6. An example of corrosion rate calculation using linear regression model

### 3.4. Step 4: Calculation of the probability of failure (POF)

The probability of failure (POF) represents the probability that the component is no longer fit for service. Definition of failure is based on the purpose of assessment. In corrosion assessment, when the thickness degraded to below a certain critical value, the component is considered not serviceable any more, i.e. the component is failed due to corrosion. The critical thickness could be the minimum required thickness ( $t_{min}$ ) retrieved from an FFS assessment or a value given by engineer such as design nominal thickness deducted by future corrosion allowance. The process of calculating the POF is shown in Figure 7.

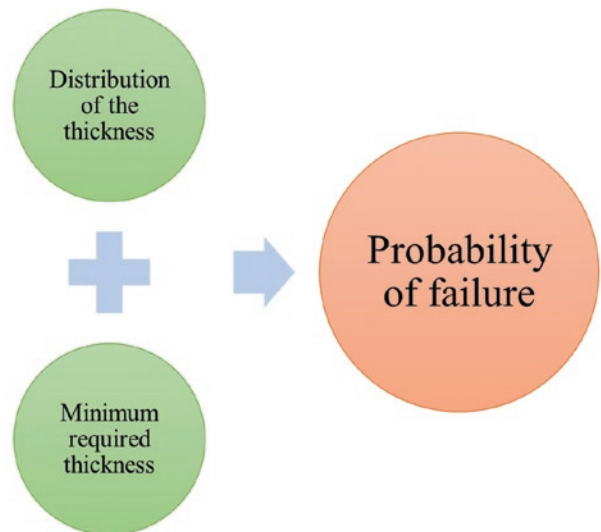


Fig. 7. Flow chart of calculating the POF

The equation of POF at the time of inspection is given by:

$$POF = \int_0^{t_{min}} f(s)ds \quad (6)$$

where  $f(s)$  is probability density function (pdf).  $t_{min}$  is the minimum required thickness.

While in case if the POF prediction is of interest, the corrosion rate (CR) should be factored in the assessment. According to the above-mentioned failure threshold, a limit state equation (LSE) for corrosion damage can be defined as [5]:

$$z(\text{time}) = (t_m - CR \cdot \text{time}) - t_{min} \quad (7)$$

where  $t_m$  is the observed thickness measurements and  $\text{time}$  is the independent variable. Considering  $t_m$  and/or CR are distributed, the future POF then can be defined as:

$$\text{POF} = P[z(\text{time}) \leq 0] \quad (8)$$

To compute the POF from above equations, the Monte Carlo Simulation (MCS) method or First / Second Order Reliability Method (FORM/SORM) are good options.

### 3.5. Step 5: Prediction of RUL

The remaining useful life (RUL) of an asset is defined as the length from the current time to the end of the useful life [26]. According to its definition, the RUL can be calculated by Equation (9):

$$\text{RUL} = \frac{t_r - t_{min}}{CR} \quad (9)$$

where  $t_{min}$  is the minimum required thickness that is obtained in Step 4,  $t_r$  is the current representative thickness that is calculated in Step 2 and  $CR$  is the corrosion rate that obtained in Step 3.

### 3.6. Step 6: Show risk assessment and prediction on a risk matrix

In a risk assessment approach, a risk matrix is developed to define the level of risk by considering the category of probability against the category of consequence severity. It increases the visibility of risks and assist management decision making. On the risk matrix, the current risk of the asset system under consideration and the RUL should be presented on the risk matrix to support the end user in inspection scheduling.

In the risk matrix, the consequence of failure (COF) is defined as the potential consequences for the component, personnel, and the environment if the adverse event occurred. In this paper, the COF is assessed using standards proposed in [33], is divided into four categories listed in Table 1; the POF has four categories as well as listed in Table 2.

Table 2 The POF ranking and value

Ranking	POF value
Low	0% - 0.25%
Moderate	0.25% - 0.5%
High	0.5% - 0.75%
Severe	0.75% - 1%

### 3.7. Step 7: Updated probability using Bayesian theory

Bayesian theory is a probabilistic approach which applies the conditional probability principals to work with uncertainties [11]. In this paper, the Bayesian theory is applied for POF updates when there is additional inspection data. The model is developed upon the assumption that higher inspection effectiveness levels will reduce the uncertainty of the damage state of the component hence improving the accuracy of the assessment. The inspection effectiveness is measured as probability of detection (POD) and false call ratio (FCR).

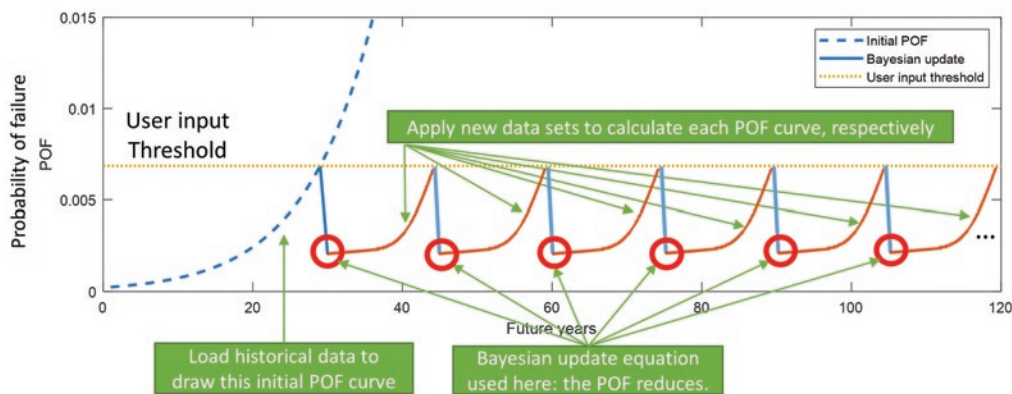


Fig. 8. Probability of failure curve when applied new data and Bayesian theory

Table 1. The COF ranking and value

Ranking	COF value	Description
Low	1	None or very minor effect on safety, health, and environment.
Moderate	2	A moderate effect. The system requires scheduled repair. A failure which may cause moderate injury, moderate property damage, or moderate system damage which will result in delay or loss of system availability or mission degradation. 100% of the mission may need to be reworked or process delayed.
High	3	System performance is degraded. Some safety functions may not operate. A failure causes injury, property damage, or system damage. Some portion of the mission is lost. High delaying restoring function.
Severe	4	The system is inoperable with loss of primary function. Failure can involve hazardous outcomes and/or noncompliance with government regulations or standards.

Table 3. Conditional probability table for driving Equation (10)

	Defect yes (A <sup>+</sup> )	Defect no (A <sup>-</sup> )	
UT results, positive (B <sup>+</sup> )	$P(A^+ \cap B^+) = \text{POF}_{\text{initial}} \times \text{POD}$	$P(A^- \cap B^+) = (1 - \text{POF}_{\text{initial}}) \times \text{FCR}$	$P(A^+ \cap B^+) + P(A^- \cap B^+)$
UT results, negative (B <sup>-</sup> )	$P(A^+ \cap B^-) = P(B^-   A^+) \cdot P(A^+)$ $P(A^+) = \text{POF}_{\text{initial}} \times (1 - \text{POD})$	$P(A^- \cap B^-) = (1 - \text{POF}_{\text{initial}}) \times (1 - \text{FCR})$	$P(A^+ \cap B^-) + P(A^- \cap B^-)$
	$\text{POF}_{\text{initial}}$	$1 - \text{POF}_{\text{initial}}$	1

Figure 8 shows the POF and its updating using Bayesian theory. The initial POF curve is obtained by using historical data, and applying the method described from Step 1 to Step 6. Accordingly, future inspection (or maintenance) is scheduled, which provides new data to update the initial POF. The new datasets provide new POF (POF<sub>new</sub>) curves corresponding to the data as processed from Step 1 to Step 6 (the orange curves shown in Figure 8).

Note that an updated POF value (POF<sub>update</sub>) can be calculated using Equation (10) if the effectiveness of inspection is available in terms of POD and FCR. An updated POF value can be obtained for POF<sub>initial</sub> or POF<sub>new</sub> although only the former is shown in Equation (10).

$$\text{POF}_{\text{update}} = \frac{\text{POF}_{\text{initial}} \times (1 - \text{POD})}{\text{POF}_{\text{initial}} \times (1 - \text{POD}) + (1 - \text{POF}_{\text{initial}}) \times (1 - \text{FCR})} \quad (10)$$

where POF<sub>initial</sub> is the POF evaluated from previous inspection; POD is the probability of detection, which is used to describe the capability of an inspection to detect flaws; and FCR is the false call ratio (or called false positive ratio), which means the rate that negative events (i.e. no defects) are wrongly categorized as positive (i.e. defects).

Table 3 shows how Equation (10) was derived following recommendations by API 581 [1]. First, assume the probability of UT results regarding unacceptable corrosion defects showing positive is P(B<sup>+</sup>) and the current probability of a component having such defects is P(A<sup>+</sup>) that is equal to POF<sub>initial</sub>. Then complete Table 3 accordingly; the updated probability of failure, according to the rule of conditional probability [13], can be calculated using Equation (11) which is equivalent to Equation (10):

$$P(A^+ | B^-) = \frac{P(B^- | A^+) \cdot P(A^+)}{P(B^-)} \quad (11)$$

#### 4. Application of software implementing the framework: a steel bridge case study

Corrosion is a major cause of deterioration in steel bridges [15]. The results of this deterioration can range from progressive weakening of a bridge structure over a long period of time to sudden bridge collapse [15]. Hence, corrosion damage must be carefully appraised and evaluated. In some cases, immediate repair or closure is necessary, while in other cases, the conditions created by corrosion can be

tolerated. In order to mitigate the risk of unexpected collapse of a bridge, key components of a bridge should be identified and inspected periodically. In this case study, we focused on one of the most essential components in bridges – girder. The implementation of the proposed model is presented in this section.

The raw data of a bridge girder was obtained from ultrasonic sensors, and was then transformed to thickness readings. The developed software provided the risk assessment of the structure by performing corrosion assessment using statistical method. One critical input of the software was the thickness readings; the histogram of thickness data is shown in Figure 9.

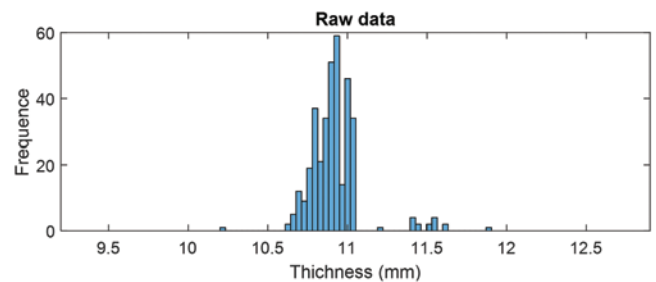


Fig. 9. Histogram of the raw thickness data

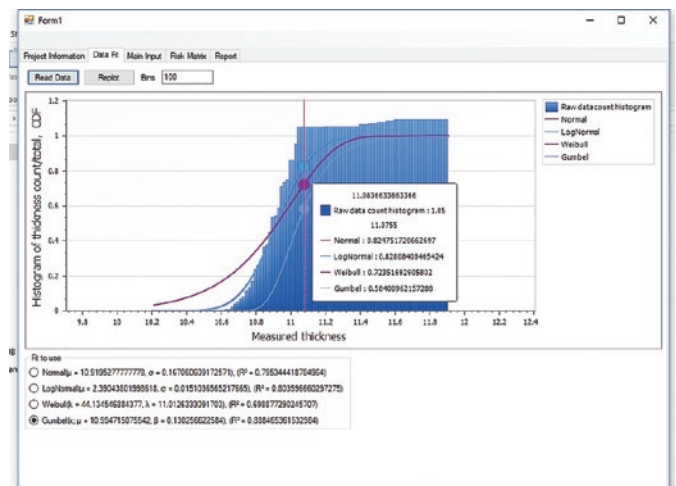


Fig. 10. CDF of real data and fitted distributions

After loading the thickness data into the software, the next step was to select a statistical distribution that best fits to the thickness data. In this case, the candidate distributions include Weibull, Gumbel, Normal and Log-normal distribution. R squared value was applied to

Table 4. R square value of different distributions

Distribution name	Normal	Lognormal	Weibull	Gumbel
parameters	$\mu = 10.9195$	$\mu = 2.3904$	$k = 44.1345$	$\mu = 10.9947$
	$\sigma = 0.1671$	$\sigma = 0.0151$	$\lambda = 11.0126$	$\sigma = 0.1303$
R square value	0.7953	0.8036	0.6989	0.8885

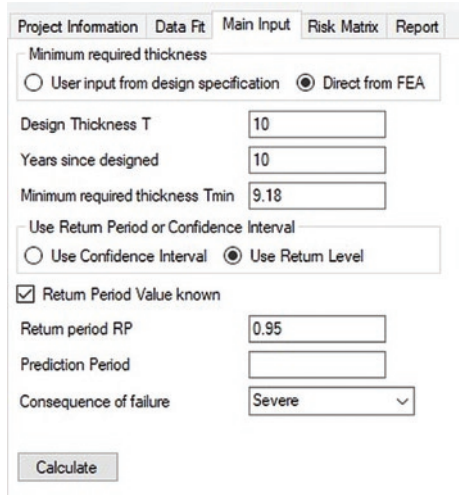


Fig. 11. User input screenshot

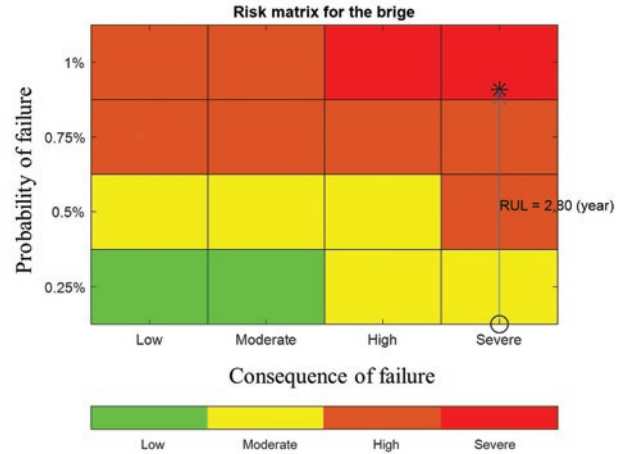


Fig. 13. Risk matrix of the bridge girder

Table 5. The output of the software

Output	POF threshold	Representative thickness (mm)	Current POF	Corrosion rate (mm/yr)	Current remaining useful life (year)
Values	7.8394e-03	10.59	6.3323e-28	0.5	2.8

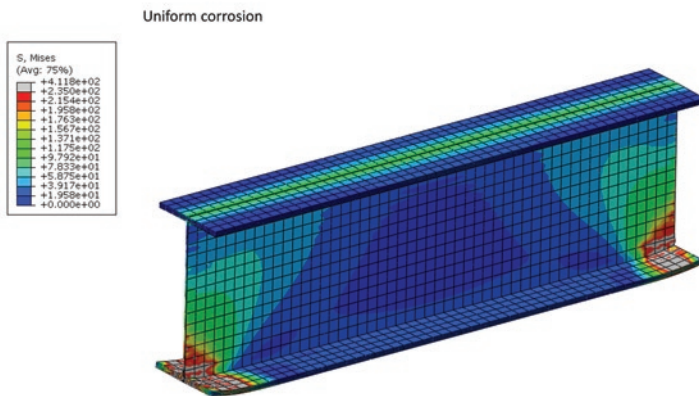


Fig. 12. Uniform corrosion by FEA method

assess the goodness of fit. A software screenshot showing the CDF of real data and fitted distributions is presented in Figure 10. The R squared value of different distributions is listed in Table 4.

In this case, the Gumbel distribution was finally selected. Other main input parameters are given in Figure 11. The design nominal thickness of the bridge girder  $t_{nom}$  was 10 mm; the minimum required thickness was directly obtained from finite element analysis (FEA), and  $t_{min} = 9.18$  mm for uniform (see Figure 12); Service age was estimated to be 10 years; the return period was given as 0.95. The COF was estimated to be 'severe' if the girder failed.

The output of the software is shown in Table 5. The current risk and predicted RUL are presented on the risk matrix in Figure 13. A risk matrix would typically show the risk profile of a system of components i.e. a system of girders to support operators to prioritise inspections based on risk. In this application, for simplicity, only one girder is shown.

The final output of this software is the prediction of POF using Bayesian theory. The method of calculating the POF curves and the predicted POF after inspection can be found in Section 5.7. In this case, we assumed that the corrosion rate remained the same. The POD was 0.75 and FCR was 0.15. The POF curves when applied to bridge girder

data and Bayesian theory is shown in Figure 14. As shown in the figure, the next inspection time was after 2.8 years, and within 14 years, given the existing risk profile, there should be four inspections to keep the bridge girder safe. After every inspection, the risk profile needs to be re-calculated based on the most updated information.

### 5. Conclusions

This paper proposes an integrated framework to carry out the corrosion assessment via statistically analyzing thickness measurements. It is flexible and widely applicable to various engineering structures that suffer from corrosion. The methodology uses Bayesian theory to update the likelihood of the event

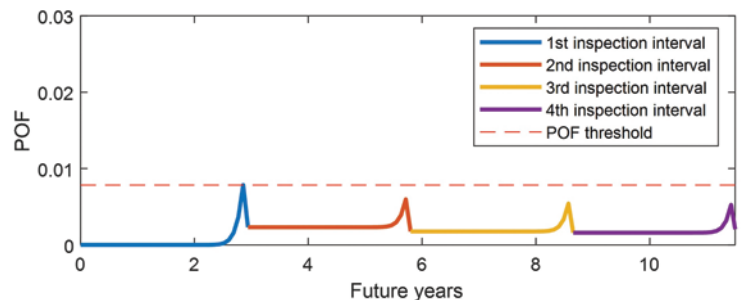


Fig. 14. POF curves when applied bridge girder data and Bayesian theory

occurrence and failure probability. The framework is aimed at supporting decision makers in optimising their asset integrity management strategy such that they can cost effectively maintain risk within tolerable levels.

The paper builds on recent work at TWI aimed at using statistical techniques on inspection data [2, 3, 6, 12]. The use of such techniques is getting increasingly common with the availability of more data. This trend has been facilitated by advances in sensor technology and computing power, and is likely to grow stronger as plant operators look for ways to improve safety and reliability of their assets.

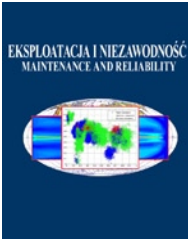
### Acknowledgement

*Data presented in this paper was partially generated within the Innovate UK (IUK) funded project - ASSAI (Development of an unmanned aerial system for advanced contact inspection of civil structures) Project, Ref.104405- in which TWI Ltd was one of the partners; the authors are grateful to IUK for providing them this opportunity. The authors also thank other project partners and JR Dynamics Ltd (JRD), for sharing UT data.*

### References

1. American Petroleum Institute. A.P.I. 581, Risk-based Inspection Methodology. Third Edition.2016.
2. Anes-Arteche F, Yu K, Bharadwaj U et al. Challenges in the application of DCVG-survey to predict coating defect size on pipelines. *Materials and Corrosion-Werkstoffe und Korrosion* 2017; 68(3): 329-337, <https://doi.org/10.1002/maco.201608917>.
3. Anes-Arteche F, Bharadwaj U, Yu K, Lee C. Correlation of Pipeline Corrosion and Coating Condition with ECDA Survey Results. in *EUROCORR*, 2016.
4. API. Fitness-For-Service (API 579 Second Edition), Part 5 Assessment of Local Metal Loss. 2007.
5. Bharadwaj U R. Risk based life management of offshore structures and equipment. 2010.
6. Bin Muhd Noor N N, Yu K, Bharadwaj U, Gan T-H. Making use of external corrosion defect assessment (ECDA) data to predict DCVG% IR drop and coating defect area. *Materials and Corrosion-Werkstoffe und Korrosion* 2018; 69(9): 1237-1256, <https://doi.org/10.1002/maco.201810085>.
7. Chikobvu D, Chifurira R. Modelling of extreme minimum rainfall using generalised extreme value distribution for Zimbabwe. *South African Journal of Science* 2015; 111(9-10): 01-08, <https://doi.org/10.17159/sajs.2015/20140271>.
8. Damodaran A. Strategic risk taking: a framework for risk management. Pearson Prentice Hall: 2007.
9. Dorafshan S, Maguire M, Collins W. Infrared thermography for weld inspection: feasibility and application. *Infrastructures* 2018; 3(4): 45, <https://doi.org/10.3390/infrastructures3040045>.
10. Ghanooni-Bagha M, Shayanfar M A, Reza-Zadeh O, Zabihi-Samani M. The effect of materials on the reliability of reinforced concrete beams in normal and intense corrosions. *Eksplatacja i Niezawodnosc – Maintenance and Reliability* 2017; 19 (3): 393–402, <http://dx.doi.org/10.17531/ein.2017.3.10>.
11. Kalantarnia M, Khan F, Hawboldt K. Dynamic risk assessment using failure assessment and Bayesian theory. *Journal of Loss Prevention in the Process Industries* 2009; 22(5): 600-606, <https://doi.org/10.1016/j.jlpi.2009.04.006>.
12. Khalili H, Oterkus S, Bartrop N, Bharadwaj U. Different Bayesian methods for updating the fatigue crack size distribution in a tubular joint. *Journal of Offshore Mechanics and Arctic Engineering* 2020, <https://doi.org/10.1115/1.4048155>.
13. Kolmogorov A. Foundations of the theory of probability: Second English Edition.
14. Kowaka M, Tsuge H. Introduction to life prediction of industrial plant materials: Application of the extreme value statistical method for corrosion analysis. Allerton Press 1994; 26(10): 559-559, <https://doi.org/10.1002/mawe.19950261013>.
15. Kulicki J M, Prucz Z, Sorgenfrei D F et al. Guidelines for evaluating corrosion effects in existing steel bridges. 1990.
16. Lee J Y, Sim C, Detweiler C, Barnes B. Computer-Vision Based UAV Inspection for Steel Bridge Connections. *Structural Health Monitoring* 2019 2019, <https://doi.org/10.12783/shm2019/32473>.
17. Mi J, Li Y-F, Beer M et al. Importance measure of probabilistic common cause failures under system hybrid uncertainty based on bayesian network. *Eksplatacja i Niezawodnosc - Maintenance and Reliability* 2020; 22(1): 112, <https://doi.org/10.17531/ein.2020.1.13>.
18. National Association of Corrosion. Internal corrosion direct assessment methodology for liquid petroleum pipelines. NACE International: 2008.
19. Nikhil V P, Wiston B R, Ashok M. Flaw detection and monitoring over corroded surface through ultrasonic C-scan imaging. *Engineering Research Express* 2020; 2(1): 015010, <https://doi.org/10.1088/2631-8695/ab618d>.
20. Pham N H, La H M, Ha Q P et al. Visual and 3D mapping for steel bridge inspection using a climbing robot. *ISARC 2016-33rd International Symposium on Automation and Robotics in Construction*, 2016, <https://doi.org/10.22260/ISARC2016/0018>.
21. Reiss R-D, Thomas M. Statistical analysis of extreme values: from insurance, finance, hydrology and other fields. Springer: 1997, <https://doi.org/10.1007/978-3-0348-6336-0>.
22. Scarf P A, Laycock P J. Applications of extreme value theory in corrosion engineering. *Journal of Research of the National Institute of Standards and Technology* 1994; 99: 313-313, <https://doi.org/10.6028/jres.099.028>.
23. Schneider C. Extreme Value Analysis and Corrosion Mapping Data. Paper presented at 4th European-American Workshop on Reliability of NDE, 2009; 24: 26.
24. Selech J, Andrzejczak K. An aggregate criterion for selecting a distribution for times to failure of components of rail vehicles. *Eksplatacja i Niezawodnosc – Maintenance and Reliability* 2020; 22 (1): 102–111, <http://dx.doi.org/10.17531/ein.2020.1.12>.
25. Shibata T. Corrosion probability and statistical evaluation of corrosion data. *Uhlig's Corrosion Handbook*, third Edition, 2011; 21: 367, <https://doi.org/10.1002/9780470872864.ch27>.
26. Si X-S, Wang W, Hu C-H, Zhou D-H. Remaining useful life estimation-a review on the statistical data driven approaches. *European Journal of Operational Research* 2011; 213(1): 1-14, <https://doi.org/10.1016/j.ejor.2010.11.018>.
27. Sobaszek Ł, Gola A, Świć A. Time-based machine failure prediction in multi-machine manufacturing systems. *Eksplatacja i Niezawodnosc – Maintenance and Reliability* 2020; 22 (1): 52–62 <http://dx.doi.org/10.17531/ein.2020.1.7>.
28. Stone M. Wall thickness distributions for steels in corrosive environments and determination of suitable statistical analysis methods. *Proc. 4th Eur.-Amer. Workshop Rel. NDE*, 2009: 1-23.
29. TWI Limited. Guidelines for use of statistics for analysis of sample inspection of corrosion. 2002.
30. Villa T F, Gonzalez F, Miljevic B et al. An overview of small unmanned aerial vehicles for air quality measurements: Present applications

- and future perspectives. *Sensors* 2016; 16(7): 1072, <https://doi.org/10.3390/s16071072>.
31. Wang R, Kawamura Y. Development of climbing robot for steel bridge inspection. *Industrial Robot: An International Journal* 2016; 43(4): 429-447, <https://doi.org/10.1108/IR-09-2015-0186>.
  32. Wang R, Kawamura Y. An automated sensing system for steel bridge inspection using GMR sensor array and magnetic wheels of climbing robot. *Journal of Sensors* 2016, <https://doi.org/10.1155/2016/8121678>.
  33. Washer G, Nasrollahi M, Applebury C et al. Proposed guideline for reliability-based bridge inspection practices. 2014, <https://doi.org/10.17226/22277>.



Article citation info:

Konowrocki R, Kalinowski D, Szolc T, Marczewski A. Identification of safety hazards and operating conditions of the low-floor tram with independently rotating wheels with various drive control algorithms. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 21–33, <http://dx.doi.org/10.17531/ein.2021.1.3>.

## Identification of safety hazards and operating conditions of the low-floor tram with independently rotating wheels with various drive control algorithms

Indexed by:



Robert Konowrocki<sup>a,\*</sup>, Dariusz Kalinowski<sup>a,b</sup>, Tomasz Szolc<sup>a</sup>, Artur Marczewski<sup>b</sup>

<sup>a</sup>Institute of Fundamental Technological Research, Polish Academy of Sciences, ul. Pawińskiego 5b, 02-106 Warsaw, Poland

<sup>b</sup>PESA Bydgoszcz SA, ul. Zygmunta Augusta 11, 85-082 Bydgoszcz, Poland

### Highlights

- Bogies with independently rotating wheels increase travel comfort.
- System of independently rotating wheels (IRW) reducing the value of wheel-rail forces.
- Higher efficiency of tram drive control through additional braking.
- New geometry of simulational test track for light rail vehicles (LRVs).
- New control of an innovative drive with a crank axle for light rail vehicles (LRVs).

### Abstract

The aim of the article is to develop a method for the analysis of tram dynamics related to safety during operation. To achieve this, a mathematical model of the vehicle represented by a multibody simulation MBS system is used. Models of tram with a classic and innovative drive, based on a system of independently rotating wheels on crank axles are analyzed. A new configuration of an innovative drive control of the considered vehicle with the use of braking of independent wheels is proposed. A new geometry of test track is presented. During numerical investigation the values of 'Y' leading forces of tram wheels with the considered innovative drive proved to be lower than in the corresponding vehicle with standard wheelsets. It has been demonstrated that the active control systems are of key importance and should be applied in such innovative tram drives.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

maintenance of safety, reliability of trams, derailment of tramcar, numerical tests, drive with independently rotating wheel, SIMPACK Rail.

## 1. Introduction

In years 2014 to 2020, the European transport infrastructure began to change. This transformation was triggered by introducing common EU policy aimed at effective, safe and economically efficient movement of people and goods as well as reduction of environmental pollution as a part of a new EU sustainable development transport policy. Poland is known as the EU member state that, within the framework of the operational program Infrastructure and Environment, submitted the largest number of applications by the year 2020 (the total number of 178). In the period 2014-2020 (fig. 1) the projects were financed from the European Structural and Investment Funds (ESIF), the the Cohesion Fund (CF) and the European Regional Development Fund (ERDF). Great amount of submitted applications concerned the design and construction of new public transport infrastructure as well as modernization of the existing one, including the tram rolling stock [47, 33]. In Poland, very good examples of projects financed from the ESIF were tram projects approved for implementation, e.g. "Construction of a tram line along Kujawska street, from Kujawskie roundabout to Bernardyńskie roundabout, along with the extension of road network, reconstruction of rail transport system and rolling stock purchase in Bydgoszcz" (cost EUR 86 579 938), "Reconstruc-

tion of tramway tracks in Szczecin, stage II" (cost EUR 77 014 409), "Construction of a tramline to Gocław in Warsaw with the rolling stock purchase" (cost EUR 91 019 802), "Comprehensive integration programme for the low-emission public transport system in Łódź Metropolitan Area, including the purchase of means of transport for service on the W-Z (East-West) and other transport routes, as well as modernisation of tram depots in Łódź" (cost EUR 151 759 547).

The above-mentioned facts prove, that there is the necessity of an intensive development in the area of tram transport, and greater commitment to conduct research and tests related to the improvement of reliability, safety and maintenance of tram fleet and its infrastructure.

## 2. Modern stock of light rail vehicles

One of the above-mentioned methods to develop and increase the efficiency of the public transport, especially in urban area, are modern tram vehicles. One category of the Light Rail Vehicles (LRV) [9] are low-floor trams which contributed to the increase of transport efficiency and passenger comfort improvement. By reducing the total time spent by the vehicle at tram stop and facilitating boarding and unboarding of passengers, the low-floor trams – compared to con-

(\*) Corresponding author.

E-mail addresses: R. Konowrocki - [rkonow@ippt.pan.pl](mailto:rkonow@ippt.pan.pl), D. Kalinowski - [dariusz.kalinowski@pesa.pl](mailto:dariusz.kalinowski@pesa.pl), T. Szolc - [tszolc@ippt.pan.pl](mailto:tszolc@ippt.pan.pl), A. Marczewski - [artur.marczewski@pesa.pl](mailto:artur.marczewski@pesa.pl)

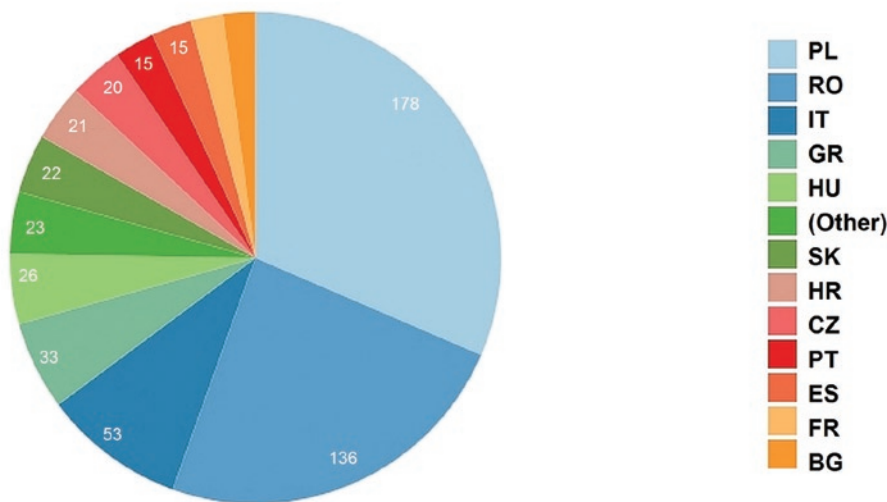


Fig. 1. The number of applications submitted by the EU member states for the development of transport infrastructure; applications under the ESIF 2014-2020 ERDF CF projects [46]

ventional trams [39] - affected the growth of public transport commuters number. The very important feature of low-floor vehicles is an easy access for People with Reduced Mobility (PRM), e.g. traveling in wheelchairs, or parents with baby strollers. Due to new passenger mobility requirements [13] a fundamental change in production technology of tram vehicles has been introduced. Greater internal volume of tram bodies as well as lowering and flattening of the floor [25] were expected. Allocation of equipment on tram roof and crucial redesign of car bodies structure, as well as independently rotating wheels (IRW) (Fig. 2) [5] were implemented. Operational introduction of trams with independently rotating wheels allowed lowering the floor height relative to the level of rail head along the entire length of the tram without lifting the level above the bogies [15]. Exemplary constructions of bogies with IRW are presented in papers [25, 23]. Such a wheelset has wheels mounted on a common axle by means of bearing system which enables independent rotation of wheels. A continuous development of IRW trams could have been observed at international fairs where tram producers presented their new vehicle designs: PESA Jazz, Siemens Avenio, Alstom Citadis and Škoda Forcity Plus. The Hyundai Rotem company is on the way to finish their work on an IRW tram, capable of traveling on curves with radius  $R = 15$  m. The vehicle is expected to be equipped with individual motor torque control and hydraulic actuators that reduce the angle of the wheelset of the bogie running on rail. Unfortunately, the research results are not commonly available. The facts presented above indicate that the research topics discussed in this article are still current, investigated and important for the transport development both on Polish and international level.

As opposed to vehicles with standard wheelsets, the application of independently rotating wheels entails direct control of wheel pulsions [4, 18]. Such control is required due to the fact that in specific conditions the IRW bogie can be characterized by worse kinematic and dynamic properties compared to standard solution with a rigid

axle that couples the angular velocities of both wheels [38, 29]. As a result of independence from the velocity constraints mentioned above, wheelsets with independently rotating wheels present difficulties in self-centering, e.g. after leaving the curve arc [2, 30]. Such behaviour has been proven in many studies, e.g. [5, 7]. The quality of operating conditions and maintenance of light tram vehicles could be significantly reduced thanks to such behaviour of a tram [19]. The occurrence of the non-centering of IRW wheelsets also reduces safety against derailment by increasing the  $Y/Q$  derailment ratio and a greater  $D_z$  wheel lift that consequently enhances the probability of a vehicle derailment [26, 14]. Co-authors of this publication developed their own method of determining the safety level against tram derailment; their assumptions are presented in article [19]. So far none of the methods of determining the safety level against tram derailment has been standardized.

The analyses of derailment safety were the subject of many publications [37, 11]. However, it should be emphasized that the quasi-static ride of a tram on a curve with a specified twist is a conservative approach to the simulation derailment analysis. Due to tram reduced velocity the conditions of such ride are particularly disadvantageous since in case of smaller velocity the dynamic friction coefficient is greater at smaller skids [24] and the curve radius of the railway track is smaller than in case of standard railway lines [46]. The motor bogies of IRW trams are equipped with motorized wheelsets (Fig. 2), each one driving one of the wheels. The difficulty in proper steering and synchronization of the system of independently rotating wheels is the main drawback of the solution in question. The existing algorithms [5, 18] allow to control such system in model conditions with a standard set of rail arrangements. However, due to different state of tracks and tram infrastructure in many municipalities [44], it is very important to adapt such type of tram drive systems to different conditions. Therefore an analysis was made to check the impact of the track gauge change on vehicle dynamics with various control configurations, ensuring right tram safety conditions. By addressing this topic the present article partially develops and supplements the subjects presented by the authors of the works [2, 12, 43]. Based on obtained results of numerical analysis, numerous researchers [8] recommended active steering and optimal control strategies for independently rotating wheels. Unfortunately, experimental validations did not take place in their case. As a result of experimental research it was proved [28] that - compared to standard wheelset - active control of the motors wheelset with independently rotating wheels gives a better stability of the vehicle. Authors of [29] showed that the application of an indirect method of controlling induction motors through the orientation of the stator magnetic field is sufficient, both in terms of response and controllability of the rail vehicle. In terms of wheel steering, braking was not taken into account in these papers.



Fig. 2. Low-floor tram vehicle motor bogie with cranked axles [50,16]



So far the approach of controlling the propulsion of trams with independently rotating wheels and the braking possibility has not been utilized. Therefore authors of this paper made attempt to include braking into approach. The main example of the execution and application possibilities of the propulsion control configuration proposed in this article is the algorithm consisting of a simultaneous indirect torque and flux control of the drive motor [4, 36]. This approach enables smooth setting of the driving torque and the braking torque of motors. In this way electrodynamic braking is achieved in the entire rotational speed range [50]. For a given vehicle velocity the braking torque value can be the same as the driving torque value. The feature of asynchronous motors applied in tram vehicles was implemented by the authors of this paper for the control variant with the option of braking a given wheel discussed in the article.

This paper presents the results of studies of the implementation of several drive control algorithms in a tram as well as the results of an analysis carried out on their basis of these studies. Additionally, the impact of this control on the track guiding force and on safety against derailment is presented. A novelty in the presented method is the implementation of braking procedure while riding in small radius curves. The results of the considered algorithms of controlling the propulsion of independently rotating bogie wheels were compared with the results of standard tram wheel sets. In addition to the above, the authors proposed a new track geometry with an additional twist for tram tests which is an alternative to the the geometry presented in the EN 14363:2016 standard [10], used for extra-urban passenger rail vehicles and freight vehicles. The results of the analysis presented in this article are of great importance for the development of expertise and innovation in the field of transport.

### 3. Theoretical studies of a tramway vehicle dynamics

#### 3.1. Description of mathematical model

The model of a tramway vehicle consists of non-deformable bodies with given masses and moments of inertia that represent the dynamic properties of its individual elements. Rigid solids representing the inertial properties of the vehicle body were connected with analogous solids corresponding to bogie frames by means of massless elastic-damping elements replacing the secondary suspension.

On the other hand, crank axles with railway wheels were attached to the rigid bodies representing bogie frames by means of massless elastic-damping elements replacing the primary suspension. The rigid bodies corresponding to the individual carriages of the tramway vehicle body were interconnected by means of massless connectors. The interaction between the vehicle and railway track as

elastic normal forces and tangential friction forces in the wheel-rail contact zone were described. The presented approach to building a vehicle model is called multibody simulation method (MBS) and is utilised by many authors investigating the interaction and dynamics of railway vehicles [1, 37, 38]. In order to determine the effect of a control algorithm of vehicle drive torque on their dynamic behaviour, the considered model based on the above-described method was prepared.

The structure of the modeled tram is shown in Figure 3. The vehicle model under consideration consists of five carriages marked with the letters A-E and is based on three bogies - two motor-bogies MB and one trailer-bogie TB. Adjacent carriages are articulated. Their lower joint constitutes a spherical joint. The upper joint also uses ball joints. However, in case of an upper link between the carriages C and D, a Panhard rod is installed. This allows the carriages to rotate relative to the transverse  $y$  and vertical  $z$  axes of global coordinate. The angular displacements of the carriage relative to each other are denoted by the symbol  $\beta$  for rotation about the  $y$  axis and the symbol  $\gamma$  for rotation about the  $z$  axis.

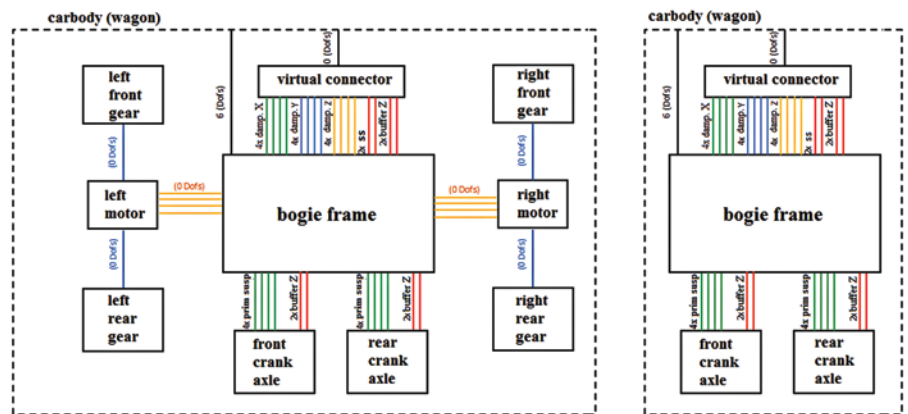


Fig. 4. Topology of the motor-bogie and trailer-bogie of the tram under consideration

The tram model considered in this paper has been divided into three basic elements, which are: vehicle parts (bodies), bogie frames and wheelsets with cranked axles. The elements of the considered vehicle represent rigid bodies with six degrees of freedom (DOF) relative to the stationary Cartesian inertial Oxyz coordinate system associated with the track (Fig. 3). The degrees of freedom define the lateral displacement  $y_{nk}$ , vertical displacement  $z_{nk}$ , longitudinal displacement  $x_{nk}$ , attack angle  $\varphi_{nk}$ , hunting angle  $\theta_{nk}$ , and galloping angle  $\phi_{nk}$ , where  $k$  takes the value  $k=1 \div 5$ , and  $n$  – corresponds to the number of vehicle units. The bogie frame is also replaced by a rigid body with five degrees of freedom that describes the lateral displacement  $y_{rj}$ , vertical displacement  $z_{rj}$ , attack angle  $\varphi_{rj}$ , hunting angle  $\theta_{rj}$ , and galloping angle  $\phi_{rj}$  (Fig. 3). In this case, the index takes values  $j=1 \div 3$ . The model of each wheelset is a subsystem consisting of 3 rigid bodies, a cranked axle to which two rigid discs representing railway wheels are rotatably attached (Fig. 4). Lateral displacements of the wheelset are marked with the symbol  $y_i$ , attack angle  $\varphi_i$ , and rolling angle of wheel  $\phi_{Li}$ ,  $\phi_{Ri}$ , different for left and right wheel. In the description of these symbols, the index  $i$  corresponds to the individual wheelsets, taking the values  $i=1 \div 6$ .

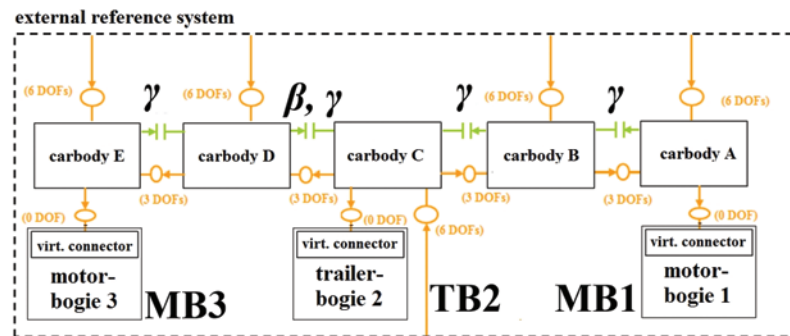


Fig. 3. Topology of considered tram model with indication of interconnections and degrees of freedom

The dynamics of the mathematical model of the considered vehicle was described by the second order ordinary differential equations. Assuming that the oscillations of the individual rigid bodies of the model relative to the reference system are small, their movement can be described by means of a linearized system of equations. This system

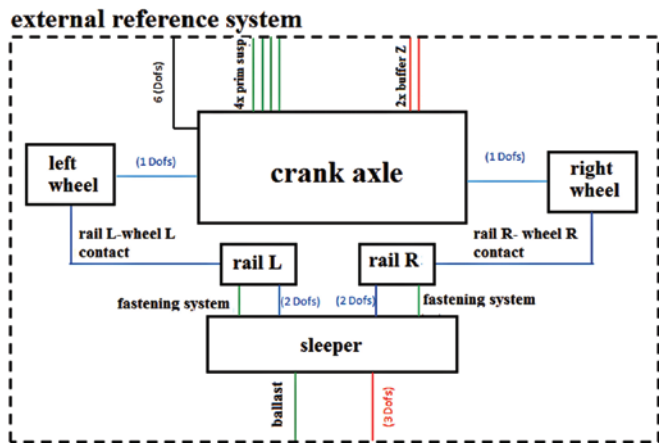


Fig. 5. Topology of a cranked axle with independently rotating wheels and railway track

of equation in the following matrix-operational form is presented below (1):

$$\left[ \mathbf{M} \frac{d^2}{dt^2} + \mathbf{D} \frac{d}{dt} + \mathbf{K} \right] \cdot \mathbf{q} = \mathbf{F} + \mathbf{W} \quad (1)$$

where:  $\mathbf{M}_{N \times N}$  denotes the mass matrix,  $\mathbf{D}_{N \times N}$  is the damping matrix,  $\mathbf{K}_{N \times N}$  describes the stiffness matrix, and  $\mathbf{q}_{N \times 1} = \{ y_i, z_i, \varphi_i, \phi_i, y_{rj}, z_{rj}, \varphi_{rj}, \theta_{rj}, \phi_{rj}, y_{nk}, z_{nk}, \varphi_{nk}, \theta_{nk}, \phi_{nk}, \dots \}^T$  is the vector of generalized coordinates,  $\mathbf{F}$  is the vector of contact forces,  $\mathbf{W}$  denotes the vector of forces from kinematic excitations caused by e.g. unevenness of the railway tracks,  $d/dt$  is the differential operator, and  $N$  denotes the number of degrees of freedom of the considered model. The number of degrees of freedom in the vehicle model under analysis was 127. As a result of the Newmark method, the system of motion equations (1) is reduced to a system of linear algebraic equations which are numerically solved with the time step  $\Delta t = 0.005$  s.

### 3.2. Wheel-rail contact model

The numerical procedures presented below have been integrated with the above-described mathematical model of a tramway vehicle, enabling the determination of the forces of dynamic interaction in the wheel-rail contact zones of the vehicle. These procedures are based on the Polach contact model [41]. In the approach used, the contact forces induced by micro-slips  $F_\xi$  and  $F_\eta$  are determined by the tangential force  $F$  acting in the contact zone, caused by the longitudinal and transverse micro-slip  $v_\xi$ ,  $v_\eta$ , as well as by the transverse force  $F_{\eta S}$  which is caused by the presence of a spin moment

$S$  resulting from the usually conical tread of the wheel profile of the considered railway vehicle. Mathematically, it can be described by equations (2), (3) and (4) as follows:

$$v_C = \sqrt{v_\xi^2 + (v_\eta + \varphi \cdot a)^2}, \quad (2)$$

$$F_\xi = F \frac{v_\xi}{v_C}, \quad (3)$$

$$F_\eta = F \frac{v_\eta}{v_C} + F_{\eta S} \frac{\varphi}{v_C}, \quad (4)$$

where:  $F$  denotes the tangential contact force caused by the longitudinal  $v_\xi$  and transverse  $v_\eta$  micro-slip,  $v_C$  is the modified translational micro-slip that takes into account the effect of the spin-induced spin torque  $\phi$  on the arm of the semi-axis  $a$ , and the contact zone ellipse (in the longitudinal direction),  $F_{\eta S}$  is the transverse tangential force caused by the spin torque  $S$ . The algorithm [41] used to determine the contact forces requires the introduction of micro-slips  $v_\xi$ ,  $v_\eta$  and  $F$  normal contact force  $N$ , semi-axes  $a$  and  $b$  of the contact zone ellipse, combined transverse modulus of the wheel and the rail  $G$ , friction coefficient  $\mu$  and creep force Kalker's coefficients  $f_{11}$ ,  $f_{12}$ ,  $f_{22}$ ,  $f_{33}$  expressed in terms of equations (5) [20].

$$\begin{aligned} f_{11} &= (ab) Gc_{11}, & f_{23} &= (ab)^{3/2} Gc_{23}, \\ f_{22} &= (a/b)^2 Gc_{22}, & f_{33} &= (ab)^2 Gc_{33}, \end{aligned} \quad (5)$$

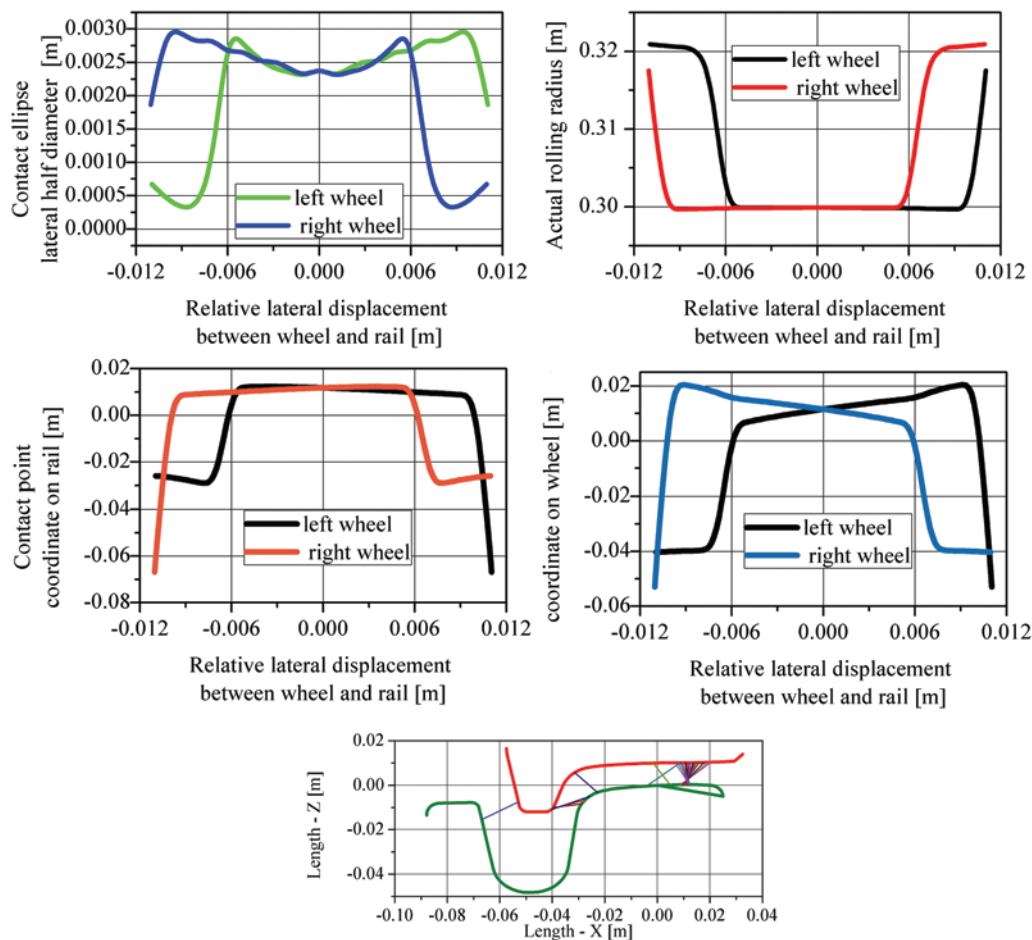


Fig. 6. Mutual positions of the wheel-rail profiles of the IRW tram with the PST wheel profile and the Ri60N rail profile with equivalent conicity of  $\lambda = 0.013$  and lateral displacement of 3 mm

Table 1. Kalker's linear rolling contact theory coefficients for  $\kappa = 0.25$  and  $a > b$  [21]

b/a	1	0.8	0.6	0.4	0.2	0.1
$c_{11}$	4.12	4.36	4.78	5.57	7.78	11.7
$c_{22}$	3.67	3.99	4.5	5.48	8.14	12.8
$c_{23}$	1.47	1.75	2.23	3.24	6.63	14.6
$c_{33}$	1.19	1.04	0.892	0.747	0.601	0.526

where:  $f_{11}$  is the lateral creep force coefficient,  $f_{12}$  denotes the lateral/spin creep force coefficient,  $f_{22}$  is the spin creep force coefficient,  $f_{33}$  describes the longitudinal creep force coefficient, and  $c_{ij}$  is the micro-slip coefficient [21] and the values of which are given in Table 1.

The above-described approach was chosen since the crank axis of the wheelset with independently rotating wheels may rotate relative to the vertical axis. It allows to determine the tangential forces in the wheel-rail contact zone, whereas the spin moment caused by the conicity of the wheel tread is taken into account.

For numerical tests using the adopted mathematical model of the vehicle, the nominal profile of the PST tram wheel was adopted. This profile is consistent with the standard [40]. According to technical guidelines included in the materials [46] grooved rails can be utilised in track curves with a radius under or equal to 150 m. Therefore, the Ri60N profile rail for the simulation was used. The positions of the wheel with the PST external profile in relation to the Ri60N rail was shown in Fig. 6.

### 3.3. Wheel drive control algorithms

Algorithms for controlling the operation of traction motors in trams equipped with IRW bogies have been the subject of many scientific publications [14, 3]. The control algorithm described in this paper was implemented to rail vehicle model by SIMAT module, which enables the co-simulation of MATLAB software, in which the algorithm was created, and the SIMPACK software, where the vehicle model was built. The calculation procedure is based on the fact that the angular velocities of independently rotating wheels of the vehicle's motor bogies are read in a single step while solving the equations of motion of the vehicle model. The angular velocities of the trailer bogie wheels are not processed due to the fact that they do not have any activators (motors) that could be controlled by the proposed algorithm. This is performed by the SIMAT module. This module transmits discrete values of the angular velocity of wheels  $\omega_i$  to MATLAB software, where they are introduced to the control algorithm. The above mentioned velocities are compared in this algorithm with the set velocity  $\omega_{set}$  of independently rotating wheels of the vehicle model. If a difference between their velocities occurs, i.e. when values of instantaneous velocities  $\omega_i$  are lower than the defined velocity, the driving torque  $T$  transmitted to the particular wheel is changed.

For the simulated cases, some variants of controlling the motor torque  $T$  were used. In the first variant the driving torque was reduced to a minimum value of 0. In the second variant of the algorithm, where the braking process was taken into account, a negative braking torque of a given wheel was applied. In case when the angular velocity of a particular wheel exceeded nominal pre-set value, a negative braking torque was initiated. Previously determined value of the variable during data exchange (co-simulation) is transferred via SIMAT module to the tram model in the SIMPACK program to the appropriate actuator of the drive in a particular motor bogie.

The algorithm is repeated in the next calculation step until the simulation end time is reached. The general diagram of the traction motor drive control system using the MATLAB co-simulation with SIMAT and SIMPACK modules is shown in Fig. 7.

The topology of traction motors in tram bogies has significant impact on the wheel-rail contact forces. In order to verify it and deter-

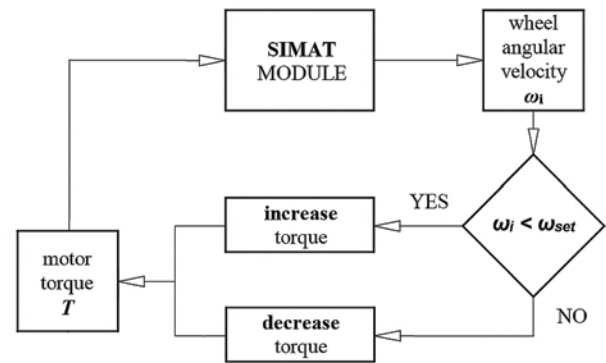


Fig 7. General wheel drive algorithm in bogies with IRW in the co-simulation software SIMPACK - MATLAB

mine the level of safety against derailment, three main configurations of torque distribution between traction motors of a given bogie were prepared. In the first configuration (marked as IRW-1), each of the four traction motors is controlled independently. The torque applied in one motor has no effect on the driving torque in other motors. In two subsequent configurations there is a group power supply of motors; for a group of two motors identical drive torques are applied. In this system, marked as IRW-2, the right or left group of wheels is supplied with the same torque. In the system marked as IRW-3, a group of wheels mounted on front or rear crank axles of a particular bogie is supplied.

Each control configuration was developed in two variants. The first one does not take into account wheel braking in the case when its angular velocity is higher than the defined velocity, while in the second it is taken into account. The simulation variants taking into account wheel braking have been marked with an additional "B" index.

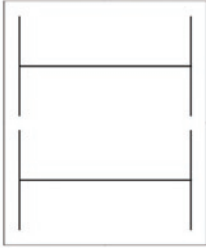
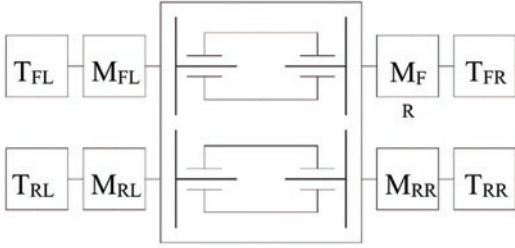
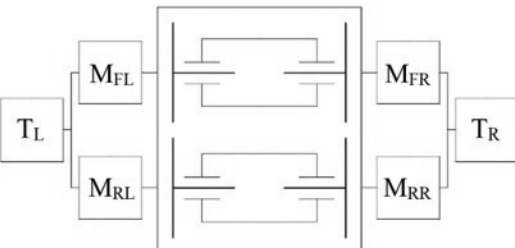
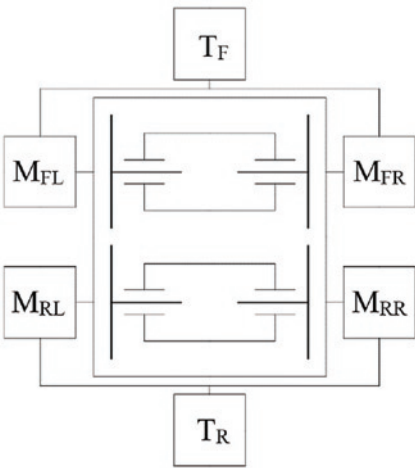
For a comparison with the classic wheelset, additional configuration of bogie with the same weight as the bogie with IRW was developed. The configuration - marked as "CWS" - is presented in Table 2 along with other configurations.

### 3.4. Description of testing methodology and simulation cases

To determine the level of safety against derailment of a railway vehicle it is necessary to measure the safety against derailment coefficient  $Y/Q$  (ratio of lateral force to vertical force of a particular wheel) and the leading wheel lift value  $D_z$ . Tests are performed on a track curve with a radius  $R = 150$  m with a twisted section of 3%, achieved by changing the height of the outer rail. Rails have the UIC60 profile. The guidelines for tests on safety against derailment of railway vehicles are set forth in standard EN 14363:2016 [10].

The validity of the safety against derailment test depends on the determined dry friction coefficient between the wheel and the rail. The test can be accepted if during a given run the measured friction coefficient is equal to or greater than 80% of the value of the previously determined dry friction coefficient  $\mu_{dry}$ . The  $\mu_{dry}$  coefficient depends on the value of the static vertical load of a particular wheel and on the attacking angle of a wheelset. Using the methodology from Annex A to EN 14363:2016 [10] and the above input data, the friction coefficient on dry rails  $\mu_{dry}$  could be determined.

Table 2. Layouts of various motor drive systems of IRW bogies

symbol	layout	description
CWS		classic wheelset in motor-bogies (for a comparison)
IRW-1 IRW-1B		each of the traction motors powered separately
IRW-2 IRW-2B		group power system for traction motors (right / left)
IRW-3 IRW-3B		group power system for traction motors (front / back)

Usually the most unfavourable load state of a vehicle subject to the safety against derailment tests is tare condition, when vertical loads on particular wheels are the lowest. However, if the vehicle is equipped with suspension with a nonlinear characteristic, its influence on the obtained results should be determined. In order to increase the differences between vertical wheel loads in particular wheels standard EN 14363:2016 [10] demands that additional twist in the primary suspension should be introduced. This could be achieved by using steel shims.

Fig. 8 displays a cross-section of a wheel with the S1002 profile in critical (but still safe) position on the UIC60 profile rail. Both the wheel and the rail are in new, unused condition. The wheel contacts the rail at a point on the flange which connects straight section with a curved one. For such combination the distance between the wheel

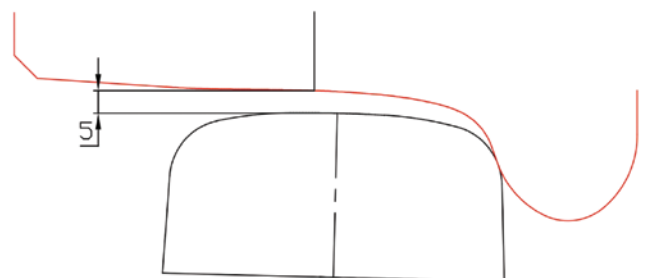


Fig. 8. Critical position of wheel profile S1002 on rail UIC 60 which defines the limit of wheel lift in safety against derailment test

tread at the point of the rolling circle and the rail surface is 5 mm. In standard EN 14363:2016 [10] it was adopted as the limit value for the wheel lift  $D_{z,lim}$  for the safety against derailment tests on a twisted track curve.

Methods described in standard EN 14363:2016 [10] cannot be applied to tram vehicles. Compared to rail vehicles, trams are characterized by a much greater variety in kinematic systems of connections between carriages and bogies. Tram vehicles mainly operate in cities and their infrastructure is adapted to local urban conditions, so in many cases it was necessary to plan track curves with a radius smaller than twisted track curve with a radius of  $R = 150$  m, required in standard [10]. Results of the safety against derailment tests of a tram on such curve may not be critical.

The above-mentioned issues contributed to the development of a proprietary methodology for testing safety against derailment of tram vehicles on curves of tracks with different radii, which was first presented in article [19]. Based on document [46], which is the main handbook for designing tram infrastructure in Poland, the radius of curves to be used for simulation tests and the cant values for each of these curves were determined. The guidelines of the document [46] also define the inclination of the superelevation ramp equal to 1:300.

The curves with radii of 18, 25, 50 and 150 m were selected as representative for the tram infrastructure for the safety against derailment tests. To ensure operational safety as the most critical case, the inclination of the superelevation ramp equal to 1:150 was assumed (less favourable than in the document [46]); it corresponds to a track twist of approx. 6.67‰. At the exit of each of the analysed curves additional twist was added - the outer rail over a section of 6 m changed its height by 20 mm, which corresponds to 3.33‰ twist. Such geometrical parameters of the track curve are shown in Fig. 9. The total twist on each curve equals to 10‰ and corresponds to poor quality track infrastructure awaiting periodic renovation.

Rubber-metal elements with specific force-displacement characteristics dominate in the suspension systems of tram bogies. Stiffness of these elements depends both on atmospheric conditions (the lower the ambient temperature, the higher the stiffness in relation to the stiffness

in the reference temperature) and manufacturing tolerance (usually +/- 15%). Therefore, it was decided to increase the nominal stiffness of all rubber-metal elements in the simulation model twice to ensure adequate level of safety against derailment also in conditions of low operating temperatures.

The nominal tram track gauge is 1435 mm, and the maximum - 1450 mm [46]. For the simulation purposes the range of rail gauge was extended to 1455 mm. In the above-mentioned range numerical simulations were carried out for cases of track gauge gradation every 5 mm. Assumed combinations of geometric parameters of the test tracks used for the proposed research method are presented in Table 3.

Table 3. Track geometry parameters for considered simulation cases

Curve radius [m]	Track twist on curve	Additional twist (outer rail)	Track gauge [m]
18,25,50,150	6.67‰	3.33‰	1.435; 1.440; 1.445; 1.450; 1.455

In order to reduce the length of the S-type track, and thus reduce the calculation time, the symmetry of the weight distribution to the right and left side of the tram was adopted. Thanks to it the level of safety against derailment could have been tested for only one direction of the track curve, i.e. the right one.

The key element of the developed research methodology for safety against derailment is the definition of a limit criterion for assessing the level of safety for particular tram. As already proved in the paper above, the decisive factor for the safety against derailment is the value of wheel lift  $D_z$  in relation to the top of rail. Wheel and rail profiles used for tram vehicles are completely different from those ones applied for railway vehicles. The outer profile of the rail was already mentioned before, i.e. Ri60N. Polish tram operators utilise trams mainly with two outer contours of wheels: T profile or PST. Their limit position on the Ri60N rail is shown in Fig. 10. The minimum wheel lift value achieved for T or PST profile combined with the Ri60N profile rail was  $D_z = 8.76$  mm. Rounding

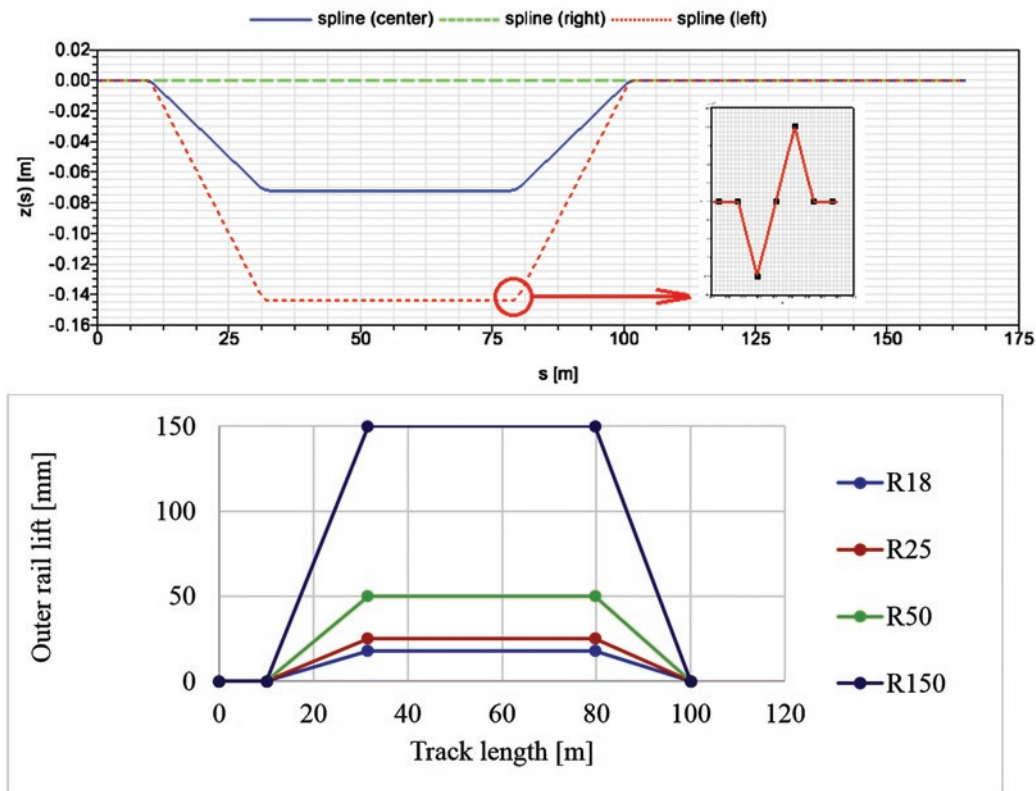


Fig. 9. A novel geometry of test track for proposed research methodology

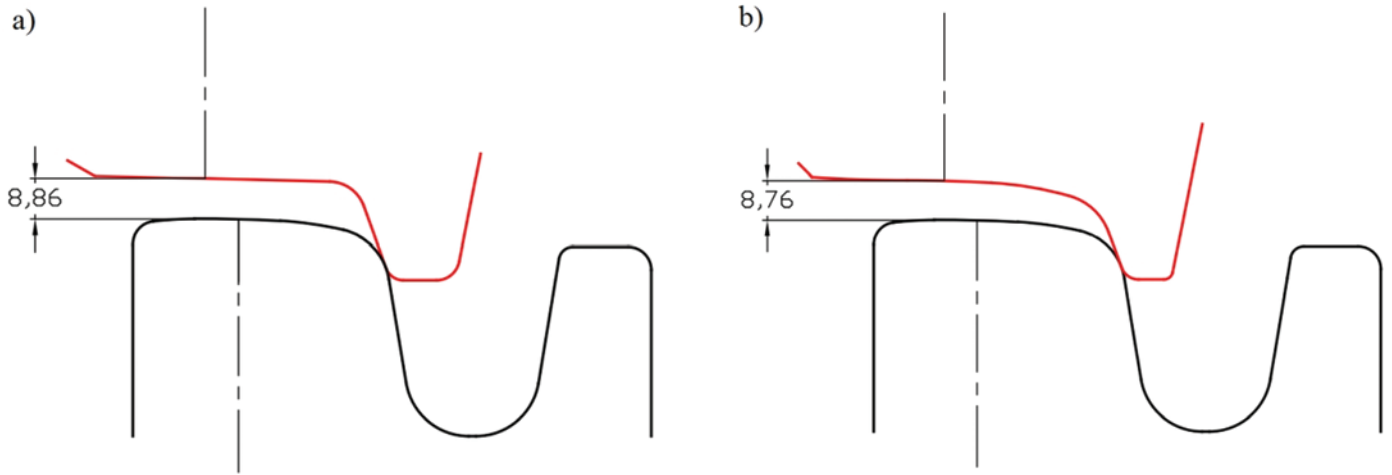


Fig. 10. Setting limit position of safety against derailment for wheels with profile T (a) or PST (b) with rail profile Ri60N

Table 4. List of parameters used for particular cases of developed test methodology

Steering algorithm	Curve radius [m]	Track gauge [m]
CWS, IRW-1, IRW-1B, IRW-2, IRW-2B, IRW-3, IRW-3B	18, 25, 50, 150	1.435; 1.440; 1.445; 1.450; 1.455

this value down to an integer, a higher level of safety against derailment can be obtained. Therefore, it is recommended that the value  $D_{z,lim} = 8 \text{ mm}$  should be set as the limit value when determining the safety margin against derailment of tram vehicles. As to analysis of trams from other countries, a separate analysis should be performed for the wheel and rail profiles used there.

Table 5. Charts of transverse wheel-rail Y contact forces for the leading wheel (front left) of the MB1 bogie in the right curve of the track with a track gauge of 1435 mm

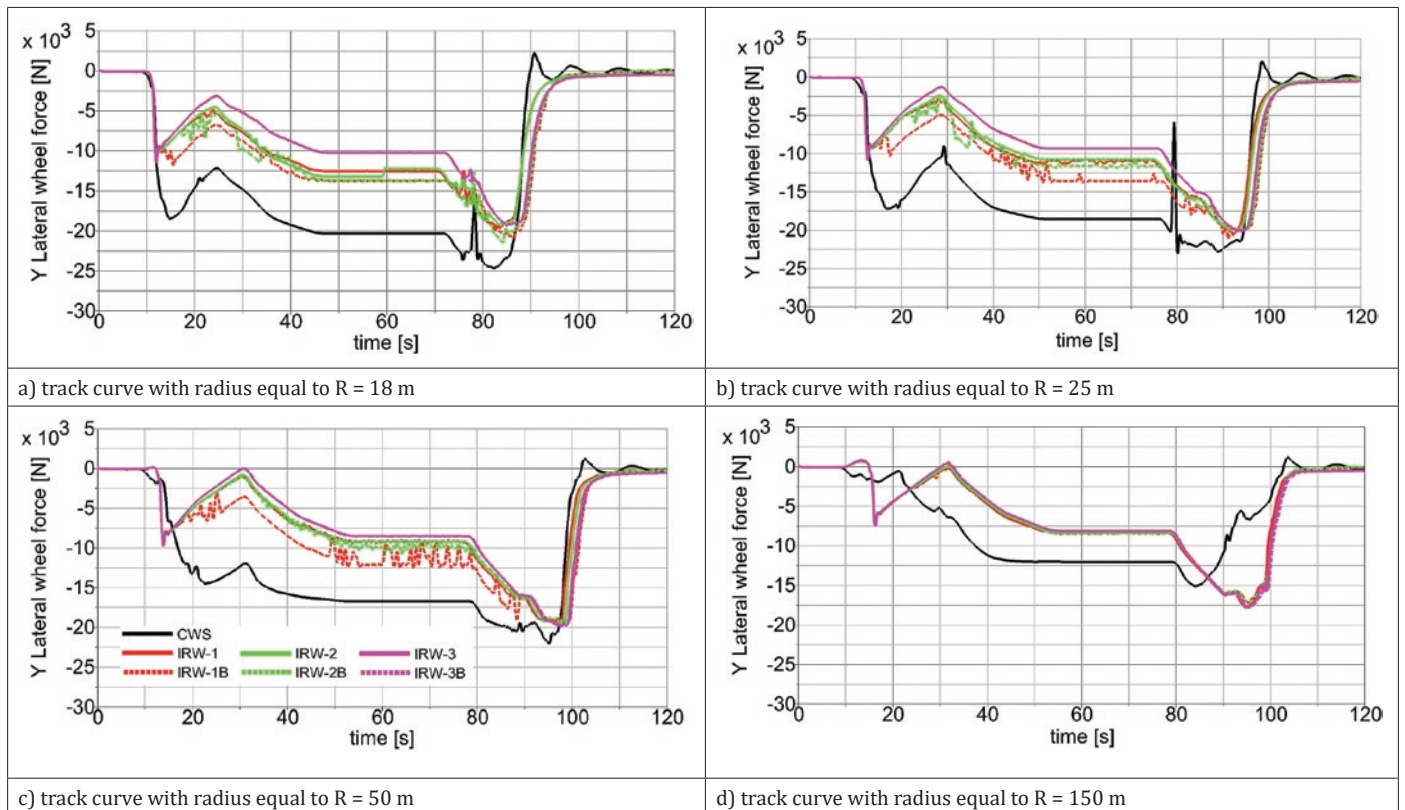


Table 4 includes parameters used for particular calculation cases, i.e. radius of curves, track gauges and variants of algorithms controlling the torque of traction motors. Results presented in the next chapter were obtained on the basis of simulation scenarios carried out with combinations of parameters from each of the columns in Table 4.

#### 4. Results of numerical study

The result values of the wheel lift  $D_z$  of the front wheelset of the MB1 bogie for the left (FL) and right (FR) wheels obtained on the basis of the test results, were presented separately for each of the control algorithms used as a function of the curve radius and track gauge. Graphical representation with top values for wheel lift  $D_z$  is displayed in Tables 6-8. Moreover, signals of transverse wheel-rail contact forces for the leading wheel (left front) of MB1 bogie in the track curve with the assumed track gauge of 1435 mm were compared and presented in the graphs in Table 5. For the rear wheels of the MB1 bogie, regardless of analysed loadcase and the type of the assumed

Table 6. Graphical representation of the wheel lift  $D_z$  for a classic wheelset CWS

Layout	Results [mm]											
CWS												
	R/s	1435	1440	1445	1450	1455	R/s	1435	1440	1445	1450	1455
	18	1.77	2.44	0	0	0	18	0	0	1.35	0	0
	25	3.76	4.71	0	0	0	25	0	0	0	0	0
	50	1.17	1.76	0	0	0	50	0	0	0	0	0
	150	0	0	0	0	0	150	0	0	0	0	0

drive control algorithm, their lift was equal to zero and therefore they are not presented in this summary. In the result Tables 6-8 the maximum values for a given wheel are marked in grey. In each of the numerical simulations loadcase the tram velocity was assumed as 1 m/s through the right curve of the track (Fig. 9) without a transition curve. In the geometric shape of the track, twist was assumed with a total value of 10%, reflecting bad condition of maintenance. Based on the obtained results, the level of safety against derailment was defined by determining the vehicle wheel lift  $D_z$  during track curve negotiation. In all numerical simulations the vehicles had the same mass, including the movement simulation of bogies with standard wheelsets (CWS). Thanks to it an analogous signal of changes of vertical loads of tram wheels for an identical track system (regardless of the adopted control algorithm) could have been obtained.

For track gauge up to 1445 mm the front left wheel was mainly responsible for guiding bogie in the track. For larger track gauge values the lateral guidance in the track was taken over by the wheel on the inside of the curve, i.e. the right front wheel. This is reflected in the increase of Y/Q ratio for this wheel and for all analysed curve radii, as well as in the decrease of Y/Q ratio for the left front wheel. The reason for the above was the shape of the outer rail profile and rotation of bogie around its axis because, as confirmed by the papers [5, 27], the Ri60N profile applied in these tram simulation scenarios is grooved.

Lateral force applied to the guide wheel (known as the “guiding force” of rail vehicle wheel) in a standard wheelset is often greater than in an IRW bogie during entering and passing the curve, regardless of algorithm used. As the curve radius increases, the difference of guiding force value for standard wheelset loadcase and the loadcase of IRW bogie decreases. In a curve with radius R150 guiding forces for a bogie with IRW in the moment of entering and exiting the curve are greater than in the loadcase of a bogie with standard wheelset. No differences in the guiding force value resulting from the application of algorithms for controlling the operation of motor torque distribution in motor bogies with IRW can be found in curves up to R50 radius. Larger values can be achieved for IRW-1B and IRW-2B systems with motor braking. For IRW-3 and IRW-3B systems (corresponding to the group supply of the front-rear wheels drive), the force waveforms for each of the analysed curves were almost identical. As to curve with radius R150 no significant differences in the guiding force values can be identified, regardless of the applied torque control algorithm.

For all standard wheelset simulation loadcases the lowest values of wheel lift were obtained. The greatest wheel lift of the left front wheel (FL), equal to 4.71 mm, was achieved on a curve with a radius

R25 and a track gauge of 1440 mm, and the highest value for the right front wheel, equal to 1.35 mm, was obtained for a curve with a radius R18 and track width of 1445 mm. The results for the MB1 bogie from IRW turned out to be different. In all analysed loadcases, the highest values of the left front wheel (FL) lift were achieved on R50 curve with a track gauge of 1440 mm. The maximum value, i.e. 9.52 mm was obtained by applying the IRW-3 algorithm. The right front wheel (FR) reached maximum lift on R18 curve with a track gauge in the range of 1450-1455 mm. The highest value was achieved by using the IRW-2B algorithm. All values obtained for the front right wheel (FR) concerning the wheel lift on the rail groove matched the assumed limit  $D_z \leq 8$  mm. The application of motor braking in the IRW control systems resulted in the reduction of maximum values of the left front wheel lift (running in the right curve) from 9.25 mm to 7.33 mm, from 9.50 mm to 9.32 mm, and from 9.52 mm to 7.82, corresponding with the variant of algorithm IRW-1, IRW-2 and IRW-3.

## 5. Summary

Based on numerical tests of the light rail vehicle model with independently rotating wheels, very high sensitivity of trajectory of this vehicle motion caused by the method of controlling the vehicle’s wheel drive has been identified.

According to the results of papers [12, 44] describing the dynamics of rail vehicles with independently rotating wheels, the effect of unstable bogie in the central position of the track was observed. Such cases occurred only in the absence of wheel drive control, as opposed to correct vehicle stabilization in the axis of track for bogies with standard wheelsets [17, 42]. This instability effect influences the value of contact forces between wheel flanges and rail head adversely. Such phenomena deteriorate technical conditions of operation and increase maintenance costs of the rolling stock [6, 38, 49].

It should be also pointed out that the drive control system with the option of braking wheels with the motor - proposed by the authors - allows to reduce angular velocity differences between individual wheels. An induced speed reduction of a given wheel increases the forces of lateral interaction with the track. This, consequently, leads to a slight growth of the lateral force Y relative to control systems without the braking option. Bearing this in mind, a conclusion can be drawn that this phenomenon reduces wheel lift and improves both operating conditions and driving safety. This is one of the ways in designing an effective stabilizing control. Authors proved that, regardless of selected drive and braking control system, the determined

Table 7. Graphical representation of the wheel lift Dz with various control algorithms of IRW without wheels motor braking

Layout	Results [mm]											
IRW-1												
	R/s	1435	1440	1445	1450	1455	R/s	1435	1440	1445	1450	1455
	18	0	0	0	0	0	18	0	0	1.75	4.38	4.05
	25	3.44	4.35	0	0	0	25	0	0	0	0	0
	50	8.82	9.25	0	0	0	50	0	0	0	0	0
	150	1.37	1.25	0	0	0	150	0	0	0	0	0
IRW-2												
	R/s	1435	1440	1445	1450	1455	R/s	1435	1440	1445	1450	1455
	18	0	0	0	0	0	18	0	0	1.77	4.48	3.77
	25	4.02	4.65	0	0	0	25	0	0	0	0	0
	50	9.41	9.50	0	0	0	50	0	0	0	0	0
	150	1.37	1.26	0	0	0	150	0	0	0	0	0
IRW-3												
	R/s	1435	1440	1445	1450	1455	R/s	1435	1440	1445	1450	1455
	18	0	0	0	0	0	18	0	0	1.88	2.72	4.13
	25	4.18	4.83	0	0	0	25	0	0	0	0	0
	50	9.34	9.52	0	0	0	50	0	0	0	0	0
	150	1.37	1.25	0	0	0	150	0	0	0	0	0

values of guiding forces Y of tram wheels with IRW for all analysed radii curves of the track were smaller than in the similar vehicle with standard wheelsets. The greater the difference in the value of leading force between the analysed vehicles (standard wheelset and IRW), the smaller the curve radius of the track used for simulations. This proves

the advisability of further development of the IRW bogie structure which allows for the reduction of wheel-rail transverse impacts on the urban track infrastructure, where many curves have radius smaller than  $R = 150$  m (e.g. on tram loops). Additional positive aspect of



Table 8. Graphical representation of the wheel lift Dz with various control algorithms of IRW with wheels motor braking

Layout	Results [mm]											
IRW-1B												
	R/s	1435	1440	1445	1450	1455	R/s	1435	1440	1445	1450	1455
	18	0	0	0	0	0	18	0	0	1.74	2.78	4.13
	25	4.43	4.66	0	0	0	25	0	0	0	0	0
	50	6.96	7.33	0	0	0	50	0	0	0	0	0
	150	0	0	0	0	0	150	0	0	0	0	0
IRW-2B												
	R/s	1435	1440	1445	1450	1455	R/s	1435	1440	1445	1450	1455
	18	0	0	0	0	0	18	0	0	1.74	4.45	4.29
	25	4.68	5.24	0	0	0	25	0	0	0	0	0
	50	9.11	9.32	0	0	0	50	0	0	0	0	0
	150	0	0	0	0	0	150	0	0	0	0	0
IRW-3B												
	R/s	1435	1440	1445	1450	1455	R/s	1435	1440	1445	1450	1455
	18	0	0	0	0	0	18	0	0	1.80	2.71	4.06
	25	3.77	4.54	0	0	0	25	0	0	0	0	0
	50	7.60	7.82	0	0	0	50	0	0	0	0	0
	150	1.03	0	0	0	0	150	0	0	0	0	0

the lower lateral force Y in the wheel-rail contact is the reduction of noise level.

The analysis of the obtained results showed that vehicles with independently rotating wheels should be carefully designed along with the drive control system in order to minimize the lift and wear of wheels.

The introduction of bogies with independently rotating wheels of the IRW type to tram vehicles is a step forward in the development of the design of modern chassis of such vehicles. They ensure greater driving comfort for passengers and, thanks to proper drive control

systems, forces of the lateral interaction of the wheel with the rail and can be reduced, which consequently, slows down the track wear.

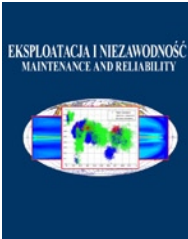
The study results presented in this paper may be helpful in the processes of designing tram vehicles and their drive control systems. They can also be useful for their users and operators who are respon-

sible for the reliability and maintenance of the rolling stock of tram vehicles with independently rotating wheels. This is of particular importance since the number of variants of tram structures with the drive system under consideration will grow in the near future due to their operational advantages.

## References

1. Bogacz R, Konowrocki R. On new effects of wheel-rail interaction. *Archive of Applied Mechanics* 2012; 82: 1313-1323, <https://doi.org/10.1007/s00419-012-0677-6>.
2. Cho Y. Verification of control algorithm for improving the lateral restoration performance of an independently rotating wheel type railway vehicle. *International Journal of Precision Engineering and Manufacturing* 2020; 21: 247-1258, <https://doi.org/10.1007/s12541-020-00346-4>.
3. Chudzikiewicz A, Firlirk B. Light rail vehicle dynamics from a running safety perspective *Archives of Transport* 2009; 21: 39-49.
4. Chudzikiewicz A, Sowińska M, Krzyżyński T, Maciejewski I. Modeling of wheel set movement with independently rotating wheels including the wheel control system. *Logistyka* 2015; 4: 32-139.
5. Chudzikiewicz A, Sowiński B. Modelling and simulation of trams bogies with fully independently rotating wheels. *Dynamics of vehicles on roads and tracks* 2016; 1427-1434, <https://doi.org/10.1201/b21185-151>.
6. Chudzikiewicz A, Korzeb J. Simulation study of wheels wear in low-floor tram with independently rotating wheels. *Archive of Applied Mechanics* 2018; 88: 75-192, <https://doi.org/10.1007/s00419-017-1301-6>.
7. Chudzikiewicz A, Sowińska M. Modelling and simulations of dynamics of the low-floor tramcar with independently rotating wheels. *Communications - Scientific Letters of the University of Zilina* 2017; 17 (4): 45-52.
8. Goodall RM and Mei TX. Mechatronic strategies for controlling railway wheelsets with independently rotating wheels, in *Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2005; 1: 225-230, <https://doi.org/10.1109/AIM.2001.936458>.
9. Guerrieri M. Tramways in urban areas an overview on safety at road intersections. *Urban Rail Transit* 2018; 4: 223-233, <https://doi.org/10.1007/s40864-018-0093-5>.
10. EN 14363:2016 - Railway applications - Testing and Simulation for the acceptance of running characteristics of railway vehicles - Running behaviour and stationary tests.
11. Firlirk B, Staśkiewicz T, Jaśkowski W, Wittenbeck L. Optimisation of a tram wheel profile using a biologically inspired algorithm *Wear*, 2019; 430-431: 12-24, <https://doi.org/10.1016/j.wear.2019.04.012>.
12. Hauser V, Nozhenko O, Kravchenko K, Loulová M, Gerlici J, Lack T. Proposal of a steering mechanism for tram bogie with three axle boxes. *Procedia Engineering* 2017; 192 (41): 289-294, <https://doi.org/10.1016/j.proeng.2017.06.050>.
13. Gill A, Firlirk B, Kobaszynska-Twardowska A. Analysis of the Distribution of Passengers inside a Tram, in J. Pombo, (Editor), *Proceedings of the Third International Conference on Railway Technology: Research, Development and Maintenance*, Civil-Comp Press, Stirlingshire, UK, Paper 303, 2016.
14. Grether G, Looye G, Heckmann A. Lateral guidance of independently rotating wheel pairs using feedback linearization, *Proceedings of the Fourth International Conference on Railway Technology: Research, Development and Maintenance*, Spain, 2018.
15. Hoshi M, Ookubo Y, Murakami N, Arai T, Kono H, Aruga H. Development of bogie for user friendly, extra low floor, light rail vehicle (LRV) using independent wheel system and next generation LRV, Mitsubishi Heavy Industries Ltd., *Technical Review* 2007; 44 (2): 1-4. <https://lucchinirs.com/wp-content/uploads/2017/05/limoset-1.jpg> (access on 07.07.2020).
16. Iwnicki S. *Handbook of railway vehicle dynamics*. CRC Press. 2006, <https://doi.org/10.1201/9781420004892>.
17. Ji Y, Ren L, Zhou J. Boundary conditions of active steering control of independent rotating wheelset based on hub motor and wheel rotating speed difference feedback. *Vehicle System Dynamics* 2018; 56 (12): 1883-1898, <https://doi.org/10.1080/00423114.2018.1437273>.
18. Kalinowski D, Szolc T, Konowrocki R. The new simulation approach of tramway safety against derailment evaluation in term of vehicle dynamics. *Transbaltica XI: Transportation Science and Technology*, Vilnius (LT), 2020: 245-254, [https://doi.org/10.1007/978-3-030-38666-5\\_26](https://doi.org/10.1007/978-3-030-38666-5_26).
19. Kalker JJ. Review of wheel - rail rolling contact theories. The general problem of rolling contact. Eds. A.L. Browne and N.T. Tsai. Vol. 40. New York: American Society of Mechanical Engineers, 1980.
20. Kalker JJ. A strip theory for rolling contact of two elastic bodies in presence of dry friction. Ph. D. dissertation, Delft University of Technology, 1976.
21. Klamka J, Grzesiak R. Drive bogie for low-floor rail vehicle (Wózek napędowy do niskopodłogowego pojazdu szynowego), Patent PL225456, 28.04.2017 WUP 04/17.
22. Kolar J. Modern trends in the drive wheelsets of rail vehicles. In: Dinybyl V., Berka O., Petr K., Lopot F., Dub M. (eds) *The Latest Methods of Construction Design*. Springer, Cham, 2016: 27-35, [https://doi.org/10.1007/978-3-319-22762-7\\_5](https://doi.org/10.1007/978-3-319-22762-7_5).
23. Konowrocki R, Bajer Cz. Friction rolling with lateral slip in rail vehicles. *Journal of Theoretical and Applied Mechanics* 2009; 47 (2): 275-293.
24. Konowrocki R, Bogacz R. Tram vehicles with independently rotating wheels, chapter: Overview of light rail vehicle solutions used in operation, Book Warsaw University of Technology Publishing House 2019: 148-164, ISBN: 978-83-8156-010-8 (in Polish).
25. Konowrocki R, Chojnacki A. Analysis of rail vehicles' operational reliability in the aspect of safety against derailment based on various methods of determining the assessment criterion *Eksploatacja i Niezawodność - Maintenance and Reliability* 2020; 22 (1): 73-85, <https://doi.org/10.17531/ein.2020.1.9>.
26. Konowrocki R. Tram vehicles with independently rotating wheels, chapter: Simulations of tram vehicle dynamics with respect to different operating conditions, Warsaw University of Technology Publishing House 2019: 297-328, ISBN: 978-83-8156-010-8 (in Polish).
27. Liang B, Iwnicki S D. An experimental study of independently rotating wheels for railway vehicles. *International Conference on Mechatronics and Automation*, Harbin, 2007: 2282-2286, <https://doi.org/10.1109/ICMA.2007.4303908>.
28. Liang B, Iwnicki S D. Independently rotating wheels with induction motors for high-speed trains. *Journal of Control Science and Engineering* 2011: 1-7, <https://doi.org/10.1155/2011/968286>.

30. Li H, Chi M, Liang S, Wu X. Curve-passing performance of independent rotating wheels based on differential feedback. *Journal of Vibration and Shock* 2018; 37 (23): 126-132, <https://doi.org/10.13465/j.cnki.jvs.2018.23.018>.
31. Lindblom J. Trackbound vehicle with steering of wheel axles, Patent EP0939717A1, 1999.
32. Lu ZG, Sun XJ, Yang JQ. Integrated active control of independently rotating wheels on rail vehicles via observers. *Journal of Rail and Rapid Transit* 2017; 231 (3): 295-305, <https://doi.org/10.1177/0954409716629705>.
33. Macián V, Tormos B, Herrero J. Maintenance management balanced scorecard approach for urban transport fleets. *Eksplatacja i Niezawodność - Maintenance and Reliability*, 2019, 21 (2): 226-236, <https://doi.org/10.17531/ein.2019.2.6>.
34. Matej J, Seńko J, Awrejcewicz J. Dynamic properties of two-axle freight wagon with uic double-link suspension as a non-smooth system with dry friction. In: Awrejcewicz J. (eds) *Applied Non-Linear Dynamical Systems. Springer Proceedings in Mathematics & Statistics* 2014; 93: 255-268, [https://doi.org/10.1007/978-3-319-08266-0\\_18](https://doi.org/10.1007/978-3-319-08266-0_18).
35. Mei T X, Goodall R M. Practical strategies for controlling railway wheelsets independently rotating wheels. *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, 2003; 125 (3): 354-360, <https://doi.org/10.1115/1.1592191>.
36. Michajłow M, Jankowski Ł, Szolc T, Konowrocki R. Semi-active reduction of vibrations in the mechanical system driven by an electric motor. *Optimal Control Applications & Methods* 2017; 38 (6): 922-933, <https://doi.org/10.1002/oca.2297>.
37. Myamlin S, Kirilchuk O, Metyzhenko V. Mathematical model of wheelset oscillations with independent wheel rotation in the horizontal plane. *Science and Transport Progress* 2016; 4 (64): 134-141, <https://doi.org/10.15802/stp2016/77999>.
38. Opala M. Study of the derailment safety index Y/Q of the low-floor tram bogies with different types of guidance of independently rotating wheels. *Archives of Transport* 2016; 38 (2): 39-47, <https://doi.org/10.5604/08669546.1218792>.
39. Opala M. Evaluation of bogie centre bowl friction models in the context of safety against derailment simulation predictions. *Archive of Applied Mechanics* 2018; 88 (6): 943-953, <https://doi.org/10.1007/s00419-018-1351-4>.
40. PN-K-92016:1997 - Tram wheel sets, flexible - Surfaced tyres - Requirements and testings.
41. Polach O. A fast wheel-rail forces calculation computer code. *Vehicle System Dynamics* 1999; 33: 728-739, <https://doi.org/10.1080/00423114.1999.12063125>.
42. Shabana AA, Zaazaa KE, Sugiyama H. *Railroad vehicle dynamics : a computational approach*. CRC Press 2008, <https://doi.org/10.1201/9781420045857>.
43. Staśkiewicz T, Firlík B. Verification of a tram wheel new profile dynamic behavior. *Journal of Mechanical and Transport Engineering* 2017; 69 (1): 49-60, <https://doi.org/10.21008/j.2449-920X.2017.69.1.05>.
44. Suda Y, Wang W, Nishina M, Lin S, Michitsuji Y. Self-steering ability of the proposed new concept of independently rotating wheels using inverse tread conicity. *Vehicle System Dynamics* 2012; 50: 291-302, <https://doi.org/10.1080/00423114.2012.672749>.
45. Tabaszewski M, Firlík B. Assessment of the track condition using the Gray Relational Analysis method. *Eksplatacja i Niezawodność - Maintenance and Reliability* 2018; 20 (1): 147-152, <https://doi.org/10.17531/ein.2018.1.19>.
46. Technical guidelines for the design and maintenance of tram tracks, Ministry of Administration, Local Economy and Environmental Protection, Department of Public Transport and Roads, Warsaw 1983.
47. Transport in the European Union - current trends and issues, Report European Commission, Directorate-General Mobility and Transport, B-1049 Brussels, March 2019.
48. UIC 518, Testing and approval of railway vehicles from the point of view of their dynamic behaviour - Safety - Track fatigue - Ride quality 2009.
49. Wijata X, Awrejcewicz A, Matej J, Makowski M. Mathematical model for two-dimensional dry friction modified by dither. *Mathematics and Mechanics of Solids* 2017; 2 (10): 1936-1949, <https://doi.org/10.1177/1081286516650483>.
50. Zirek A, Voltr P, Lata M. Validation of an anti-slip control method based on the angular acceleration of a wheel on a roller rig. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 2020; 234 (9): 1029-1040, <https://doi.org/10.1177/0954409719881085>.



Article citation info:

Xu Z, Guo D, Wang J, Li X, Ge D. A numerical simulation method for a repairable dynamic fault tree. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 34–41, <http://dx.doi.org/10.17531/ein.2021.1.4>.

## A numerical simulation method for a repairable dynamic fault tree

Indexed by:



Zhixin Xu<sup>a</sup>, Dingqing Guo<sup>b,a</sup>, Jinkai Wang<sup>a</sup>, Xueli Li<sup>c,d</sup>, Daochuan Ge<sup>c\*</sup>

<sup>a</sup>State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment, China Nuclear Power Engineering Co., Ltd., Shenzhen 518172, China

<sup>b</sup>School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>c</sup>Key Laboratory of Neutronics and Radiation Safety, Institute of Nuclear Energy Safety Technology, HFIPS, Chinese Academy of Sciences, Hefei, Anhui 230031, China

<sup>d</sup>University of Science and Technology of China, Hefei, Anhui 230026, China

### Highlights

- The adapted sequential failure region is developed to characterize the failure mechanism of a minimal cut sequence.
- The proposed approach is applicable for nonexponential distribution situations.
- The proposed approach is more efficient than the Markov chain state space methods.

### Abstract

Dynamic fault trees are important tools for modeling systems with sequence failure behaviors. The Markov chain state space method is the only analytical approach for a repairable dynamic fault tree (DFT). However, this method suffers from state space explosion, and is not suitable for analyzing a large scale repairable DFT. Furthermore, the Markov chain state space method requires the components' time-to-failure to follow exponential distributions, which limits its application. In this study, motivated to efficiently analyze a repairable DFT, a Monte Carlo simulation method based on the coupling of minimal cut sequence set (MCSS) and its sequential failure region (SFR) is proposed. To validate the proposed method, a numerical case was studied. The results demonstrated that our proposed approach was more efficient than other methods and applicable for repairable DFTs with arbitrary time-to-failure distributed components. In contrast to the Markov chain state space method, the proposed method is straightforward, simple and efficient.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

repairable dynamic fault tree, numerical simulation, Monte Carlo, sequential failure region, minimal cut sequence set.

## 1. Introduction

Dynamic fault trees are extended from the traditional static fault trees (SFTs) by integrating several dynamic logic gates, such as Warm Spare (WSP) gate, Priority AND (PAND) gate, and Function Dependent (PDEP) gate. With the help of integrating these dynamic gates, DFTs can model industrial systems with sequential failure behaviors that are not permitted in SFTs. Currently, dynamic fault trees are successfully applied to system safety design, reliability evaluation, and risk management [14, 26, 32]. During the past few years, researchers have done much work in this field, making fruitful achievements [6, 16, 18]. However, it is not easy to quantify a repairable DFT when applying the Markov chain state space method, because the executive process is time consuming and error prone, especially for a large-scale repairable DFT.

The primary analyzing techniques for quantifying a DFT are divided into three main categories: Markov chain state space methods [1, 7, 24, 29], combinatorial methods [21, 23, 31, 34], and numerical simulation methods [8, 25, 35]. Markov chain-based and combinatorial approaches are analytical methods that can provide exact solutions. Markov chain-based methods are not only applicable for

a nonrepairable DFT but are also applicable for a repairable DFT. However, Markov chain state space methods are vulnerable to state space explosion when analyzing large-scale DFTs. In addition, Markov chain-based methods require the components' time-to-failure to follow exponent distributions. By contrast, combinatorial approaches (such as inclusion exclusion principle (IEP) methods [19], sequential binary decision diagram (SBDD) methods [30, 31], improved SBDD methods [11], dynamic binary decision tree (DBDT) methods [13], and adapted K.D.Heidtmann methods [12]) are seldom trapped into state space explosion and often more efficient than the Markov chain methods. Nevertheless, most existing combinatorial approaches are limited in solving nonrepairable DFTs. It is worth noting that some researchers have tried to develop a kind of combinatorial method to solve a repairable DFT with PAND gates under a steady state [33]. Hence, for a repairable DFT, combinatorial approaches need to be further studied and improved.

Numerical simulation analyzing techniques are also commonly used to deal with DFTs, such as Monte Carlo (MC) numerical simulation [4, 5, 27]. Compared with the analytical approaches, numerical simulation methods can either provide great generalities on a DFT structure and failure distribution of their input events or reduce the

(\*) Corresponding author.

E-mail addresses: Z. Xu - [xuzhixin@cgnpc.com.cn](mailto:xuzhixin@cgnpc.com.cn), J. Wang - [wangjinkai@cgnpc.com.cn](mailto:wangjinkai@cgnpc.com.cn), D. Guo - [guodingqing@cgnpc.com.cn](mailto:guodingqing@cgnpc.com.cn), X. Li - [1328931871@qq.com](mailto:1328931871@qq.com), D. Ge - [daochuan.ge@inest.cas.cn](mailto:daochuan.ge@inest.cas.cn)

scale of the problem to be handled. Generally, simulation methods are more versatile than analytical approaches, especially for probability density functions (PDFs) of input events' time-to-failure, which are quite complex and lack explicit primitive functions. Rao KD et al. ever applied the Monte Carlo simulation technique to analyze repairable DFTs based on failure logics of various dynamic logic gates [25], and calculated the reliability index (i.e., unavailability) of a given system. However, the state of a DFT's top event could not be determined until all dynamic gates' logic states were simulated, which meant more simulating time might be needed due to the redundant logic terms. Zhang P et al. then used a similar Monte Carlo simulation technique to evaluate the reliability of a Phasor Measurement Unit [35]. Ge D et al. ever proposed a Monte Carlo simulation approach based on the coupling of DFTs' minimal cut sequence set and sequence failure regions (SFR) to analyze a nonrepairable DFT, but this method was not extended to repairable DFTs [10]. Merle G et al. developed a Monte Carlo simulation method based on DFTs' structure functions, but this method is only applicable for nonrepairable DFTs [22]. DFTsim [3] and MatCarloRE [20] are two analyzing tools for DFTs, and both tools use Monte Carlo simulation for solving DFTs, but do not allow repairable basic events. Recently, Gascard E et al. proposed an event-driven Monte Carlo simulation approach for quantitative analysis of DFTs [9], but the authors also assumed that the basic events are nonrepairable.

As mentioned above, for repairable DFTs, the accessible analyzing tools are Markov chain state space-based methods and dynamic logic gates-based Monte Carlo numerical simulation methods. For a large-scale repairable DFT, the feasible methods are Monte Carlo numerical simulation approaches. However, the existing Monte Carlo numerical simulation methods for repairable DFTs are dependent on dynamic logic gates' failure logics, which means more simulation time might be needed due to redundant logic terms. In this study, an MCSS-based Monte Carlo simulation method that couples the DFTs' minimal cut sequence set (MCSS) and sequence failure regions is proposed, which can be the main research contribution. Compared with existing methods, the merits of our proposed method are: 1) in contrast to the Markov chain state space method, the proposed numerical simulation method is versatile and not limited to particular distribution types; 2) it can provide more reliability indices for a concerned system, such as uncertainty of system reliability and component importance; and 3) by comparison with dynamic logic gates-based numerical simulation methods, the proposed method can reduce the unnecessary redundant logic terms based on minimal cut sequence set and hence improve computing efficiency.

The remainder of this paper is organized as follows. The concepts of dynamic logic gates and repairable DFTs are clarified in section 2. In section 3, the proposed MCSS-based Monte Carlo numerical simulation method is provided. Numerical case study is chosen and implemented to demonstrate the reasonability of the proposed method in section 4. Section 5 is devoted to the final conclusion.

## 2. Dynamic fault trees

### 2.1. Dynamic logic gates and repairable dynamic fault trees

To capture the sequence failure behaviors in the industrial systems, researchers have developed several dynamic logic gates such as Function Dependent (FDEP) gates, Priority AND (PAND) gates, Sequence Enforcing (SEQ) gates, and spare gates including Cold Spare (CSP) gates, Warm Spare (WSP) gates, and Hot Spare (HSP) gates, as shown in Fig. 1. A FDEP gate (Fig. 1 (a)) has a single trigger event (basic event or the output of another gate) and several dependent basic events. It characterizes a situation where the failure of the trigger event would cause all dependent basic events to fail, yet failure of any dependent basic event does not have effects on the trigger event. The PAND gate in Fig. 1 (b) is a special case of the AND gate. The PAND gate fires if its input events fail in a left-to-right order. A SEQ gate

(Fig.1 (c)) has only one failure order (i.e., from left to right), and only when all the input events fail can the SEQ gate occur. For a spare gate, it often has one primary event and some spare events. Only when the primary event fails can spare events start to replace the primary one. As all input events under a spare gate lose, the spare gate fires. Specifically, the CSP gate in Fig. 1 (d) allows modeling of the case where cold spares always stay at an unpowered state when the primary event functions. This means the primary event,  $e_1$ , must fail first; then the first cold spare  $e_2$  fails; and finally the last one  $e_n$  fails. The WSP in Fig. 1 (e) is unlike CSP gates, the spares stay at a reduced power when the primary event is normal. That is, the input events under a WSP can fail in any sequence. With regard to the HSP in Fig. 1 (f), the spares stay at full power when the primary event operates normally. Hence, the failure logic for an HSP is equivalent to the AND gate.

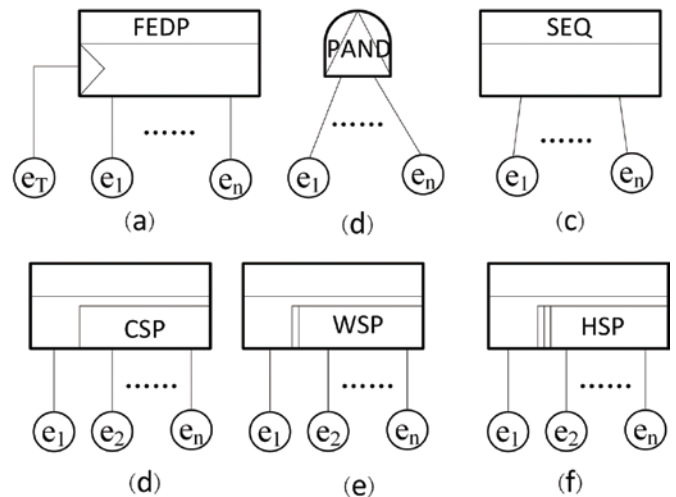


Fig. 1. Dynamic logic gates

To evaluate the reliability of systems with sequential failure behaviors, DFTs are proposed and developed. DFTs are defined by researchers as static fault trees integrating at least one dynamic logic gate. According to input events, regardless of whether or not they have reparability behaviors, DFTs can be classified into two categories: nonrepairable and repairable DFTs. A non-repairable DFT is defined as a DFT whose input events do not have any reparability behavior. A repairable DFT is defined as having input basic events with reparability behaviors.

### 2.2. Failure logic expressions of a repairable DFT

As mentioned above, the occurrence of a DFT's top event not only depends on combinations of its basic events but also depends on their failing orders. Therefore, the minimal cut set used to express failure behaviors in traditional static fault trees is not available. To settle this problem, Tang et al. developed the concept of minimal cut sequence for DFT analysis [28]. The minimal cut sequence (MCS) is defined as the minimal failure order that leads to occurrence of the top event of a DFT, and all the minimal cut sequences can form a universal set (i.e., minimal cut sequence set (MCSS)). The MCSS can be applied to capture the complete failure information in a DFT. In this contribution, MCSS is applied to characterize the failure logic expressions (FLE) of a DFT. Suppose a DFT has a MCSS with  $m$  minimal cut sequences, then the failure logic expression  $FLE_{dft}$  of this DFT can be written as:

$$FLE_{dft} = MCS_1 + MCS_2 + \dots + MCS_m \quad (1)$$

To explicitly formulate an MCS, some special symbols are introduced. We use the symbol " $\rightarrow$ " to represent sequential failure, which means the left basic event fails before the right one. It is defined as:

$$\sigma(a \rightarrow b) = \begin{cases} 1, & 0 \leq t(a) \leq t(b) \leq T \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where  $\sigma(\cdot)$  represents the state of a considered basic event or a sequential failure event, "1" denotes the failure state, "0" denotes the normal state,  $t(a)$  and  $t(b)$  indicates the failure time of  $a$  and  $b$ , and  $T$  is the mission time.

It should be noted that the symbol " $\rightarrow$ " just reflects the order of time-to-failure of the components represented by basic events. In fact, the start times of some components are also sequence dependent, such as cold spares, warm spares, and even the basic events under a SEQ gate. To characterize the sequence of the start time of some components, in this work, the special symbols  $A$ ,  ${}^0_A B$ ,  ${}^\beta_A B$ , and  ${}^1_A B$  are also introduced, where  $A$  denotes a general basic event,  ${}^0_A B$  represents  $B$  is a cold spare of  $A$  or any one of the second and subsequent input events under a SEQ gate,  ${}^\beta_A B$  indicates  $B$  as a warm spare of  $A$  that fails before  $A$ ,  $\beta$  is the a dormant factor ( $0 < \beta < 1$ ), and  ${}^1_A B$  expresses  $B$  as a warm spare of  $A$  that fails after  $A$  at full power. Therefore, the minimal cut sequences of dynamic gates, each having two input events, can be written as:  $FLE_{fdep} = e_1 + e_2$ ,  $FLE_{pand} = e_1 \rightarrow e_2$ ,  $FLE_{seq} = e_1 \rightarrow e_2$ ,  $FLE_{csp} = e_1 \rightarrow e_2$ ,  $FLE_{wsp} = ({}^\beta e_2 \rightarrow e_1) + (e_1 \rightarrow e_2)$ ,  $FLE_{hsp} = e_1 \cdot e_2 = e_1 \rightarrow e_2 + e_2 \rightarrow e_1$ , where the symbol "+" means the logical operator OR and " $\cdot$ " means the logical operator AND. The failure logic of the FDEP gate is equal to the OR gate, and the HSP gate is equivalent to an AND gate. In addition, the SEQ and CSP gates have similar failure behaviors. The only differences lie in the fact that input events under a SEQ can be an event representing a system, and the input events under a CSP are constrained to basic events representing components. For a DFT, the MCSS is unique regardless of whether its input events have reparability.

### 2.3. Logic operation rules in a repairable DFT

To obtain the FLE of a DFT, several logic operation rules are developed and applied. Liu et al. developed a set of inference rules to obtain the FLE of a given DFT, and Merle et al. presented several logic operation rules to deduce a DFT's structure function. In contrast to Merle's methods, Liu's inference rules are straightforward and simple [17]. In our approach, Liu's inference rules were introduced to obtain the FLE of a DFT. The detailed fundamental inference rules are listed as follows:

$$(A \rightarrow B) \rightarrow C \rightleftharpoons A \rightarrow B \rightarrow C \quad (3)$$

$$A \rightarrow (B \rightarrow C) \rightleftharpoons A \rightarrow B \rightarrow C + B \rightarrow A \rightarrow C \quad (4)$$

$$A \rightarrow (B + C) \cdot A \rightarrow B + A \rightarrow C \quad (5)$$

$$(A + B) \rightarrow C \cdot A \rightarrow C + B \rightarrow C \quad (6)$$

$$(A \rightarrow A) \rightleftharpoons A \quad (7)$$

$$A \rightarrow B \rightarrow A \rightleftharpoons \Phi \quad (8)$$

$$A \cdot B \rightleftharpoons A \rightarrow B + B \rightarrow A \quad (9)$$

where " $\rightleftharpoons$ " represents that the left are the necessary and sufficient conditions for the right, " $\Rightarrow$ " means that the left are sufficient but not necessary conditions for the right, and " $\Phi$ " denotes an empty set. Based on these fundamental inference rules, ten additional deductive inference rules are also offered. Interested readers are encouraged to consult reference 17. Through applying these inference rules, we can obtain the FLE of a repairable DFT.

## 3. The proposed numerical simulation method

### 3.1. Adapted sequence failure region and its formulation for a repairable DFT

The sequence failure region concept has been proposed in our previous contributions [8] and has already been used to analyze the reliability of a nonrepairable DFT. However, in a repairable DFT, the sequence failure regions are even more complex due to reparability behaviors. When a nonrepairable component enters a failure state, it never recovers again. However, for a repairable component, successful and failed states appear alternatively due to reparability. Its running state diagram is shown in Fig. 2, where  $T_i$  is the running time (i.e., time to failure),  $O_i$  is the repair time (i.e., time to recovery) and  $i = 1, 2, k$ . Hence, the sequence failure regions of repairable DFTs are different from those of nonrepairable ones.

The time to failure  $T_i$  and time to recovery  $O_i$  can be obtained by the following equations:

$$\begin{cases} T_i = F^{-1}(\varepsilon) \\ O_i = G^{-1}(\eta) \end{cases} \quad (10)$$

where the  $F(x)$  and  $G(x)$  are Cumulative Distribution Functions (CDFs) of  $T_i$  and  $O_i$ ,  $F^{-1}(x)$  and  $G^{-1}(x)$  are the corresponding inverse functions.  $\varepsilon$  and  $\eta \in (0,1)$  are uniformly distributed random numbers generated by standard random generators. For example, suppose a component follows exponent time-to-failure distribution with parameter  $\lambda$ , and its probability density function  $f(x)$  and cumulative probability distribution function  $F(x)$  are obtained as follows:

$$f(x) = \lambda \cdot e^{-\lambda \cdot x} \quad (11)$$

$$F(x) = \int_0^x f(x) dx = 1 - e^{-\lambda \cdot x} \quad (12)$$

Then, the  $x$  formulated as a function of  $F(x)$  is obtained as:

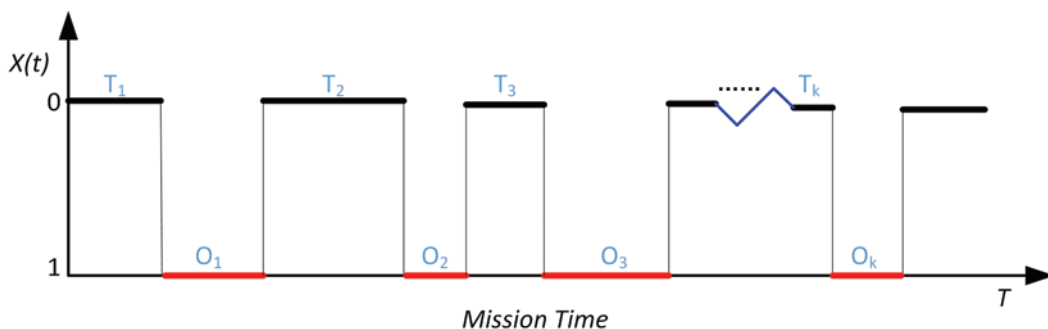


Fig. 2. Running state diagram of a repairable component

$$x = \frac{1}{\lambda} \cdot \ln\left(\frac{1}{1-F(x)}\right) \quad (13)$$

Let  $\lambda$  be  $1.0 \times 10^{-3}/\text{h}$ , and the generated random number is 0.5 which is used to replace the  $F(x)$ . Through applying Eq. (13), the time-to-failure of the component is simulated as 693.1 h. In the same way, we can get the time-to-recovery of the component. Alternately, the component's running state diagram can be derived.

A repairable DFT's failure state is determined by its MCSS. According to the semantics of an MCS, its failure state should satisfy two requirements: 1) the time-to-failure of components must occur in a sequential order; and 2) under the conditions of 1), all the components follow in a failure state at the common failure time interval. The sequential failure region of an MCS is demonstrated by a general minimal cut sequence  $e_1 \rightarrow e_2 \rightarrow \dots \rightarrow e_n$ , which is shown in Fig. 3.

The variables  $t_{ij}$ ,  $\mu_{ij}$  represent the  $j$ th time-to-failure and time-to-recovery of the  $i$ th component respectively. The variables  $T_{ij}$  and  $U_{ij}$  represent the  $j$ th failure time and recovery time located at the sequential failure region of the  $i$ th component, respectively. As observed in Fig. 3,  $T_{1,1} = t_{1,1} + \mu_{1,1} + t_{1,2} + \mu_{1,2} + t_{1,3}$ ;  $T_{2,1} = t_{2,1} + \mu_{2,1} + t_{2,2} + \mu_{2,2} + t_{2,3}$ ;  $T_{3,1} = t_{3,1} + \mu_{3,1} + t_{3,2}$ ;  $\dots$ ;  $T_{n,1} = t_{n,1} + \mu_{n,1} + t_{n,2} + \mu_{n,2} + t_{n,3}$ , and  $T_{1,1} < T_{2,1} < T_{3,1} < \dots < T_{n,1}$  satisfying the sequence failure requirement, where the "+" is the notation of summation. Under this condition, the lower boundary of the failure time interval of this MCSS is  $L_{sfr} = \text{Max}\{T_{1,1}, T_{2,1}, T_{3,1}, \dots, T_{n,1}\} = T_{n,1}$ , and the upper boundary is  $U_{sfr} = \text{Min}\{U_{1,1}, U_{2,1}, U_{3,1}, \dots, U_{n,1}\} = U_{3,1}$ . Therefore, the failure time interval (FTI) of this MCSS can be expressed as:  $FTI_{MCSS} = (L_{sfr}, U_{sfr}) = (T_{n,1}, U_{3,1})$ . Suppose a DFT has  $n$  minimal cut sequences (i.e., MCSS). Each MCSS has  $m$  failure time intervals, and the adapted sequence failure region (SFR) for this repairable DFT can be expressed as:

$$SFR_{dft} = \bigcup_{i=1}^n \bigcup_{j=1}^m FTI_{i,j} \quad (14)$$

In the simulation process, the obtained failure time intervals may have overlapping parts that may lead to a wrong reliability analysis result and should be merged and deleted. Four overlapping scenarios

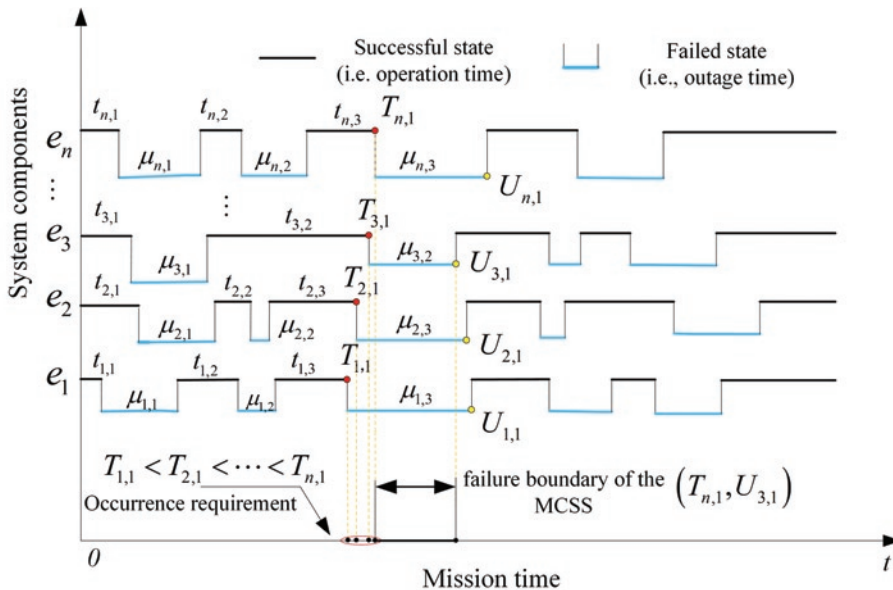


Fig. 3. Sequence failure region of a general MCSS

are identified (as shown in Fig. 4). The corresponding merging rules are provided for two overlapping failure time intervals ( $FTI_1 = (T_{11}, T_{12})$ ,  $FTI_2 = (T_{21}, T_{22})$ ) in Table 1. For two overlapping failure time intervals, the boundaries of the merged FTI are the lower and upper times of the two FTIs.

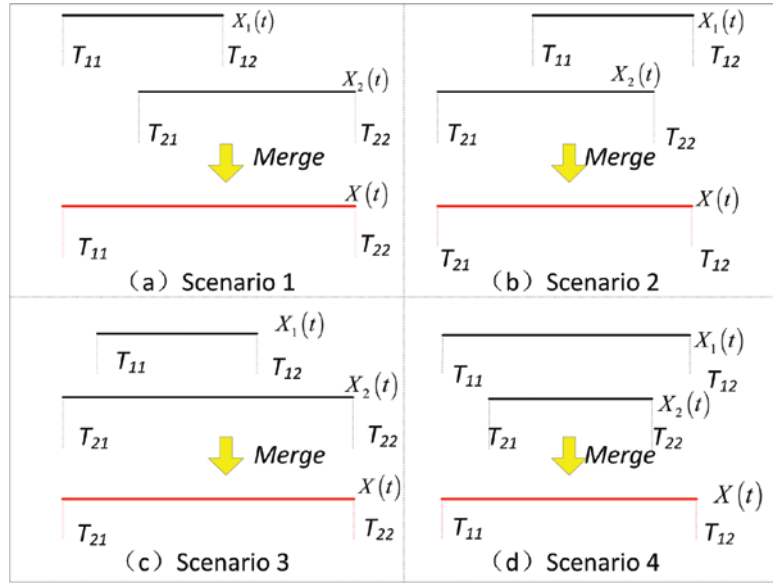


Fig. 4. Four identified overlapping scenarios

Table 1. Merging rules for overlapping failure time intervals

No.	Overlapping Conditions	Merging Rules
Scenario 1	$0 \leq T_{11} \leq T_{21}; T_{21} \leq T_{12} \leq T_{22} \leq T^*$	$FTI_1 \cup FTI_2 = (T_{11}, T_{22})$
Scenario 2	$0 \leq T_{21} \leq T_{11} \leq T_{22}; T_{22} \leq T_{12} \leq T^*$	$FTI_1 \cup FTI_2 = (T_{21}, T_{12})$
Scenario 3	$0 \leq T_{21} \leq T_{11}; T_{12} \leq T_{22} \leq T^*$	$FTI_1 \cup FTI_2 = (T_{21}, T_{22})$
Scenario 4	$0 \leq T_{11} \leq T_{21}; T_{22} \leq T_{12} \leq T^*$	$FTI_1 \cup FTI_2 = (T_{11}, T_{12})$

\* Mission time

### 3.2. Statistical formulas for reliability indices

A system's reliability indices are the indicators that can be applied to measure the degree of reliability. In a system reliability assessment, the system indices primarily include MTBF, MTTR, availability and a component's importance.

#### (1) MTBF, MTTR, Availability and Unavailability indices

MTBF is defined as the mean working time between two failure scenarios, MTTR is defined as the mean repair time between two working periods, and  $T_i$  is defined as the mission time ( $T_i = T$ ). Based on the merged failure time intervals, we can also obtain the working time intervals as  $\{[t_{1,2}^i, t_{2,1}^i], [t_{2,2}^i, t_{3,1}^i], [t_{3,2}^i, t_{4,1}^i], \dots, [t_{m_i-1,1}^i, t_{m_i,1}^i]\}$ . According to the obtained failure and working time intervals, the statistical indices of MTBF and MTTR can be expressed as:

$$MTBF_{sf} = \frac{\sum_{i=1}^N \sum_{j=1}^{m_i} (t_{j,2}^i - t_{j,1}^i)}{\sum_{i=1}^N T_i} \quad (15)$$

$$MTTR_{sf} = \frac{\sum_{i=1}^N \sum_{j=1}^{m_i-1} (t_{j+1,2}^i - t_{j,1}^i)}{\sum_{i=1}^N T_i} \quad (16)$$

The availability of a system is a very important reliability index, and it not only reflects the safety of a system but also reflects its economy. Based on MTBF and MTTR, the statistical index of a system's availability ( $A_{sf}$ ) and unavailability ( $UA_{sf}$ ) can be described as:

$$A_{sf} = \frac{MTBF_{sf}}{MTBF_{sf} + MTTR_{sf}} = \frac{\sum_{i=1}^N \sum_{j=1}^{m_i-1} (t_{j,2}^i - t_{j,1}^i) / \sum_{i=1}^N T_i}{\sum_{i=1}^N \sum_{j=1}^{m_i-1} (t_{j,2}^i - t_{j,1}^i) / \sum_{i=1}^N T_i + \sum_{i=1}^N \sum_{j=1}^{m_i-1} (t_{j+1,2}^i - t_{j,1}^i) / \sum_{i=1}^N T_i} \quad (17)$$

$$UA_{sf} = \frac{MTTR_{sf}}{MTBF_{sf} + MTTR_{sf}} = \frac{\sum_{i=1}^N \sum_{j=1}^{m_i-1} (t_{j+1,2}^i - t_{j,1}^i) / \sum_{i=1}^N T_i}{\sum_{i=1}^N \sum_{j=1}^{m_i-1} (t_{j,2}^i - t_{j,1}^i) / \sum_{i=1}^N T_i + \sum_{i=1}^N \sum_{j=1}^{m_i-1} (t_{j+1,2}^i - t_{j,1}^i) / \sum_{i=1}^N T_i} \quad (18)$$

## (2) Importance index

The importance index of a component can be used to arrange it according to its decreasing or increasing order of importance. In our proposed method, a simulation-based importance index for a component is introduced, namely, the failure criticality importance index ( $I^{FC}$ ) [15]. The fundamental idea of this concept is to divide the number of system failures caused by failure of component  $j$  with the failure number of the system in  $(0, t)$ , and the statistical formula  $I_{j,sf}^{FC}$  is defined as:

$$I_{j,sf}^{FC} = \frac{n_j}{\sum_{i=1}^N m_i} \quad (19)$$

where  $n_j$  represents the number of system failures caused by the considered component  $j$ , the variable  $m_i$  indicates the total number of system failures in the  $i$ th simulation round, and "caused" here means the final event that makes the system fail.

### 3.3. MCSS-based Monte Carlo numerical simulation methodology

Based on the aforementioned statements, the proposed MCSS-based Monte Carlo numerical simulation methodology can be implemented as shown in *Algorithm 1*.

#### Algorithm 1.

- Step 1.** Apply Liu's inference rules to obtain the MCSS of a DFT.
- Step 2.** Simulate the time-to-failure and time-to-recovery of each component contained in MCSS.
- Step 3.** Merge the overlapping parts to obtain the FTI of the MCSS.
- Step 4.** Establish statistical formulas for reliability indices.
- Step 5.** Calculate the reliability indices based on the merged FTI of the MCSS.
- Step 6.** Output the simulated reliability results.

The detailed simulation procedure is provided as shown in Fig. 5.

## 4. Numerical validation

To illustrate the reasonability of the proposed MCSS-based Monte Carlo numerical simulation methodology, a DFT from an adapted hypothetical cardiac assist system (HCAS) was chosen as a numerical validation case since it contained many kinds of dynamic gates [2].

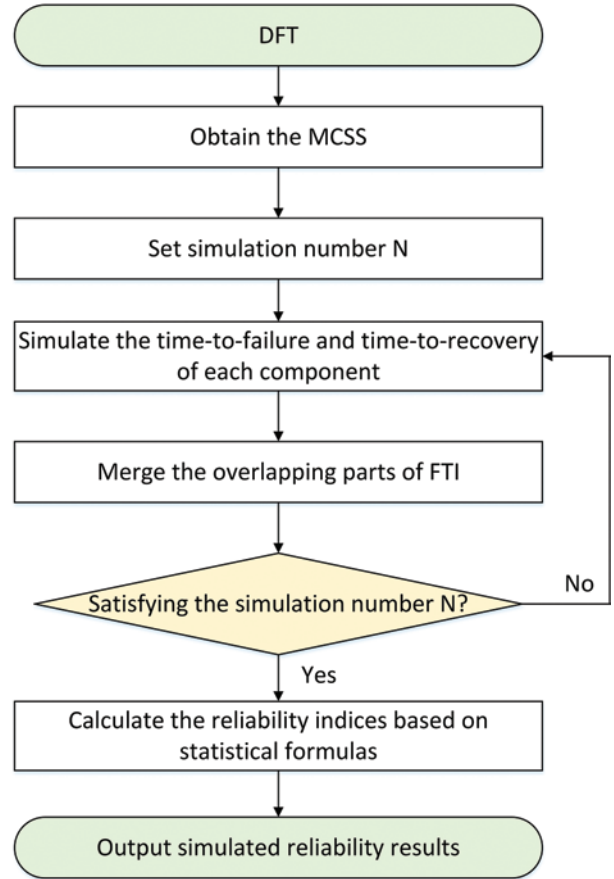


Fig. 5. Procedure for implementing the proposed numerical simulation methodology

### 4.1. Reliability evaluation

The original DFT model of HCAS is shown in Fig. 6. The reliability parameters  $\lambda_i$ ,  $\mu_i$  were the failure and repair rates of the  $i$ th component ( $e_i$ ) and were assumed to be:  $\lambda_i=10^{-3}$ ,  $\mu_i=0.5$  ( $i=1, 2, 3, \dots, 9$ ). Based on the inference rules for temporal operations, the MCSS of this DFT could be obtained as:

$$MCSS_{dft} = e_1 + e_2 + e_3 \rightarrow e_4 + e_4 \rightarrow e_3 + e_5 e_6 + e_7 \rightarrow e_7 e_9 \rightarrow e_8 + e_8 \rightarrow e_7 \rightarrow e_8 e_9 \quad (20)$$

where the hot spare gate (CUP) was logically equivalent to a static logic AND gate. Hence, its FLE ( $e_3 \cdot e_4$ ) could be expanded to  $(e_3 \rightarrow e_4) \cup (e_4 \rightarrow e_3)$  in the simulation process. In the same way, the AND gate (MOTOR) with input event  $e_5$  and  $e_6$  was expanded to  $(e_5 \rightarrow e_6) \cup (e_6 \rightarrow e_5)$ . The input event  $e_9$  was a repeated event and was contained in two different cut sequences,  $e_7 \rightarrow e_7 e_9 \rightarrow e_8$  and  $e_8 \rightarrow e_7 \rightarrow e_8 e_9$ .

In our study, reliability indices such as availability and the components' importance were evaluated by the proposed MCSS-based Monte Carlo numerical simulation method. To show the reasonability of the proposed methodology, the derived calculation results were compared with those obtained from the Markov chain state space-based approaches. All computations were implemented on a portable computer with an Intel (R) Core (TM) i5-4200M 2.5 GHz CPU and MATLAB programming platform.

### 4.2. Results and Discussions

We set the simulation number  $N$  as 10,000 rounds. The unavailability results at different mission times calculated by the proposed MCSS-based Monte Carlo numerical simulation methodology are shown in Table 2, which were compared with those obtained by the Markov chain state space-based methods. In addition, the compo-



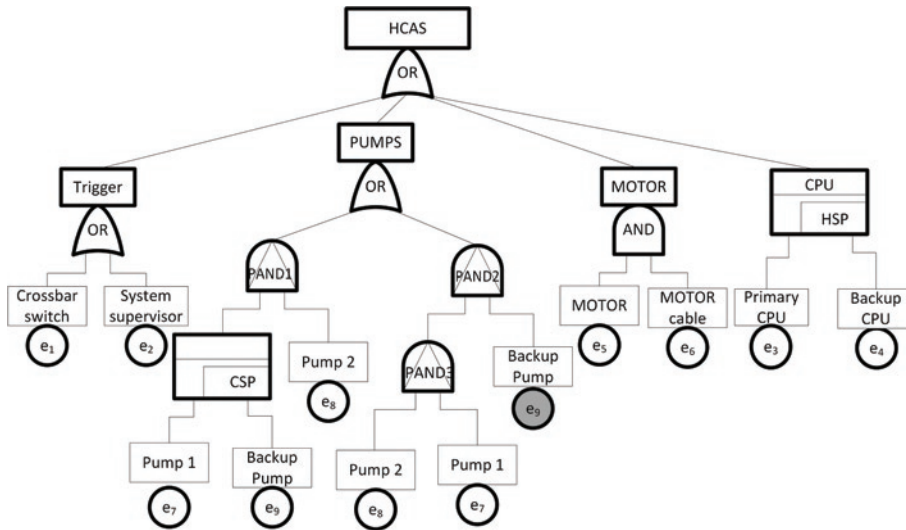


Fig. 6. DFT model of the adapted HCAS

Table 2. Unavailability results ( $N=10,000$ )

Mission time	100h	200h	400h	600h	800h	1000h
Proposed method	0.003990	0.004054	0.003998	0.004009	0.003943	0.004018
Markov method	0.003999	0.003999	0.003999	0.003999	0.003999	0.003999

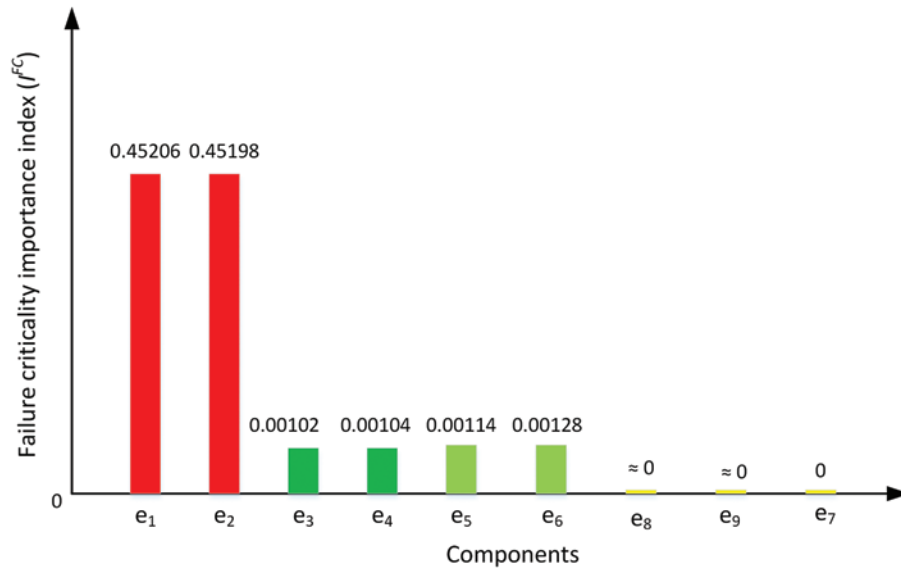


Fig. 7. Components' failure criticality importance index ( $I^{FC}$ )

Table 3. Unavailability results with A, B, and C following lognormal distributions ( $N=10,000$ )

Mission time	50h	100h	150h	200h	250h	300h
Unavailability	0.002534	0.019018	0.027261	0.029555	0.032073	0.032576

ponents' failure criticality importance index ( $I^{FC}$ ) was also calculated as shown in Fig. 7.

As observed in Table 2, the results calculated by our proposed methodology agreed with those obtained by the Markov chain state space-based methods, which demonstrated the effectiveness of our proposed methodology. Through applying the proposed methods, the components' failure criticality importance indices ( $I^{FC}$ ) were also de-

termined. As seen in the Fig. 7,  $e_1$  and  $e_2$  had almost the same  $I^{FC}$ , which was higher than those of the others and ranked as 1. The values of  $e_3$ ,  $e_4$ ,  $e_5$  and  $e_6$  had the second highest  $I^{FC}$  and were ranked as 2. The remaining components ( $e_7$ ,  $e_8$  and  $e_9$ ) were ranked as 3. In addition, when applying the pure Markov chain based methods, 272 states and 758 transitions were generated, and building the Markov Chain model manually would have cost approximately 3.5 hours. However, through applying the proposed methodology, the number of simulated minimal cut sequences was only 8, and the results could be provided in 3 ~ 4 seconds. Therefore, our proposed methodology was more effective and efficient than the Markov chain state space-based method.

To demonstrate the applicability of the proposed method for nonexponent distributions, we also assumed that the time-to-failure of components A, B, and C followed lognormal distributions, and their failure parameters were: mean  $\mu_{A,B,C}=100$  and variances  $\sigma_A=25$ ,  $\sigma_B=30$ ,  $\sigma_C=35$ . The unavailability results at different mission times calculated by the proposed MC-SS-based Monte Carlo numerical simulation methodology are shown in Table 3 (simulation number  $N=10,000$ ). However for this case, the Markov chain state space model was unavailable because the time-to-failure of some components did not follow exponent distributions.

## 5. Conclusions and future work

In our study, an MCSS-based Monte Carlo numerical simulation methodology was proposed for analyzing a repairable DFT. The main simulation ideas, procedures and statistical formulas for reliability indices were also developed. To illustrate reasonability and applicability of the proposed methods, we used a case study. With less computing time (3 ~ 4s), the results calculated by the proposed methods and Markov chain state space methods are well matched, which can demonstrate that the proposed method was straightforward and simple for analyzing a repairable DFT. In addition, the proposed methods can give more reliability indices than those provided by Markov chain state space-based methods, such as components' importance indices. Especially for a large-scale repairable DFT where some components have nonexponent time-to-failure distributions, the proposed methodology is also applicable and promising for the future.

However, the proposed MCSS-based Monte Carlo numerical simulation methodology is only suitable for repairable DFTs with time-dependent failure events, and is not applicable for demand failure events whose occurrence probabilities are independent of time. This can be viewed as a disadvantage. In the future, we

will focus on solving repairable DFTs with demand failure behaviors. Computer code development for MCSS-based Monte Carlo numerical simulation is also part of our ongoing work.

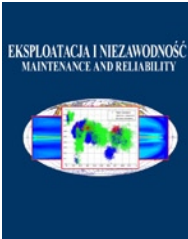
## Acknowledgments

This work is supported by the National Natural Science Foundation of China (71901203), the Natural Science Foundation of Anhui province (2008085MA23), and the National Key R&D Program of China (2018YFB1900301). In addition, the authors express their sincere gratitude to the State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment, China Nuclear Power Engineering Co., Ltd., Shenzhen.

## References

1. Alam M, Al-Saggaf UM. Quantitative reliability evaluation of repairable phased mission-time systems using Markov approach. *IEEE Transactions on Reliability* 1986; R-35(5): 498-503, <https://ieeexplore.ieee.org/abstract/document/4335529>.
2. Boudali H, Dugan JB. A discrete-time Bayesian network reliability modeling and analysis framework. *IEEE Transactions on Reliability* 2006; 55 (1):86-97, <https://ieeexplore.ieee.org/document/1603897>.
3. Boudali H, Nijmeijer A, Stoelinga M. DFTSim: A simulation tool for extended dynamic fault trees. *Proceedings of the 42nd Annual Simulation Symposium, Society for Modeling and Simulation International* (2009), pp. 31-38, <https://dl.acm.org/doi/10.5555/1639809.1639842>.
4. Budde CE, Biagi M, Monti RE, D'Argenio PR, Stoelinga M. Rare Event Simulation for Non-Markovian Repairable Fault Trees. In: Biere A., Parker D. (Eds) *Tools and Algorithms for the Construction and Analysis of Systems. TACAS 2020. Lecture Notes in Computer Science*, Vol. 12078. Springer, Cham, [https://link.springer.com/chapter/10.1007/978-3-030-45190-5\\_26](https://link.springer.com/chapter/10.1007/978-3-030-45190-5_26).
5. Budde CE, Stoelinga M. Automated Rare Event Simulation for Fault Tree Analysis via Minimal Cut Sets[M]// *Measurement, Modelling and Evaluation of Computing Systems, 20th International GI/ITG Conference, MMB 2020, Saarbrücken, Germany, March 16–18, 2020, Proceedings*. 2020, [https://link.springer.com/chapter/10.1007/978-3-030-43024-5\\_16](https://link.springer.com/chapter/10.1007/978-3-030-43024-5_16).
6. DUAN R, LIN Y, ZENG Y. Fault diagnosis for complex systems based on reliability analysis and sensors data considering epistemic uncertainty. *Eksplatacja i Niezawodność – Maintenance and Reliability* 2018; 20 (4): 558–566, <http://dx.doi.org/10.17531/ein.2018.4.7>.
7. Dugan JB, Bavuso SJ, Boyd MA. Fault trees and Markov models for reliability analysis of fault-tolerant digital systems. *Reliability Engineering and System Safety* 1993; 39(3):291–307, <https://www.sciencedirect.com/science/article/abs/pii/095183209390005J>.
8. Ejlali A, Miremadi SG. FPGA-based Monte Carlo simulation for fault tree analysis. *Microelectronics Reliability* 2004; 44(6): 1017-1028, <https://www.sciencedirect.com/science/article/abs/pii/S0026271404000769>.
9. Gascard E, Simeu-Abazi Z. Quantitative Analysis of Dynamic Fault Trees by means of Monte Carlo Simulations: Event-Driven Simulation Approach. *Reliability Engineering & System Safety* 2018; 180: 487-504, <https://doi.org/10.1016/j.ress.2018.07.011>.
10. Ge DC, Li D, Lin M, Yang YH. SFRs-based numerical simulation for reliability of highly-coupled DFTs. *Eksplatacja i Niezawodność – Maintenance and Reliability* 2015; 17(2):199–206, <http://dx.doi.org/10.17531/ein.2015.2.5>.
11. Ge DC, Lin M, Yang YH, Zhang RX, Chou Q. Quantitative analysis of dynamic fault trees using improved Sequential Binary Decision Diagrams. *Reliability Engineering and System Safety* 2015; 142: 289-299, <https://www.sciencedirect.com/science/article/abs/pii/S0951832015001763>.
12. Ge DC, Lin M, Yang YH, Zhang RX, Chou Q. Reliability analysis of complex dynamic fault trees based on an adapted K.D.Heidtmann algorithm. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 2015; 229(6): 576-586, <https://journals.sagepub.com/doi/abs/10.1177/1748006X15594694>.
13. Ge DC, Yang YH. Reliability analysis of non-repairable systems modeled by dynamic fault trees with priority AND gates. *Applied Stochastic Models in Business and Industry* 2015; 31(6): 809-822, <https://onlinelibrary.wiley.com/doi/abs/10.1002/asmb.2108>.
14. Guo J, Shi L, Zhang K, Gu K, and et al. Dynamic fault tree analysis based fault diagnosis system of power transformer. *The 10th World Congress on Intelligent Control and Automation (WCICA)*, 6-8 July 2012, Beijing China, pp. 3077-3081, <https://ieeexplore.ieee.org/abstract/document/6358400>.
15. Hilber P, Bertling L. A method for extracting reliability importance indices from reliability simulations of electrical networks. In *Proc. 15th Power Syst. Comput. Conf. (PSCC)*, Liege, Belgium, Aug. 2005, <https://services.montefiore.uliege.be/stochastic/pssc05/papers/fp475.pdf>.
16. LI J, DUAN R. Dynamic diagnostic strategy based on reliability analysis and distance-based VIKOR with heterogeneous information. *Eksplatacja i Niezawodność – Maintenance and Reliability* 2018; 20 (4): 610–620, <http://dx.doi.org/10.17531/ein.2018.4.12>.
17. Li Y, Wang B, Liu D, Yang H, Yang F. Complete Temporal Rules for Cut Sequence Generation in Dynamic Fault Tree Analysis. *Proceedings of the World Congress on Engineering 2013 Vol II, WCE 2013, July 3-5, 2013, London, U.K.*, [http://www.iaeng.org/publication/WCE2013/WCE2013\\_pp903-908.pdf](http://www.iaeng.org/publication/WCE2013/WCE2013_pp903-908.pdf).
18. Li YF, Huang HZ, Liu Y, Xiao N, Li H. A new fault tree analysis method: fuzzy dynamic fault tree analysis. *Eksplatacja i Niezawodność – Maintenance and Reliability* 2012; 14 (3): 208-214, <http://ein.org.pl/2012-03-04>.
19. Liu D, Zhang C, Xing L, Li R, Li H. Quantification of cut sequence set for fault tree analysis. *HPCC lecture notes in computer science*. V.4728. Houston, USA: Springer-Verlag; 2007. pp. 755–65, [https://link.springer.com/chapter/10.1007/978-3-540-75444-2\\_70](https://link.springer.com/chapter/10.1007/978-3-540-75444-2_70).
20. Manno G, Chiacchio F, Compagno L, D'Urso N, and Trapani N. MatCarloRe: An integrated FT and Monte Carlo Simulink tool for the reliability assessment of dynamic fault tree. *Expert Systems with Applications* 2012; 39 (12): 10334-10342, <https://doi.org/10.1016/j.eswa.2011.12.020>.
21. Merle G, Roussel JM, Lesage JJ, Bobbio A. Probabilistic Algebraic Analysis of Fault Trees with Priority Dynamic Gates and Repeated Events. *IEEE Transactions on Reliability* 2010; 59(1): 250-261, <https://ieeexplore.ieee.org/document/5361394>.
22. Merle G, Roussel JM, Lesage JJ, Perchet V, Vayatis N. Quantitative Analysis of Dynamic Fault Trees Based on the Coupling of Structure Functions and Monte Carlo Simulation. *Quality and Reliability Engineering International* 2014; 32(1): 7-18. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qre.1728>.
23. Merle G, Roussel JM, Lesage JJ. Quantitative analysis of dynamic fault trees based on the structure function. *Proceedings of Annual Reliability and Maintenance Symposium* 2011; 24-27 Jan. 2011, Lake Buena Vista, FL, USA, pp.1-6, <https://ieeexplore.ieee.org/document/5754452/authors#authors>.
24. Meshkat L, Dugan JB and Andrews JD. Dependability analysis of systems with on-demand and active failure modes, using dynamic fault trees. *IEEE Transactions on Reliability* 2002; 51(2): 240-251, <https://ieeexplore.ieee.org/abstract/document/1011531>.
25. Rao KD, Gopika V, Rao VVSS, and et al. Dynamic fault tree analysis using Monte Carlo simulation in probabilistic safety assessment. *Reliability Engineering and System Safety* 2009; 94(4): 872-883, <https://www.sciencedirect.com/science/article/abs/pii/S0951832008002354>.

26. Ren Y, Dugan JB. Design of reliable systems using static and dynamic fault trees. *IEEE Transactions on Reliability* 1998; 47(3): 234-244, <https://ieeexplore.ieee.org/document/740491>.
27. Ruijters E, Guck D, Drolenga P, Peters M, Stoelinga M. Maintenance Analysis and Optimization via Statistical Model Checking. In: Agha G., Van Houdt B. (Eds) *Quantitative Evaluation of Systems. QEST 2016. Lecture Notes in Computer Science, Vol 9826*. Springer, Cham, [https://link.springer.com/chapter/10.1007/978-3-319-43425-4\\_22](https://link.springer.com/chapter/10.1007/978-3-319-43425-4_22).
28. Tang Z, Dugan JB. Minimal Cut Set/Sequence Generation for Dynamic Fault Trees. *Proceedings of Annual Reliability and Maintenance Symposium 2004*; 26-29 Jan.2004, Los Angeles, CA, USA, pp.1-5, <https://ieeexplore.ieee.org/document/1285449>.
29. Xing L, Fleming KN, Loh WT. Comparison of Markov model and fault tree approach in determining initiating event frequency for systems with two train configurations. *Reliability Engineering and System Safety* 1996; 53(1): 17-29. <https://www.sciencedirect.com/science/article/abs/pii/S0951832096000336>.
30. Xing L, Shrestha A, Dai Y. Exact combinatorial reliability analysis of dynamic systems with sequence-dependent failures. *Reliability Engineering and System Safety* 2011; 96(10): 1375-1385, <https://www.sciencedirect.com/science/article/abs/pii/S0951832011001050>.
31. Xing L, Tannous O, Dugan JB. Reliability Analysis of Nonrepairable Cold-Standby Systems Using Sequential Binary Decision Diagrams. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 2012; 42(3): 715-726, <https://ieeexplore.ieee.org/document/6059511>.
32. Yao Y, Yang X, Li P. Dynamic fault tree analysis for digital fly-by-wire flight control system. *The 15th AIAA/IEEE Digital Avionics Systems Conference*, 27-31 Oct. 1996, Atlanta GA, pp. 479-484, <https://ieeexplore.ieee.org/abstract/document/559203>.
33. Yuge T, Yanagi S. Fault tree analysis considering sequence dependence and repairable input events. *Journal of Quality in Maintenance Engineering* 2013; 19(2): 199-214, <https://www.emerald.com/insight/content/doi/10.1108/13552511311315986/full/html>.
34. Yuge T, Yanagi S. Quantitative analysis of a fault tree with priority AND gates. *Reliability Engineering and System Safety* 2008; 93(11): 1577-1583. <https://www.sciencedirect.com/science/article/abs/pii/S0951832008000409>.
35. Zhang P, Chan KW. Reliability Evaluation of Phasor Measurement Unit Using Monte Carlo Dynamic Fault Tree Method. *IEEE Transactions on Smart Grid* 2012; 3(3): 1235-1243, <https://ieeexplore.ieee.org/document/6151874>.



Article citation info:

Ambroźkiewicz B, Syta A, Meier N, Litak G, Georgiadis A. Radial internal clearance analysis in ball bearings. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 42–54, <http://dx.doi.org/10.17531/ein.2021.1.5>.

Indexed by:



## Radial internal clearance analysis in ball bearings

Bartłomiej Ambroźkiewicz<sup>a,c\*</sup>, Arkadiusz Syta<sup>b</sup>, Nicolas Meier<sup>c</sup>, Grzegorz Litak<sup>a</sup>, Anthimos Georgiadis<sup>c</sup>

<sup>a</sup>Department of Automation, Lublin University of Technology, ul. Nadbystrzycka 36, 20-618 Lublin, Poland

<sup>b</sup>Institute of Technological Systems of Information, Lublin University of Technology, ul. Nadbystrzycka 36, 20-618 Lublin, Poland

<sup>c</sup>Institute of Product and Process Innovation (PPI), Leuphana University, Universitätsallee 1, 21335 Lüneburg, Germany

### Highlights

- Radial internal clearance in ball bearings is studied.
- Spectral and recurrence-based methods are applied for the time-series analysis.
- The optimal operating conditions of ball bearing are found.

### Abstract

Radial internal clearance (RIC) is one of the most important parameters influencing on rolling bearing exploitation in mechanical systems. Lifetime of rotary machines strongly depends on a condition of applied rolling elements, thus a study on applied clearance is very important in terms of maintenance and reliability. This paper proposes, a novel approach of studying RIC, based on a nonlinear dynamics method called recurrences. The results are confronted with standard analyses, i.e. statistical condition indicators, Fast Fourier Transform and Continuous Wavelet Transform. The application of the mentioned methods allowed us to find the optimal radial clearance for operating bearings. To ensure precise measurements of the clearance, an automated setup for RIC measurements is applied and next mounted in a plunger block and tested to finally measure vibration acceleration. The proposed methods are useful for a condition monitoring and lifetime prediction of bearings or bearing-based systems in which a proper value of radial clearance is crucial.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

ball bearings; radial internal clearance; statistical analysis; recurrences; fast fourier transform; continuous wavelet transform.

## 1. Introduction

The trend of health condition monitoring (HCM) is introduced in many areas of industrial processes to ensure constant and maintenance free operation of machines and mechanical systems. Among many and different types of bearings, ball bearings play an important role in various mechanical mechanisms as they provide seamless rotational movement and carry radial load in most cases. However, bearing failure modes have various origins such as obsolescence, accidents or surface degradation and regarding this last, mechanical wear can be connected with corrosion, abrasion, adhesion or fatigue. Bearing lifetime and durability strongly depend on whether a bearing is correctly adjusted to a specific application regarding mostly operating conditions and several other features such as lubrication, loads, power losses, operating velocities, bearing characteristic frequencies etc. are taken into account in its selection.

One of the most important parameters in ball bearings being crucial for operation and having a strong influence on its dynamic response is radial internal clearance (RIC), determined as the play between

rings and rolling elements. The selection of RIC in the ball bearing regarding its application is very important for both tribology and dynamics, and its optimal value has a strong impact on noise, generated vibrations, thermal expansion and fatigue [37]. In several works, the analysis of internal clearance in ball bearings was performed mostly based on the mathematical models. Tiwari et al. [41] discussed the influence of RIC on the dynamic response of a horizontal rotor, Harsha [14] studied the dynamics of a balanced rotor by variable radial clearance in time instead. Moreover, the influence of internal clearance on bearing life and damping ratio was referred by Yakout et al. [49] and Chudzik et al. [7] combined the FEM (Finite Element Method) model with the Lundberg-Palmgren fatigue theory. Miskovic et al. [32] analysed by thermo-graphic inspection how an amount of contamination in grease can influence RIC. Those works show the importance of internal clearance on bearing dynamics and nonlinear mathematical models of rolling bearings always refer to this parameter [42, 52].

Recently, the influence of bearing clearance on bearing performance has been increasingly studied to satisfy the high market demand for a reliable product [6, 17], especially for maintenance free applications. The development of online condition monitoring systems

(\* ) Corresponding author.

E-mail addresses: B. Ambroźkiewicz - [b.ambrozkiwicz@pollub.pl](mailto:b.ambrozkiwicz@pollub.pl), A. Syta - [a.syta@pollub.pl](mailto:a.syta@pollub.pl), N. Meier - [nmeier@leuphana.de](mailto:nmeier@leuphana.de), G. Litak - [g.litak@pollub.pl](mailto:g.litak@pollub.pl), A. Georgiadis - [georgiadis@leuphana.de](mailto:georgiadis@leuphana.de)

supposed to take into account changes of radial clearance [44] as its undesirable change in time can lead to premature wear. Xu et al. [48] applied the vibration analysis for studying bearing response with different clearance pointing the need for using optimal indicators in the analysis of spectra. Liu et al. [27] signalizes the fact that the obtained frequency spectra have stochastic character and one of the factors influencing them is radial clearance. Studies on bearing clearance are very demanding, and currently there is no accurate methodology for its detection in time. Conclusions coming from the state of the art works show the need to apply more accurate methods to analyse bearing dynamic response with different radial clearances.

Measurement of radial clearance in ball bearings is normalized and its principles are precisely described in the norm ISO-1132-2 regarding “Measuring and gauging principles and methods”. In our research, an automated setup [30, 31] measures RIC in a self-aligning ball bearing, and the experiment is run with a precisely and accurately determined clearance in a bearing, analysing its dynamic response. For the analysis of experimental acceleration time series, recurrence analysis is proposed and evaluated both in qualitative (recurrence plots) and quantitative (recurrence quantifiers) ways. Moreover, standard methods are used to confirm the dynamic states present in a bearing. For this case, statistical condition indicators and results of the Fast Fourier Transform (FFT) and Continuous Wavelet Transform (CWT) are discussed.

The recurrence method is effective at diagnostics in physiology, of manufacturing processes and faults in rotary systems. Focusing on the mechanical systems, Litak et al. [23, 25, 26] proposed in his works the application of recurrences to analyse nonlinear time series obtained from cutting and milling manufacturing processes. Kwuimy et al. [20-22] used the recurrence analysis to diagnose different systems by analysing results of mathematical models and experimental data. Fault diagnostics of rotary systems is discussed in other several papers [2, 24, 39]. The aim of our research is to analyse the dynamic response of a self-aligning ball bearing with different values of radial internal clearance.

According to the literature, radial clearance has a strong influence on bearing response, which is reflected in mathematical models. Currently, measurements of clearance are only performed on special test rigs or with instruments. There is no accurate method of its identifica-

tion in-situ, except from using a gauge feeler on an assembled and non-rotating bearing. For this purpose, the recurrence analysis is applied in order to evaluate bearing dynamic response related to internal clearance. The results obtained from recurrences are compared with calculations of statistical indicators and frequency and time-domain digital signal processing (DSP) methods.

The remainder of this paper is structured as follows. The second section describes the experimental setup consisting of the automated system for measuring radial clearance and the test rig. Certain essential principles of the clearance measurement and experimental features are also specified here. The third section presents fundamentals of all applied diagnostics methods, focusing on the analysis of nonlinear time series and significance of the applied recurrence-based methods in the analysis of the systems. In the fourth section, the data is processed and the experimental results are discussed. The last section summarizes the paper and presents the next steps in the research.

## 2. Experimental setup and experiment description

The experimental setup consists of two test rigs: 1) an automated setup to measure radial clearance in ball bearings and 2) a gearbox equipped with plummer blocks dedicated to self-aligning ball bearings and a high precision inverter system to precisely set system’s velocity. Both setups are in the laboratory of the Institute of Product and Process Innovation (PPI) at the Leuphana University of Lüneburg.

### 2.1. The automated setup for measuring bearing clearance

The first system presents a novel, very precise and fully automated way of RIC measurements. The test rig is adapted to a wide range of ball bearings varying in external dimensions and type. For the experiment, a (SABB) 2309SK tapered bore double row self-aligning ball bearing is used in which bearing clearance can be roughly set with a clamping sleeve and nut. The digital measurement of RIC is performed according to the standardised method and its flowchart is presented in Figure 1. The concept of the setup can be introduced into industrial practice as it meets the requirements for Measurement System Analysis (MSA) by providing reproducible and repeatable clearance measurements. The software applied in the test rig is written in C language and its execution is managed with the Arduino Due

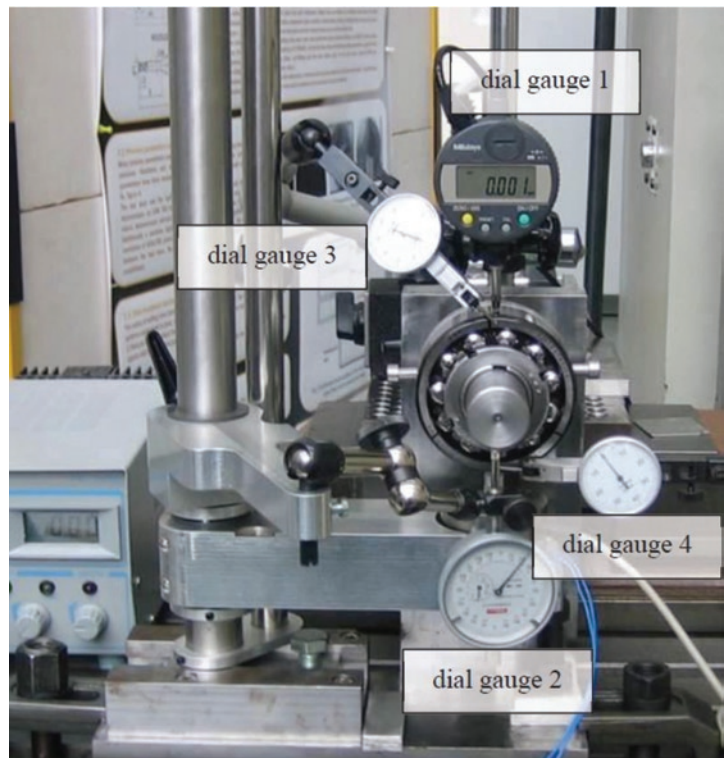
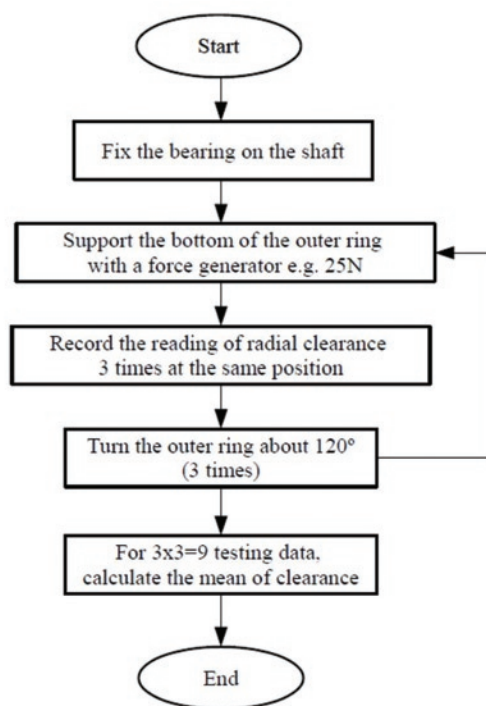


Fig. 1. Flowchart of RIC measurement procedure (left hand side), the automated setup for measuring bearing clearance (right hand side)

microcontroller. Four dial gauges in the setup are used to measure clearance, the distance of shifting the bearing onto the adapter sleeve and the influence of the test force on the displacement on the shaft.

## 2.2. Experimental gearbox

The bearing after the measurements is mounted into the plummer block (Figure 2) equipped with two accelerometers measuring bearing acceleration in two axes. The data is recorded with the piezoelectric (y-axis) and MEMS-based (x-axis) sensor with sampling time equal to 10.24ms, which corresponds to approximately 98Hz sampling frequency. The acceleration results are recorded with the data acquisition card (DAQ) in the CSV format and used in the further DSP analysis. Rotational velocity up to 50Hz (3000rpm) can be set on the test rig, but for the needs of the experiment, only one rotational velocity of 25Hz (1500rpm) is applied. The measurement for a single clearance lasts for around 10 minutes and is based on the obtained time series of acceleration, and the further DSP analysis is performed. The time of the experiment is fairly short so it is motivated to neglect a thermal expansion of the bearings during operation and ignore its influence on the clearance in contrary to the specially prepared cases [16, 33]. The influence of variable load is neglected as the measurements are focused on the impact of clearance on response. There are only constant loads generated by a housing and torque as a reaction from the rotating shaft. For the experiment, only brand-new bearings are used and it is assumed that there is no influence of characteristic bearing defects on the measured acceleration. In total, 6 cases with different clearances are studied. The detailed information on the applied testing setup and performed experiment are collected in Table 1 and 2, respectively.

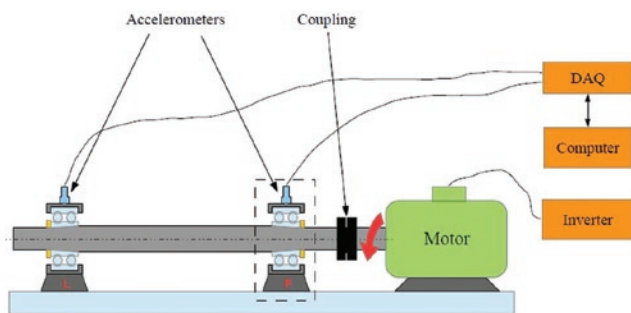


Fig. 2. Gearbox used for the bearing testing, the data were recorded in two axes only for the fixed bearing (right hand side)

Table 1. Equipment for the experimental setup.

Component	Type
Ball Bearing	NTN 2309SK
Adapter Sleeve	NTN H2309
Locating Ring	NTN FR 100x4
Bearing Housings	NTN SNC 511-609
MEMS Sensor	IFM VSA001
Piezoelectric Sensor	IFM VSP001
Data Acquisition System	IFM VSE100
3-Phase Motor	Siemens 1LA5090-4AA60-Z
Frequency Converter	Siemens SIMOVERT 6SE2103

Table 2. Duty cycle applied in the experiment.

Rotational velocity [rpm]	25Hz					
RIC [ $\mu\text{m}$ ]	7	20	22	34	41	46

## 2.3. Ball bearing characteristic frequencies

The classical method used in diagnostics of ball bearing damages is based on the determination of characteristic frequencies related to the specific part [9, 35]. Each characteristic frequency has its own abbreviation in the industrial nomenclature. Their values depend on bearing's internal dimensions, features and rotational velocity. The basic characteristic frequencies for the NTN 2309SK bearing (Figure 3) are specified in Table 3 and the geometrical data are collected in Table 4.

Table 3. Ball bearing characteristic frequencies.

Characteristic frequency	Formula	Values of characteristic frequencies for SABB NTN 2309SK [Hz]
Fundamental Train Frequency (FTF)	$FTF = \frac{s}{2} \left( 1 - \frac{b_d}{d_p} \cos\beta \right)$	0.394
Ball Spin Frequency (BSF)	$BSF = \frac{d_p}{b_d} s \left( 1 - \left( \frac{b_d}{d_p} \cos\beta \right)^2 \right)$	2.160
Ball Pass Frequency (Inner Ring - BSFI)	$BSFI = \frac{n_b}{2} s \left( 1 + \frac{b_d}{d_p} \cos\beta \right)$	7.884
Ball Pass Frequency (Outer Ring - BSFO)	$BSFO = \frac{n_b}{2} s \left( 1 - \frac{b_d}{d_p} \cos\beta \right)$	5.116

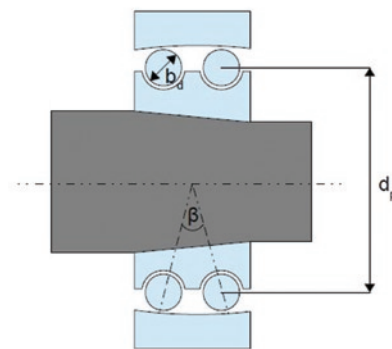


Fig. 3. Double row self-aligning ball bearing with its relevant dimensions marked taken for the calculation of its characteristic frequencies

The FFT and Wavelet results give the information about characteristic frequencies in the spectra and value of their amplitudes, and especially, the amplitude's value is taken into account to determine bearing damage. In this analysis, it is expected not to obtain high values of frequency peaks as the brand-new bearing is tested. The variable parameter is radial clearance and its different value influences the obtained frequency spectra.

Table 4. Ball bearing characteristic frequencies.

Parameters for the NTN 2309SK bearing	
Ball diameter $b_d$ [mm]	15.870
Pitch diameter $d_p$ [mm]	71.810
Contact angle $\beta$ [°]	15.52
Number of rolling elements $n_b$ [-]	26 (13 per row)
Rotational velocity $s$ [Hz]	25

### 3. Diagnostics methods

Over the years, mostly frequency-domain DSP (Digital Signal Processing) methods such as FFT (Fast Fourier Transform), EMD (Empirical Mode Decomposition), SK (Spectral Kurtosis), etc. were applied in diagnostics of rotational systems. The relevant support for the frequency-based methods is the analysis performed in time domain as it allows us to observe and detect dynamic changes present in the system for a specific period of time. In this section, the principles of the applied methods to process experimental data are presented. Acceleration time-series are analysed with statistical condition indicators, the Fast Fourier Transform, the Continuous Wavelet Transform and Recurrences.

#### 3.1. Diagnostics indicators applied to bearings diagnostics

The application of statistical condition indicators (CI) to the obtained time series from the experiment shows its accuracy in diagnostics of different kind of damages or transient states occurring in gears, bearings, shafts etc. [11, 28]. The mentioned indicators may be divided into time, frequency and time-frequency domain. In our experiment, we use only some of them, supposing their most significant impact on an identification of bearing response with different RICs.

- 1) Standard deviation ( $x_{std}$ ) measures the dispersion of results regarding its mean and is calculated as the square root of variance [45]. The value of standard deviation corresponds to the amplitude of misalignment in the planar section:

$$x_{std} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

- 2) Kurtosis ( $x_{kur}$ ) defines how heavily the tails of a given distribution differ from the tails of a normal distribution. This indicator is applied to detect fracture [38] and wear [3] of components:

$$x_{kur} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(n-1)x_{std}^4} \quad (2)$$

- 3) Crest factor (CF) is the ratio of instantaneous peak amplitude of a signal, to its root mean square RMS value. The mentioned parameter can diagnose different sources of impulsive vibrations (snaps and jerks) in the system [10]:

$$CF = \frac{x_0 - x_{pk}}{RMS_x} \quad (3)$$

- 4) Energy ratio (ER) is the ratio between a residual signal and a deterministic periodic signal which is separated using an auto-regressive model. With the ER, there is a high possibility to detect strong wear or defect [51]:

$$ER = \frac{RMS_d}{RMS_{y_d}} \quad (4)$$

#### 3.2. Fast Fourier Transform (FFT)

The Fast Fourier Transform as the most popular DSP method used in diagnostics of mechanical structures provides the information on characteristic frequencies present in spectra [1]. The analysed time-series is decomposed into a number of sinusoidal signals characterised by specific frequency. Information of potential damage of the system is presented on a spectrogram which consists of characteristic frequency peaks with determined amplitude. In the case of strongly nonlinear and non-stationary data, the FFT loses the information about time, so there is no possibility to define how long a specific state lasts. Additionally, noise present in experimental time-series introduces additional frequency peaks which can mix with other relevant peaks denoting damage so a band-pass filter is necessary [8].

In the analysis of discrete signals  $x[n]$ , the discrete Fourier transform (DFT) can be described in the following way [34]:

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-j \left( \frac{2\pi nk}{N} \right)} \quad (5)$$

where:  $N$  is the harmonic index regarding the exponential function for  $k=0,1,\dots,N-1$ .

In the final analysis with the FFT, the scaling of the spectrum is necessary in frequency units, and each signal  $X(k)$  is attributed to frequency  $f[k]$  according to the following formula:

$$f[k] = k \frac{f_p}{N} \quad (6)$$

where:  $f_p$  is sampling frequency.

#### 3.3. Continuous Wavelet Transform (CWT) and Scale Index (SI)

The disadvantage of the FFT is eliminated with an application of the Wavelet Analysis by complementing it with a localisation of periodicities in time. In the analysis of mechanical systems, it is important to detect transient and intermittent phenomena which can occur in short periods of time. The window size is crucial in time series analysis in time and frequency resolution. Owing to wavelets analysis, the window size is adapted to the identified state in the signal: it narrows when focusing on high-frequency or small-scale features and widens on low-frequency or large-scale features like in zoom lens with adjustable focus [15].

The mathematical definition of the continuous wavelet transform regarding acceleration time series has a following form:

$$W_{s,n}(Acc) = \sum_{i=1}^N \frac{1}{s} \left( \frac{i-n}{s} \right) \frac{(Acc(t_i) - \langle Acc \rangle)}{\sigma_{Acc}} \quad (7)$$

where  $\langle Acc \rangle$  and  $\sigma_{Acc}$  are average and standard deviations of acceleration,  $s$  denotes the scale index and  $n$  is the time index.

The wavelet power spectra (WPS) of the signal is defined as the squared modulus of the CWT and is the following:

$$P_w = |W_{s,n}|^2 \quad (8)$$

Another wavelet-based tool is the Scale Index (SI) proposed by Benitez et al. [4, 5]. Notably intended for non-stationary time-series, this index provides the quantitative measure of signal's non-periodicity. A scalogram of a given analysed signal has the following form:

$$S(s) = \left( \int_{c(s)}^{d(s)} |WAcc(u,s)|^2 du / (d(s) - c(s)) \right)^{\frac{1}{2}} \quad (9)$$

where  $WAcc(u,s)$  is the CWT of the signal at specified time  $u$  and scale  $s$ .

The scale index interval  $[s_{\rho}, s_l]$  is determined as the quotient:

$$i_{scale} = \frac{S(s_{min})}{S(s_{max})} \quad (10)$$

where  $s_{max}$  is the maximum in the scale index interval and  $s_{min}$  is the minimum defined in the left scale interval after the maximum and  $s_l$ . The result of the scale index is from zero to one, which quantifies the measure of non-periodicity of the signal. The value close to one denotes the strongly nonlinear signal and value close to zero denotes the periodic signal.

### 3.4. Recurrence-based methods

In the analysis of strongly nonlinear experimental time series, the statistical and DSP analyses should be supported with an alternative method showing changes in short-time intervals. One of the possible methods is recurrence analysis. Experimental data are mostly characterized by strong nonlinearities and noise, so recurrences are a promising method in our analysis. A change in the RIC value can cause changes in the dynamic response of the analysed system that can be identified by recursion analysis. Recurrence is understood here as the property of the dynamical oscillator when a specific state of the system returns to the previous one after some time. Firstly, the idea of recurrences was proposed by Henri Poincare in his seminal work to analyse conservative systems. In order to visualise trajectories of dynamic systems, Eckmann et al. [12] introduced the recurrence plots (RP) method designed to locate hidden recurring patterns, non-stationarity and structural changes. In further research, this method was extended by Webber et al. [46, 47] and Marwan et al. [29] by statistics of points as well as diagonal and vertical lines from recurrence plots.

The next step in the development of recurrence-based methods was the introduction of the recurrence quantification analysis (RQA) [50] in which recurrence quantifiers show quantitative information from recurrence plots. Nowadays, in many research, both recurrence plots and recurrence quantification analysis approaches are applied to both experimental and modelled nonlinear systems from many fields.

#### 3.4.1. Recurrence plots method

The principle of the RP-based method is revealing all times when the phase space trajectory of the dynamical system meets roughly the same area in the phase space. Two points on the trajectory are considered as neighbours when they are close enough to each other and this can be expressed with the  $R$  distance matrix with its element  $R_{ij}^{\epsilon}$  calculated as follows:

$$R_{i,j}^{\epsilon} = H\left(\epsilon - \|\mathbf{x}_i - \mathbf{x}_j\|\right), i, j = 1, \dots, N \quad (11)$$

where  $N$  is the number of considered states  $\mathbf{x}_i, \mathbf{x}_j$  for a threshold distance  $\epsilon$ ,  $\|\cdot\|$  a norm and  $H$  the Heaviside function. The obtained matrix consisting of zeros and ones denotes corresponding recurrence plots by:

$$R_{i,j} = \begin{cases} 1: \{\mathbf{x}_i\} \approx \{\mathbf{x}_j\} \\ 0: \{\mathbf{x}_i\} \neq \{\mathbf{x}_j\} \end{cases}, i, j, \dots, N \quad (12)$$

where  $\{\mathbf{x}_i\} \approx \{\mathbf{x}_j\}$  indicates points coordinates in the embedding dimension space belonging to the neighbourhood of radius  $\epsilon$ . To obtain the recurrence plot, we need three following parameters, i.e. threshold -  $\epsilon$ , embedding dimension -  $m$  and time delay -  $\tau$  to create missing coordinates [40]. The threshold value is a maximal distance that marks two points as a recurrence. There are several approaches including percentage of standard deviation of data, but usually a few percent of recurrence points is sufficient. The time delay should be optimal to preserve both short and long term correlations (autocorrelation method or mutual information method [13]). The value of embedding dimension must ensure the situation when there are no false neighbours in the reconstructed attractor (false neighbours method [18]). On the other hand, instead of reconstructing the attractor, it is possible to apply the recorded data in the plumper block. In our case, these are acceleration time series measured in two directions:  $x$  and  $y$  as components of the state vector that clearly describe the dynamics of the system at a given time. This approach avoids errors due to improper reconstruction and was adopted in this work.

#### 3.4.2. Recurrence quantification analysis

The recurrence quantification analysis (RQA) provides a quantitative interpretation of obtained recurrence plots quantifying the number and the duration of recurrences for the analysed dynamic system within its state space trajectory. The three main recurrence quantifiers are the recurrence rate (RR) based on a recurrence points density, determinism (DET) based on diagonal lines statistics and laminarity (LAM) based on vertical lines. The appearance of a diagonal line on the recursive chart indicates periodicity and the length of this line. The appearance of a vertical line on the recursive graph indicates a laminar state when the system is stuck for a time equal to the length of the line in one specific state. Both statistics can be used to detect the system switching between its various states.

- (1) Recurrence Rate (RR) describes the density of recurrence points on the recurrence plot:

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j} \text{ for } |i-j| \geq 1 \quad (13)$$

- (2) Determinism (DET) expresses the percentage of recurrence points creating diagonal lines in the recurrence plot of minimal length  $l_{min}$ :

$$DET = \frac{\sum_{l=l_{min}}^N IP(l)}{\sum_{l=l_{min}}^N P(l)} \quad (14)$$

- (3) Laminarity (LAM) represents the ratio between recurrence points forming vertical structures:



$$LAM = \frac{\sum_{v=v_{min}}^N vP(v)}{\sum_{v=v_{min}}^N vP(v)} \quad (15)$$

- (4) Diagonal line (L)  $R_{i+k,j+k} \equiv 1 \Big|_{k=0}^{l-1}$  occurs when a segment of the trajectory runs almost in parallel to another segment for  $l$  time units:

$$\mathbf{x}_i \approx \mathbf{x}_j, \mathbf{x}_{i+1} \approx \mathbf{x}_{j+1}, \dots, \mathbf{x}_{i+l-1} \approx \mathbf{x}_{j+l-1} \quad (16)$$

A diagonal line of length  $l$  is then defined by:

$$(1 - R_{i-1,j-1})(1 - R_{i+l,j+l}) \prod_{k=0}^{l-1} R_{i+k,j+k} \equiv 1 \quad (17)$$

The length of this diagonal is determined by the duration of such similar local evolution of trajectory segments.

- (5) Vertical line (V)  $R_{i,j+k} \equiv 1 \Big|_{k=0}^{v-1}$  marks a time interval in which a state does no change or changes very slowly:

$$\mathbf{x}_i \approx \mathbf{x}_j, \mathbf{x}_{i+1} \approx \mathbf{x}_{j+1}, \dots, \mathbf{x}_{i+v-1} \approx \mathbf{x}_{j+v-1} \quad (18)$$

The formal definition of a vertical line is:

$$(1 - R_{i,j-1})(1 - R_{i,j+v}) \prod_{k=0}^{v-1} R_{i,j+k} \equiv 1 \quad (19)$$

#### 4. Data analysis and results

In this section, we report the results of all applied methods for the acceleration results obtained from the experiment for bearing response differing from each other in radial clearance values. In order to distinguish the differences between experimental cases, firstly, the statistical indicators were calculated for time series consisting of 15000 data points, the results from each sensor were analysed separately, rotational velocity during the test was chosen  $n=1500$  rpm.

Table 5 (x-axis) and Table 6 (y-axis) show the results of the most statistical indicators. The results obtained differ in values in a certain order. The standard deviation results for both axes have a decreasing tendency till  $RIC=22\mu m$ , whereas its value has an increasing trend in the next cases. The low value of STD can be related to low amplitudes obtained in the signal, and consequently, there are low amplitudes of horizontal and vertical accelerations in the bearing.

The results of kurtosis are different in nature in each axis. For the x-axis in the area of interest,  $RIC=22\mu m$  kurtosis has the maximum value. In contrast, for the y-axis for the same value of RIC, the kurtosis has the minimum. However, the trend of that value is kept and the values of kurtosis in all cases are close to 3, which indicates a Gaussian-like peakedness. The leptokurtic (more peaky distribution,  $Kurt > 3$ ) and platykurtic (flatter than Gaussian distribution,  $Kurt < 3$ ) shapes of distribution in different axes are caused by the subjected load and the torque acting on the bearing from the shaft, which influences differently on each accelerometer. For the larger kurtosis of time series, the dynamical system is close to pass intermittency between two solutions [36].

The distribution of the crest factor (CF) values is similar to the previous indicator: single transient peaks in the time course cause an increase of the CF in relation to the small standard deviation value of the signal. The correlation between the RIC value and the mentioned indicator is observed in the energy ratio values of the signal, for both axes. This means that the residual signal is richer for the smaller values of clearance, which is related to a tight deployment of rolling surfaces resulting in increased friction or amplitude of contact stresses.

Based on statistical indicators results, we conclude that there is a correlation between increasing RIC values and corresponding indicators. Note, standard deviation and kurtosis show non-monotonic

Table 5. Values of statistical indicators for the results obtained from the x-axis at 25Hz.

RIC [ $\mu m$ ]	Standard deviation $\sigma$ [ $m/s^2$ ]	Kurtosis	Crest Factor	Energy Ratio
7	18.448	3.097	1.148	0.441
20	18.201	3.126	1.132	0.391
22	9.629	3.356	1.190	0.362
34	22.003	3.071	1.140	0.378
41	32.593	2.926	1.150	0.335
46	33.839	2.972	1.177	0.337

Table 6. Values of statistical indicators for the results obtained from the y-axis at 25Hz.

RIC [ $\mu m$ ]	Standard deviation $\sigma$ [ $m/s^2$ ]	Kurtosis	Crest Factor	Energy Ratio
7	3.678	2.974	1.117	0.409
20	3.547	3.081	1.127	0.370
22	1.847	2.641	1.145	0.349
34	3.606	3.476	1.120	0.326
41	5.064	2.771	1.111	0.322
46	4.848	3.202	1.128	0.313

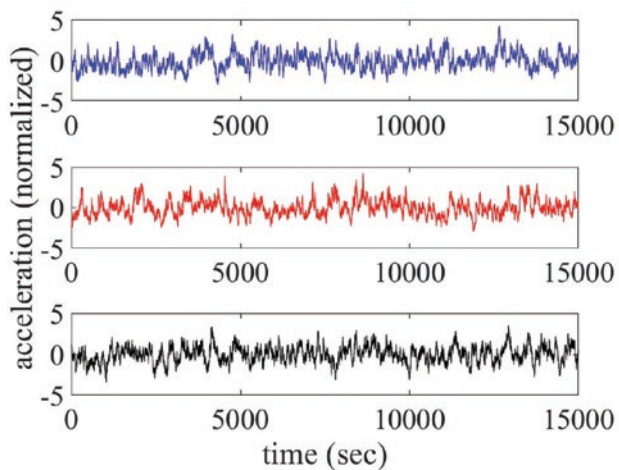


Fig. 4. Normalized acceleration time series obtained from the x-axis for the selected clearance of RIC=20 $\mu\text{m}$  – blue, 22 $\mu\text{m}$  – red and 34 $\mu\text{m}$  – black

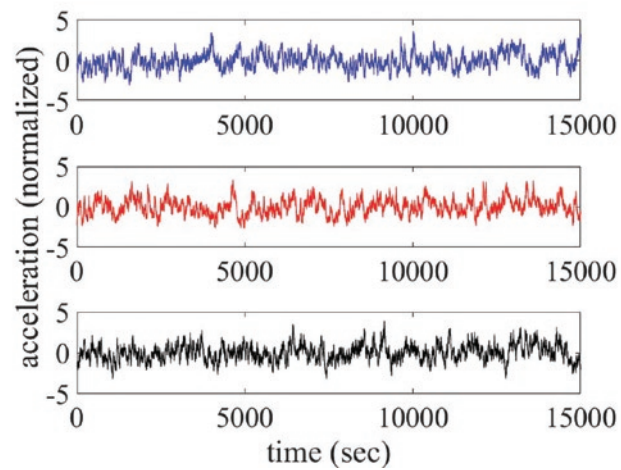


Fig. 5. Normalized acceleration time series obtained from the y-axis for the selected clearance of RIC=20 $\mu\text{m}$  – blue, 22 $\mu\text{m}$  – red and 34 $\mu\text{m}$  – black

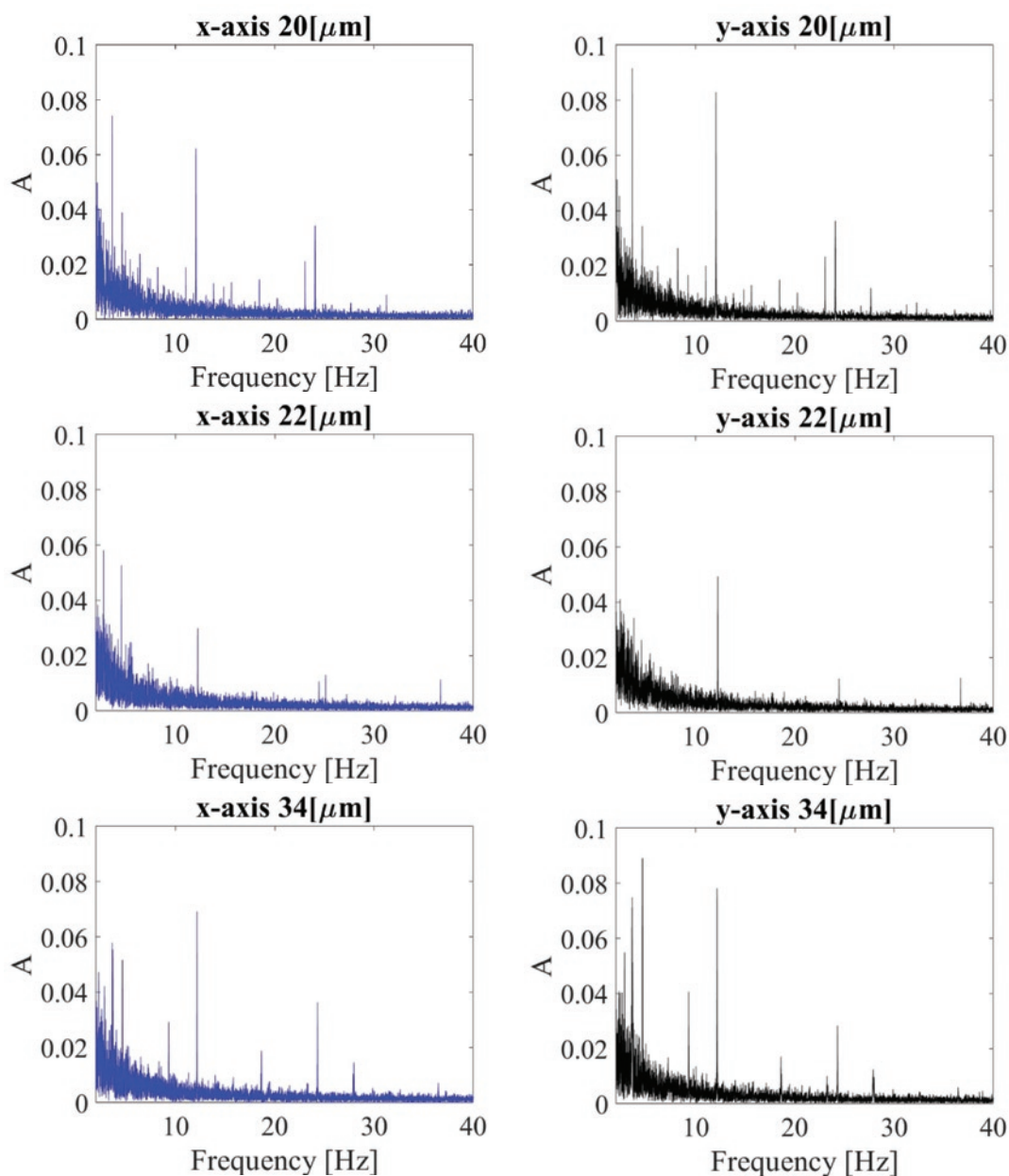


Fig. 6. Results of the FFT (Fast Fourier Transform) for the time series obtained from the x-axis (left panel) and the y-axis (right panel) for the selected clearance of RIC=20, 22 and 34 $\mu\text{m}$  (downwards)

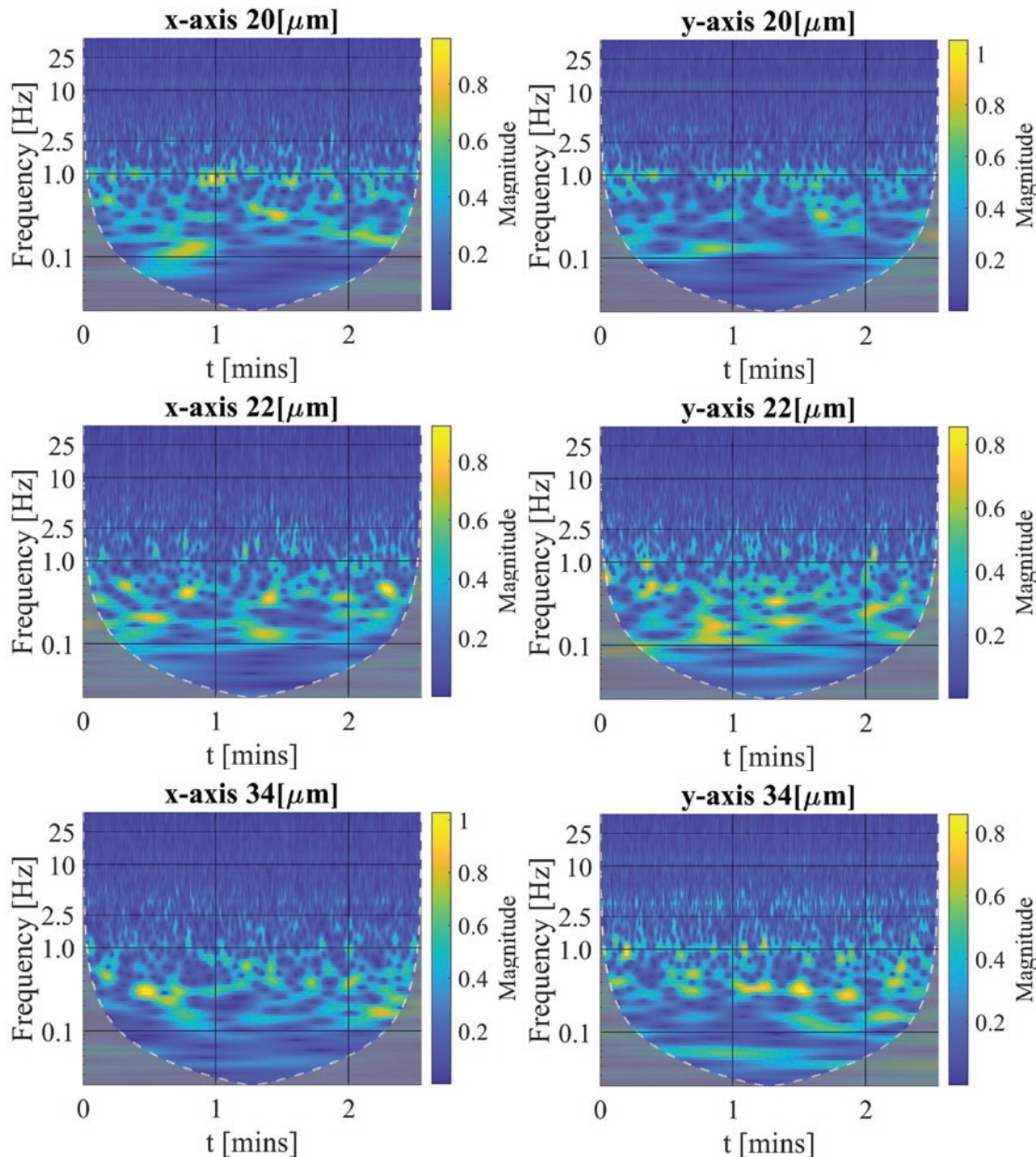


Fig. 7. Results of the CWT (Continuous Wavelet Transform) for the time series obtained from the x-axis (left panel) and the y-axis (right panel) for the selected clearances of RIC=20, 22 and 34 $\mu\text{m}$  (downwards)

dependence on RIC values, while the crest factor is almost constant and the energy ratio has a decreasing trend. The statistical indicators provide some information about dynamics of the bearing system, however it is difficult to draw more specific conclusions. One of the approaches is the calculation of the Fast Fourier Transform for the area of interest, cases RIC={20;22;34} $\mu\text{m}$  for the corresponding acceleration time courses in the x and y axis (Figure 6). In all cases, the basic bearing's characteristic frequencies specified in Table 3 are observed, all of the peaks have small amplitude, so the bearing used in the experiment is not damaged. However, the frequency spectra for the specific clearance differ from each other by the number of additional peaks enriching the response. The spectra for RIC=22 $\mu\text{m}$  have less complex spectra with the least number of frequency peaks, and the only one main harmonic and few smaller frequency peaks are observed. For RIC=20 $\mu\text{m}$  instead the share of small peaks is relatively bigger than for RIC=34 $\mu\text{m}$ , which is connected with increased nonlinearity in the response.

As the relevant information about the time is lost by the FFT analysis, the Continuous Wavelet Transform is used. The resulting WPS for acceleration time series for RIC={20,22,34} are depicted in Figure 7. The yellow and blue areas on the scalograms represent the highest and

lowest power levels, respectively, with the other colours referring to the intermediate power level. The grey area below the white dashed U-shaped line represents the cone of influence (COI), i.e. the region of wavelet spectrum in which edge effects become important and is defined as the  $e$ -folding time for the autocorrelation of wavelet power at each scale [43]. The rest of the scalogram should only be considered for further interpretation and represents the regions of greater than 95% confidence.

The characteristic frequencies found with the FFT are also depicted on the scalograms by the stronger and weaker yellow zones. The observed strong noise in the spectrogram by low frequencies is deposited on the scalograms. In all cases, the yellow areas are focused around the frequency of 0.4Hz, which relates to the Fundamental Train Frequency and is observed in the entire analysed signal with single gains in the power spectra. Besides, stronger but short periodic tendencies and intermittent behaviour can be observed in the frequency interval of [0.1, 1]Hz (Figure 7). Both sensor channels fixed on the x and y axes show similar results in terms of frequencies. The differences between clearance cases related to the amplitude of peaks are also visible, by 20 and 34 $\mu\text{m}$ , the area around the frequencies 12.5 and 25Hz is definitely stronger it is constant by entire time-course. For 22 $\mu\text{m}$ , the

power spectra are fairly diminished for the mentioned frequencies, but the dominate frequencies are shown in the numerous and very short time-intervals from the extended frequency range  $f=[0.1;10\text{Hz}]$ . This fact can be connected with the results obtained by calculated the condition indicators and FFT, assumedly by  $\text{RIC}=22\mu\text{m}$  as there is the most quiet (stable) bearing operation. The contact between balls and raceways is well fitted, resulting in low amplitude horizontal and vertical displacement.

In order to quantify the found observations in the CWT scalograms, Inner scalograms and Scale Index Analysis (SIA) are performed (Figure 8, Table 7). Note that for both sensors the inner scalograms are showing a small corrugation of the  $\text{RIC}=22\mu\text{m}$  case (see the red lines in Figure 8) in contrast to the other cases. The SIA results obtained proved that the lack of higher amplitudes of vibration, improving bearings rotational character of motion can be found in the case of  $\text{RIC}=22\mu\text{m}$  (Table 7). The course in scale domain has the most smooth character and estimated SI is almost 1.0, relating to its high regularity in the rotational motion in frequencies from the range of  $f=[10;45]\text{Hz}$ . This interval was chosen to neglect the sampling noise oscillations.

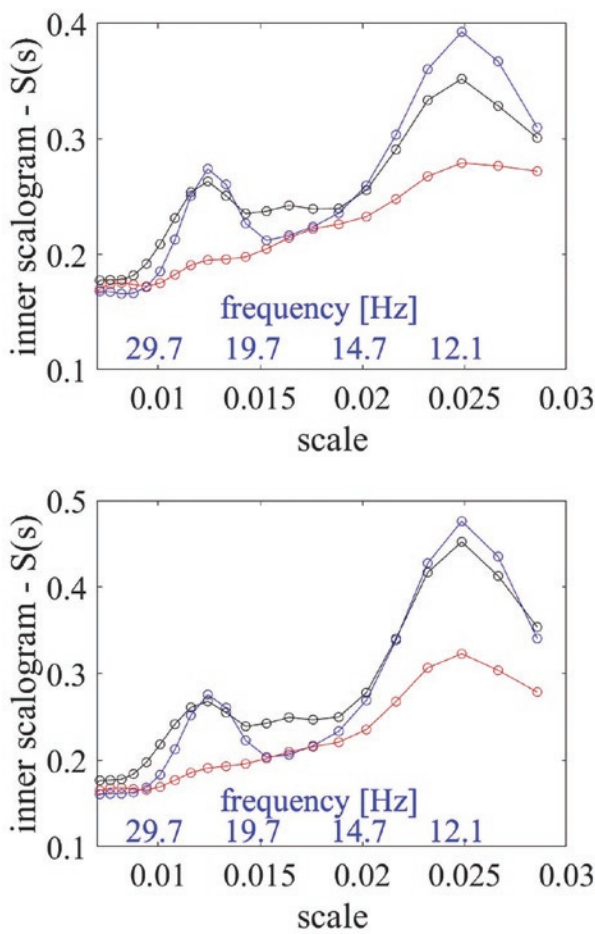


Fig. 8. Results of Scale Index Analysis (SIA) for the acceleration time courses obtained from the x-axis (top) and y-axis (bottom). Colours of the lines denote the following RIC values: blue -  $20\mu\text{m}$ , red -  $22\mu\text{m}$ , black -  $34\mu\text{m}$

Table 7. Values of Scale Index

RIC [ $\mu\text{m}$ ]	x-axis	y-axis
20	0.77	0.73
22	about 1.00	about 1.00
34	0.89	0.89

The described observations and remarks are more or less visible on the recurrence plots (Figure 9); all of them represent different kinds of patterns which are differing from each other in distribution of the recurrence points. In order to visualise them, one of the short-time series consisting of 1500 data points was taken into account. The cases of  $\text{RIC}=20$  and  $34\mu\text{m}$  represent more periodical character. The obtained recurrence points are close to each other, which means non-chaotic and ordered response. In the other case, the concentration of recurrence points has no specific structure. By almost the same percentage of RR, they are dispersed, which proves intermittences in the bearing response. The characteristic diagonal and vertical line structures clear in all recurrence plots (Figure 9) are related to periodic motion with

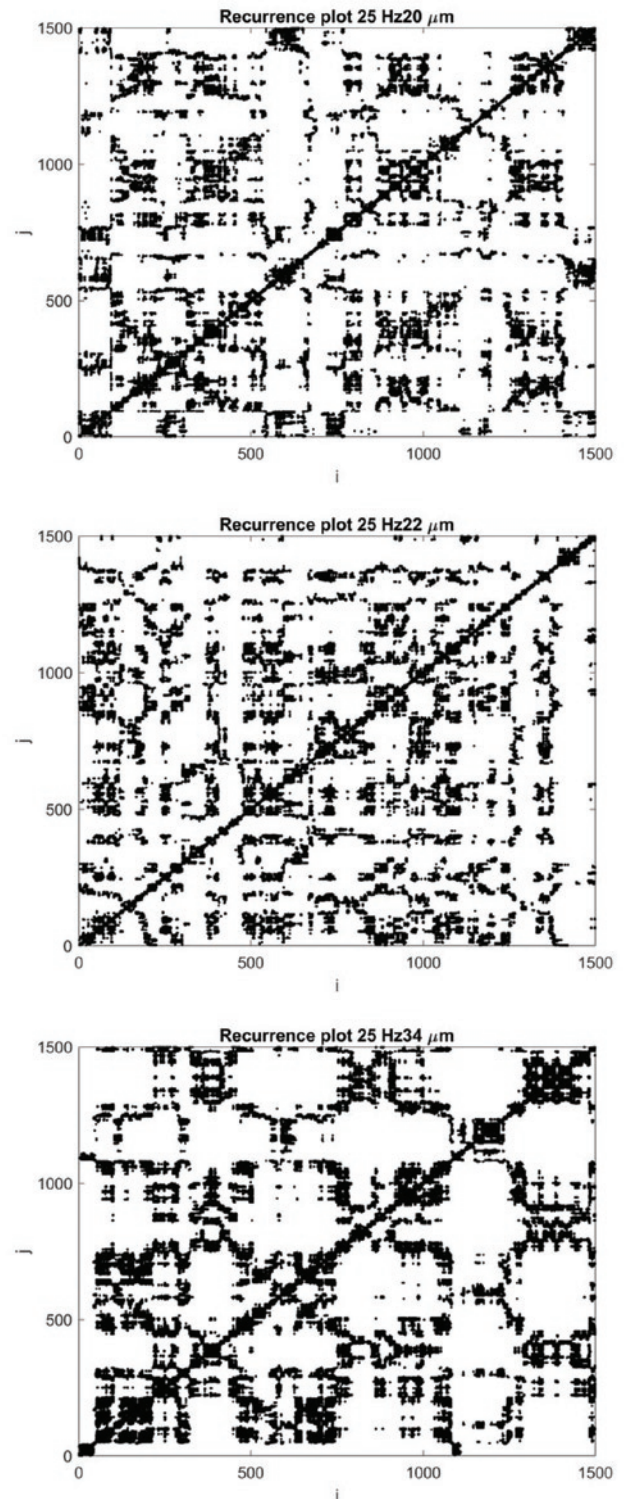


Fig. 9. Recurrence plots for  $\text{RIC}=20\mu\text{m}$  (top),  $\text{RIC}=22\mu\text{m}$  (middle) and  $\text{RIC}=34\mu\text{m}$  (bottom)

Table 8. Results of recurrence quantifiers for the considered cases derived from random short time series

RIC [ $\mu\text{m}$ ]	RR	DET	LAM	$L_{\text{MAX}}$	$V_{\text{MAX}}$
20	0.045	0.81	0.89	50	31
22	0.045	0.76	0.86	53	21
34	0.078	0.87	0.92	72	42

the periods defined by the vertical (or horizontal) space between lines and left in the current state, respectively. On the other hand, the square like structures signal occurrence of intermittences [19].

To quantify the nonlinearity of frequently appearing intermittencies in response and multiple solutions, recurrence quantification analysis is applied. In particular, the recurrence quantifiers such as RR, DET, LAM,  $L_{\text{MAX}}$ ,  $V_{\text{MAX}}$  were calculated. The method is mostly dedicated to analysing short time series so, at first, the long signal was divided into 10 parts consisting of 1500 data points each. Next, the value of a given quantificator was calculated in each short signal, then the final value taken for the evaluation was the average of the collection of results. For the calculation of each quantificator and plotting

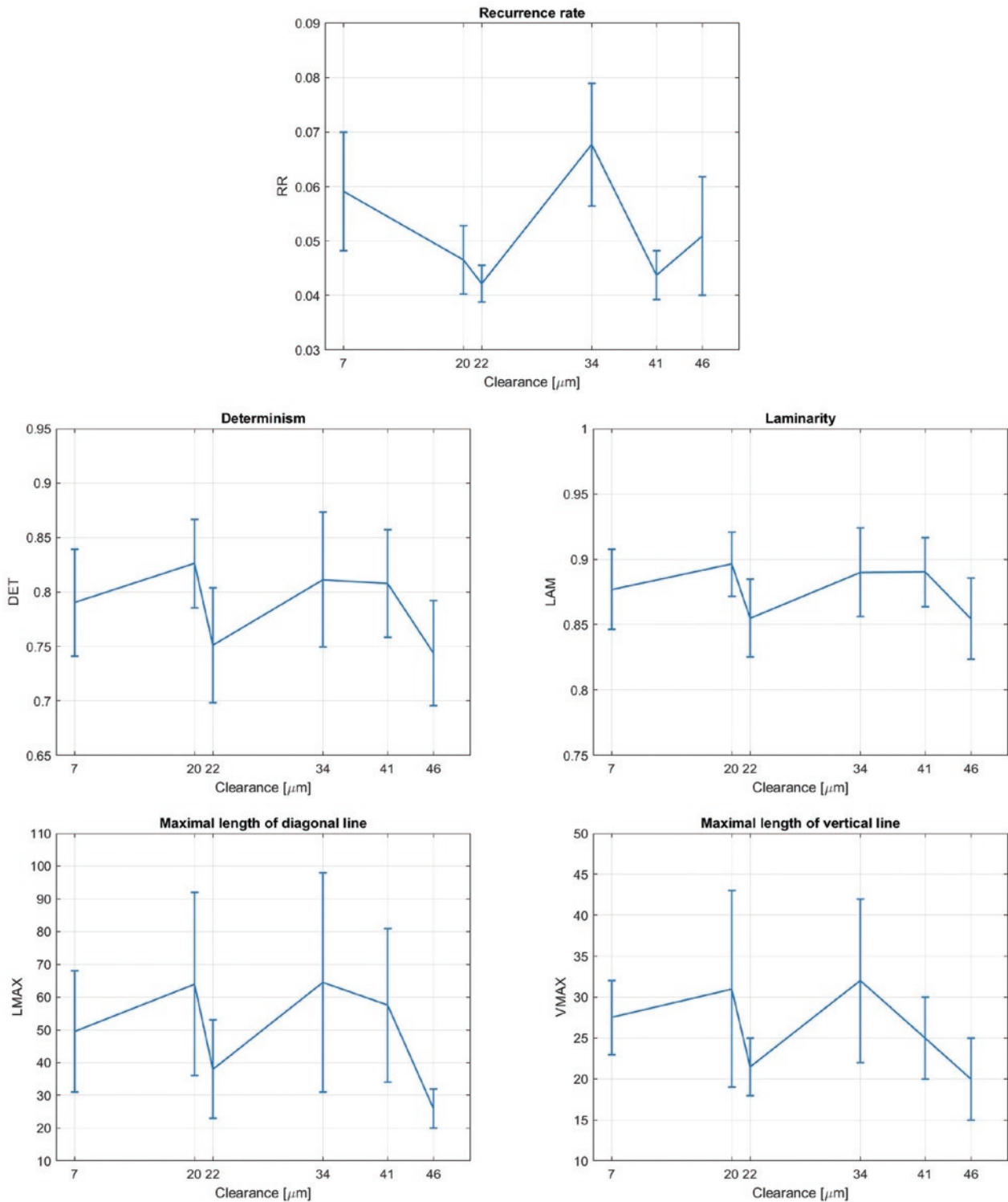


Fig. 10. Plots of the recurrence quantifiers with a marked error range versus clearance

the recurrence plots, the constant threshold value was taken as  $\epsilon=0.4$ . In Table 8, exemplary results of quantificators are collected for one random short time series. All of them are visualised on the plots with a marked error range formed out of the minimum and maximum value of each quantificator (Figure 10).

Regarding the obtained values of recurrence rate (RR), they correspond to the percentage of recurrence points obtained on the recurrence plots. The only correlation is the increase in the amount of the recurrence points with the increased RIC value by the same level of threshold.

Determinism (DET) corresponds to predictability of the system and again the averaged minimum is in  $RIC=22\mu m$ . The lower value of determinism, the system behaves more unpredictably, which proves the occurrence of the transient state in the bearing. The error value varies almost in the same range for all clearances.

The next considered quantificator is laminarity (LAM). Similarly to determinism, its averaged minimum value is in  $RIC=22\mu m$ . Laminarity corresponds to mean time when the state of the system is "trapped" for some state and indicates a switching between different states of the system.

The  $L_{MAX}$  increases with the RIC value. For the middle one, however, it can be observed that it changes in a relatively narrow range against the neighbouring. Variability of the length of diagonal lines does not clearly inform about a specific bearing clearance and does not precisely define the transient state for the clearance of  $22\mu m$ .

In the case of  $V_{MAX}$ , the length of vertical lines refers to small variations in recurrences and isolated points mean large changes. Regarding vertical lines length, there is the same values distribution as for laminarity and determinism. Moreover, the values for  $RIC=22\mu m$  change in low and narrow ranges, showing its non-linear origin transition into optimal bearing operation conditions.

## 5. Discussion and conclusion

In this paper, the tests on a double row self-aligning ball bearing with different values of radial internal clearance. The analysis of its acceleration response was evaluated with the statistical indicators, the FFT, the CWT and the recurrence-based methods. It should be emphasized, that the research system was not damaged, but with a changed operational parameter having a strong influence on its dynamic behaviour. The initial analysis of the bearing response with the statistical indicators allowed us to find the transient point in the bear-

ing characterized by one value of RIC. With the FFT and CWT, it was possible to find the characteristic frequencies present in the spectra and visualize them in time domain. The results of Scale Index proved the differences in the disturbance of the vibrations in the bearings operation between the acceleration signals obtained for the different values of radial clearance. The application of the selected recurrence quantificators and recurrence plots supported the analysis and showed the unstable behaviour of the bearing. In the performed, two specific limits of bearing behaviours can be distinguished. The first one is a frictional solution for fairly small RIC values where at small distance between rolling surfaces introduces friction. It can be connected with an undesirable skidding phenomenon in ball bearings. In the second limit, an uninhibited movement of rolling elements with contact loose and impacts is obtained. On the other hand, the point for  $RIC=22\mu m$  is the range between them, where the frequency response is changing in terms of kurtosis (Table 5 and 6) by indicating vicinity to intermittency and also in the recurrence plot (Figure 9), where the multi-scales in self-organization of the black area is stronger for  $RIC=22\mu m$ .

To sum up the four different methods allow for identifying and visualize transient states occurring in the ball bearing by different values of radial internal clearance. The state can be related to a negative phenomenon called skidding and such an unexpected behaviour was observed on the plots of all applied recurrence quantificators. The recurrence rate (Figure 10) shows that the repetitions of the states are strongly comparable to the case of  $RIC=22\mu m$  to the other neighbouring cases. Simultaneously, determinism and laminarity show a clear decrease to signal relative decrease of periodicity and the slower changes. Furthermore, the maximal diagonal and vertical lengths reductions express the unpredictability (chaoticity or stochasticity) of the optimal solution of  $RIC=22\mu m$ . Note that the other decreases of the quantificators may not correspond to the optimal choice of clearance for the other reasons as durability of bearings.

The applied methods have the potential and further steps in research will consider testing at different rotational velocities, for longer time series and on a greater number of tapered bore self-aligning ball bearings to set a specific bearing clearance in the same bearing. More complex tests would increase the power of the test and introduce the new methods (CWT, SIA and Recurrences) into diagnostics of the dynamic behaviour of bearings' dynamic behaviour. The performed tests showed more or less the range of optimal clearance can be neither small nor big to avoid spalling or pitting.

## References

1. Al-Badour F, Sunar M, Cheded L. Vibration analysis of rotating machinery using time-frequency analysis and wavelet techniques. *Mechanical Systems and Signal Processing* 2011; 25(6): 2083-2101, <https://doi.org/10.1016/j.ymssp.2011.01.017>.
2. Ambrozkiewicz B, Meier N, Guo Y, Litak G, Georgiadis A. Recurrence-based diagnostics of rotary systems. *IOP Conference Series: Materials Science and Engineering* 2019; 710 012014, <https://doi.org/10.1088/1757-899X/710/1/012014>
3. Bechhoefer E, Qu Y, Zhu J, He, D. Signal processing techniques to improve and acoustic emissions sensor. *Annual conference of the prognostics and health management* 2013; 4: 1-8.
4. Benitez R, Bolos V J, Ramirez R. A wavelet-based tool for studying non-periodicity. *Computers and Mathematics with Applications* 2010; 60: 634-641, <https://doi.org/10.1016/j.camwa.2010.05.010>.
5. Bolos V J, Benitez R, Ferrer R. A new wavelet tool to quantify non-periodicity of non-stationary economic time series. *Mathematics* 2020; 8(5), <https://doi.org/10.3390/math8050844>.
6. Castilla-Gutierrez J, Fortes J C, Pulido-Calvo I. Analysis, evaluation and monitoring of the characteristic frequencies of pneumatic drive unit and its bearing through their corresponding frequency spectra and spectral density. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2019; 21(4): 585-591, <https://doi.org/10.17531/ein.2019.4.7>.
7. Chudzik A, Warda B. Effect of radial internal clearance on the fatigue life of the radial cylindrical roller bearing. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2019; 21(2): 211-219, <https://doi.org/10.17531/ein.2019.2.4>.
8. Climente-Alcaron V, Antonino-Daviu J A, Haavisto A, Arkkio A. Particle filter-based estimation of instantaneous for the diagnosis of electrical asymmetries in induction machines. *IEEE Transactions on Instrumentation and Measurement* 2014, 63(10): 2454-2463, <https://doi.org/10.1109/TIM.2014.2310113>.
9. D'Elia G, Cocconcelli M, Mucchi E. An algorithm for the simulation of faulted bearings in non-stationary conditions. *Meccanica* 2018, 53(4-5): 1147-1166, <https://doi.org/10.1007/s11012-017-0767-1>.
10. Decker J, Lewicki D. Spiral bevel pinion crack detection in a helicopter gearbox. *NASA/TM-2003-212327* 2003, ARL-TR-2958.
11. Decker J. Crack detection for aerospace quality spur gears. *NASA/TM-2002-211492* 2002, ARL-TR-2682.

12. Eckmann J, Oliffson K, Ruelle D. Recurrence plots of dynamical systems. *Europhysics Letters* 1987; 4(9): 973-977, <https://doi.org/10.1209/0295-5075/4/9/004>.
13. Fraser A M, Swinney H L. Independent coordinates for strange attractors from mutual information. *Physical Review A, General Physics* 1986; 33(2): 1134-1140, <https://doi.org/10.1103/PhysRevA.33.1134>.
14. Harsha S P. Nonlinear dynamic response of a balanced rotor supported by rolling element bearings due to radial internal clearance. *Mechanism and Machine Theory* 2006; 41(6): 688-706, <https://doi.org/10.1016/j.mechmachtheory.2005.09.003>.
15. Hou Z, Hera A, Noori M. Wavelet-Based Techniques for Structural Health Monitoring. *Health Assessment of Engineered Structures: Bridges, Buildings and Other Infrastructure* 2013, Chapter 7: 179-201, [https://doi.org/10.1142/9789814439022\\_0007](https://doi.org/10.1142/9789814439022_0007).
16. Huaitao S, Bai X, Zhang K, Wu Y, Wang Z. Effect of thermal-related fit clearance between outer ring and pedestal on the vibration of full ceramic ball bearing. *Shock and Vibration* 2019, 8357807, <https://doi.org/10.1155/2019/8357807>.
17. Huang H-Z, Yu K, Huang T, Li H, Qian H-M. Reliability estimation for momentum wheel bearings considering frictional heat. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2020; 22(1): 6-14, <https://doi.org/10.17531/ein.2020.1.2>.
18. Kennel M B, Brown R, Abarbanel H D I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A* 1992; 45(3403), <https://doi.org/10.1103/PhysRevA.45.3403>.
19. Klimaszewska K, Żebrowski J J. Detection of the type of intermittency using characteristic patterns in recurrence plots. *Physical Review E* 2009, 80, 026214, <https://doi.org/10.1103/PhysRevE.80.026214>.
20. Kwuimy C, Samadani M, Kappaganthu K, Nataraj C. Sequential recurrence analysis of experimental time series of a rotor response with bearing outer race faults. *Vibration Engineering and Technology of Machinery* 2015, 683-696, [https://doi.org/10.1007/978-3-319-09918-7\\_61](https://doi.org/10.1007/978-3-319-09918-7_61).
21. Kwuimy C, Samadani M, Nataraj C. Bifurcation analysis of a nonlinear pendulum using recurrence and statistical methods: applications to fault diagnostics. *Nonlinear Dynamics* 2014, 76: 1963-1975, <https://doi.org/10.1007/s11071-014-1261-0>.
22. Kwuimy C, Samadani M, Pankar P, Nataraj C. Recurrence analysis of experimental time series of a rotor response with bearing outer race faults. *ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2-5 August 2015, Boston, USA, <https://doi.org/10.1115/DETC2015-48106>.
23. Litak G, Gajewski J, Syta A, Jonak J. Quantitative estimation of the tool wear effects in a ripping head by recurrence plots. *Journal of Theoretical and Applied Mechanics* 2008; 46(3): 521-530.
24. Litak G, Sawicki J T, Kasperek R. Cracked rotor detection by recurrence plots. *Nondestructive Testing and Evaluation* 2009; 24(4): 347-351, <https://doi.org/10.1080/10589750802570836>.
25. Litak G, Syta A, Gajewski J, Jonak J. Detecting and identifying non-stationary courses in the ripping head power consumption by recurrence plots. *Meccanica* 2010; 45(4): 603-608, <https://doi.org/10.1007/s11012-009-9265-4>.
26. Litak G, Syta A, Rusinek R. Dynamical changes during composite milling: recurrence and multiscale entropy analysis. *The International Journal of Advanced Manufacturing Technology* 2011; 56(5): 445-453, <https://doi.org/10.1007/s00170-011-3195-8>.
27. Liu Y, Zhang Y, Wu Z. Stochastic dynamic analysis of the rotor-bearing system considering the randomness of the radial clearance. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 2019; 41: 529, <https://doi.org/10.1007/s40430-019-2038-7>.
28. Loutridis S J. Gear failure prediction using multiscale local statistics. *Engineering Structures* 2008; 30(5): 1214-1223, <https://doi.org/10.1016/j.engstruct.2007.07.022>.
29. Marwan N, Romano M C, Thiel M, Kurths J. Recurrence plots for the analysis of complex systems. *Physics Reports* 2007; 438(5-6): 237-329, <https://doi.org/10.1016/j.physrep.2006.11.001>.
30. Meier N, Georgiadis A. Automatic assembling of bearings including clearance measurement. *Procedia CIRP* 2016; 41: 242-246, <https://doi.org/10.1016/j.procir.2015.12.110>.
31. Meier N, Papadoudis J, Georgiadis A. Intelligent software system for replacing a force sensor in the case of clearance measurement. *Procedia CIRP* 2019; 79: 517-522, <https://doi.org/10.1016/j.procir.2019.02.103>.
32. Miskovic Z, Mitrovic R, Stamenic Z. Analysis of grease contamination influence on the internal radial clearance of ball bearings by thermographic inspection. *Thermal Science* 2016; 20(1): 255-265, <https://doi.org/10.2298/TSCI150319083M>.
33. Mitrovic R M, Atanasovska I D, Soldat N D, Momcilovic D B. Effects of operation temperature on thermal expansion and main parameters of radial ball bearings. *Thermal Science* 2015; 19(5): 1835-1844, <https://doi.org/10.2298/TSCI141223091M>.
34. Saribulut L, Teke A, Tumay M. Fundamentals and literature review of Fourier transform in power quality issues. *Journal of Electrical and Electronics* 2013; 5(1): 9-22, <https://doi.org/10.5897/JEEER2013.0436>.
35. Sawalhi N, Randall R B, Endo H. The enhancement of fault detection and diagnosis in rolling element using minimum entropy deconvolution combined with spectral kurtosis. *Mechanical Systems and Signal Processing* 2007; 21(6): 2616-2633, <https://doi.org/10.1016/j.ymssp.2006.12.002>.
36. Sen A K, Litak G, Edwards K D, Finney C E A, Daw C S, Wagner R M. Characteristics of cyclic heat release variability in the transition from spark ignition to HCCI in a gasoline engine. *Applied Energy* 2011; 88(5): 1649-1655, <https://doi.org/10.1016/j.apenergy.2010.11.040>.
37. Sharma A, Upadhyay N, Kankar P K, Amarnath M. Nonlinear dynamic investigations on rolling element bearings: A review. *Advances in Mechanical Engineering* 2018; 10(3): 168781401876414, <https://doi.org/10.1177/1687814018764148>.
38. Sharma V, Parey A. A review of gear fault diagnosis using various condition indicators. *Procedia Engineering* 2016; 144: 253-263, <https://doi.org/10.1016/j.proeng.2016.05.131>.
39. Syta A, Jonak J, Jedliński Ł, Litak G. Failure diagnosis of a gear box by recurrences. *Journal of Vibration and Acoustics* 2012; 134(4), <https://doi.org/10.1115/1.4005846>.
40. Takens F. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics* 1981; 898: 366-381, <https://doi.org/10.1007/BFb0091924>.
41. Tiwari M, Gupta K, Prakash O. Effect of radial internal clearance of a ball bearing on the dynamics of a balanced horizontal rotor. *Journal of Sound and Vibrations* 2000; 238(5): 723-756, <https://doi.org/10.1006/jsvi.1999.3109>.
42. Tomovic R. A simplified mathematical model for the analysis of varying compliance vibrations of a rolling bearing. *Applied Sciences* 2020; 10(2): 670, <https://doi.org/10.3390/app10020670>.
43. Torrence C, Compo G P. A practical guide to Wavelet Analysis. *Bulletin of the American Meteorological Society* 1998, 79(1): 61-7, [https://doi.org/10.1175/1520-0477\(1998\)079<0061:APGTWA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2).

44. Wang J, Xu M, Zhang C, Huang B, Gu F. Online bearing clearance monitoring based on an accurate vibration analysis. *Energies* 2020; 13(2): 389, <https://doi.org/10.3390/en13020389>.
45. Wang W. Early detection of gear tooth cracking using the resonance demodulation technique. *Mechanical Systems and Signal Processing* 2001; 15(5): 887-903, <https://doi.org/10.1006/mssp.2001.1416>.
46. Webber C L, Zbilut J P. Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology* 1994; 76(2): 965-973, <https://doi.org/10.1152/jappl.1994.76.2.965>.
47. Webber C, Marwan N. Mathematical and computational foundations of recurrence quantifications. *Understanding Complex Systems* 2015, Chapter 1: 3-43, [https://doi.org/10.1007/978-3-319-07155-8\\_1](https://doi.org/10.1007/978-3-319-07155-8_1).
48. Xu M, Feng G, He Q, Gu F, Ball A. Vibration characteristics of rolling element bearings with different radial clearances for condition monitoring of wind turbine. *Applied Sciences* 2020; 10(14): 4731, <https://doi.org/10.3390/app10144731>.
49. Yakout M, Nassef M, Backar S. Effect of clearances in rolling element bearings on their dynamic performance, quality and operating life. *Journal of Mechanical Science and Technology* 2019; 33: 2037-2042, <https://doi.org/10.1007/s12206-019-0406-y>.
50. Zbilut J P, Webber C L. Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A* 1992; 171(3-4): 199-203, [https://doi.org/10.1016/0375-9601\(92\)90426-M](https://doi.org/10.1016/0375-9601(92)90426-M).
51. Zhang X, Kang J, Zhao J, Cao D. Features for fault diagnosis and prognosis of gearbox. *Chemical Engineering Transactions* 2013; 33: 1027-1032, <https://doi.org/10.3303/CET1333172>.
52. Zhuo Y, Zhou X, Yang C. Dynamic analysis of double-row self-aligning ball bearings due to applied loads, internal clearance, surface waviness and number of balls. *Journal of Sound and Vibration* 2014; 333(23): 6170-6189, <https://doi.org/10.1016/j.jsv.2014.04.054>.





Article citation info:

Kępczak N, Bechciński G, Rosik R. Experimental verification of the deep hole boring bar model. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 55–62, <http://dx.doi.org/10.17531/ein.2021.1.6>.

Indexed by:



## Experimental verification of the deep hole boring bar model

Norbert Kępczak<sup>a\*</sup>, Grzegorz Bechciński<sup>a</sup>, Radosław Rosik<sup>a</sup>

<sup>a</sup>Institute of Machine Tools and Production Engineering, Faculty of Mechanical Engineering, Lodz University of Technology, ul. Stefanowskiego 1/15, 90-924 Lodz, Poland

### Highlights

- In the article theoretical and experimental tests of deep hole boring bar were performed.
- During tests of the dynamic properties divergency in the results values was below 8%.
- During tests of the static properties divergency in the results values was below 3%.
- High convergence of the results of numerical and experimental research was obtained.

### Abstract

The article presents the results of the experimental verification of the deep hole boring bar tool model. The aim of the work was to obtain a verified boring bar tool model, which in further scientific research will be a starting point for creating a prototype of a tool with a new design, in which dynamic properties will be improved. The research was divided into two stages. In the first stage, modal studies of the model and the real object were carried out. The obtained discrepancy between numerical and experimental results below 8% allows to state that the model is characterized by dynamic properties occurring in the real boring bar. In the second stage of the research, static tests were carried out. The object was loaded with forces of 98.6 N, 195.0 N, 293.8 N. The obtained slight discrepancy in the results of numerical and experimental tests below 3% allows to state that the model reflects the static properties of the real boring bar. The high convergence of the theoretical and experimental results allows for the conclusion that the numerical model has been verified positively.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

boring bar, modal analysis, FEM, dynamic properties, static properties

## 1. Introduction

Various types of vibrations occur during the machining process. Vibrations that occur between the workpiece and the tool, even during undisturbed machining operations, are very important. Vibrations have a negative effect on the quality of the treated surface and can damage the machining center or the tool. Vibrations occurring during the cutting of the processed material can be divided into forced and self-excited [25, 32]. Forced vibrations arise as a result of a periodically acting excitation force. Self-excited vibrations, in contrast to forced vibrations, are not caused by external disturbances, but by the dynamic interaction between the mechanical system and the machining process [3, 20, 29].

Many studies have developed various approaches to reduce the occurrence of vibration in the cutting process. Active prevention of vibrations (damping) in the turning process is one of the possibilities of improving the stability of the machining process. One of the examples is the use of a mainly sinusoidal or triangular spindle speed change signal [2]. Another research approach is focused on the selection of parameters related to machining, such as feed, cutting speed or depth of cut [19]. In addition, directly process-oriented approaches include solutions with an additional damping system that is installed on the machine tool. The distinguishing feature of such a system is the need

to connect an external energy source and adapt it to the existing conditions of the machine tool.

The increase in damping by passive auxiliary systems is based on the conversion of the oscillation energy into another type of energy, e.g. thermal energy. For this purpose, special tools with multi-part friction dampers integrated in the central bore of the tool are developed [11, 14, 26, 29]. Moreover, there are solutions in which the damping is realized by an additional freely swinging mass [28]. Another known passive auxiliary system is an auxiliary damper that uses the damping properties of the material (polymers, elastomers). This solution is used, for example, in boring bars [29] and turning tools [31].

The auxiliary system measures the system vibrations with sensors, processes the data with the controller, and then reduces it by introducing it into the anti-vibration system. Anti-vibrations are generated by an inductor. Active energy is always supplied to the system from outside.

A completely different approach to reducing vibration is to influence the stability of the tool and machining center by means of dynamic compliance. Equation (1) for the frequency response of the transition function is based on this approach [29]:

(\*) Corresponding author.

E-mail addresses: N. Kępczak - [norbert.kepczak@p.lodz.pl](mailto:norbert.kepczak@p.lodz.pl), G. Bechciński - [grzegorz.bechcinski@p.lodz.pl](mailto:grzegorz.bechcinski@p.lodz.pl), R. Rosik - [radoslaw.rosik@p.lodz.pl](mailto:radoslaw.rosik@p.lodz.pl)

$$G(i\omega) = \frac{1}{c} \frac{1}{1 + 2iD \frac{\omega}{\omega_n} - \left(\frac{\omega}{\omega_n}\right)^2} \quad (1)$$

Value ( $\omega = \omega_n$ ) causes occurring a resonance – equation (2).

$$|G(\omega = \omega_n)| = \frac{1}{2D} = \frac{\sqrt{m}}{d} \quad (2)$$

Than equation (3) leads to equation (4).

$$\sqrt{\frac{m}{c}} = \frac{1}{\omega_n} \quad (3)$$

$$|G(\omega = \omega_n)| = \frac{1}{d\omega_n} \quad (4)$$

where:

- $c$  – static stiffness ;
- $d$  – damping ratio;
- $D$  – degree of damping;
- $G(i\omega)$  – resilience frequency response;
- $m$  – mass;
- $\omega$  – forced angular frequency;
- $\omega_n$  – natural angular frequency.

This approach shows that if the smaller ratio of the mass „m” to the static stiffness „c”, than the higher angular frequency of the natural vibrations and if the higher damping ratio „d”, than the lower value of the resilience frequency response „G”, which results a lower amplitude value at resonance.

It is important to precisely determine the frequency values where the phenomenon of resonance occurs. Currently, at the design stage, numerical analyzes can be made. The designer receives the forms of natural vibrations and their frequencies [12]. However, the numerical model is always the ideal model with no disadvantages. In order to confirm the theoretical results, the so-called an identification experiment that allows for full verification of the numerical model. This article presents the comparative results of the experimental verification of the deep hole boring bar model. Getting to know the static and dynamic properties of the real boring bar is very important, because the research is preliminary in a project aimed at designing and producing a prototype of a deep hole boring bar tool with increased dynamic properties.

Table 1. Mechanical properties of applied materials\*

	Carbon steel	High-strength steel
Tensile strength [MPa]	420.00	448.00
Yield point [MPa]	350.00	275.80
Young's modulus [GPa]	200.00	200.00
Poisson's ratio [---]	0.29	0.29
Kirchhoff's modulus [MPa]	79700.00	128700.00
Density [g/cm <sup>3</sup> ]	7.85	7.85

\*mechanical properties are listed according to the Autodesk Inventor 2020 Material Library

## 2. Dynamic properties

### 2.1. Numerical analysis

Theoretical modal analysis (TMA) is defined as the matrix self problem dependent on the mass, stiffness and damping factors. TMA requires solving an own problem for the adopted structural model. It is a commonly used practice technique for examining the dynamic properties of mechanical objects. It is used not only in mechanical engineering [10, 15, 16, 33, 34] or space engineering [30], but also successfully used in building structures [24], agriculture [8] and rail transport [17]. The sets of frequencies, damping coefficients and the modes of free vibrations allow for the simulation of the behavior of the structure with any input, selection of controls, structure modification and in other situations. It is used in the design process when it is not possible to carry out research on a real object [6, 9, 18, 21].

TMA in CAD programs consists in determining the dynamic properties of the tested structure on the basis of a structural model. The modal model of a virtual structure is determined on the basis of e.g. the finite element method (FEM). Thanks to this method, a wide range of dynamic tests can be carried out on the basis of the digital prototype without the need to build an experimental stand equipped with a real research object.

TMA performed with the Autodesk Inventor simulation module enables the obtaining of the natural mode shapes and the corresponding natural frequencies. In the analytical procedure, the software developers did not take into account information on the values of the damping coefficients for individual vibration modes. The module can be used to identify the modes of natural vibrations and their frequencies, which in theoretical analysis is sufficient at the initial stage of construction assessment. The full set of modal information (modes, frequencies, damping coefficients) can be obtained on the basis of experimental modal analysis. Due to these software limitations regarding TMA, this paper compares the results obtained on the basis of theoretical and experimental analysis, including the modes of vibrations and their frequencies.

In order to conduct numerical tests, first a three-dimensional model of the test stand was created, which is presented in Figure 1. The model consists of: (1) base, (2) mounting holes, (3) boring bar head, (4) shank, (5) support, (6) mounting bolts. A model of the real PAFANA boring bar from the Smart Head System series was prepared for the tests. This boring bar is a modular tool, which includes a head marked K40-MWLN/L08 and a shank marked A40-K40 300.

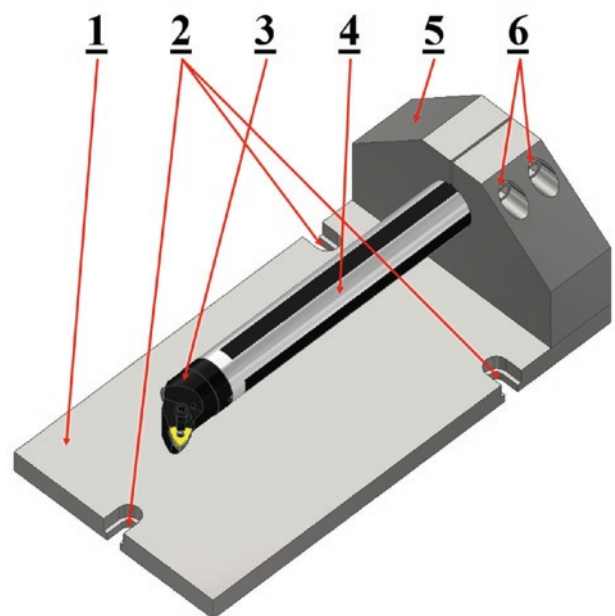


Fig. 1. Three-dimensional model of the test stand: 1 – base, 2 – mounting holes, 3 – boring bar head, 4 – shank, 5 – support, 6 – fixing screws

Next, all elements of the test stand were assigned appropriate material properties. The base and the bracket are made of carbon steel, while the boring bar is made of high-strength steel. Table 1 shows the mechanical properties for the materials used.

Then, the degrees of freedom were obtained in the places where there were mounting holes in the base (Figure 1) and a finite element mesh was implemented (Figure 2). The mesh was defined in such a way that 266329 nodes and 172856 elements were created. TMA was carried out in the frequency domain up to 3200 Hz.

**Nodes: 266329**  
**Elements: 172856**



Fig. 2. Finite element mesh view

## 2.2. Experimental research

Experimental modal analysis (EMA) is a technique that is often used in practice to investigate the dynamic properties of mechanical objects, both at the stage of designing and in operating machines. In contrast to the operational modal analysis [13], the identification experiment in the EMA consists in forcing the vibrations of the object with simultaneous measurement of the exciting force and the response of the system, most often in the form of a spectrum of vibration acceleration [6, 9, 18, 21]. One of the many applications of modal analysis is also the detection of damage to mechanical objects [35]. Apart from mechanical systems [21], experimental modal analysis is widely used in building structures [4], electronics and electrical engineering [5] and music [7].

The EMA procedure can be performed by the SISO (single input single output), SIMO (single input multiple outputs) and MIMO (multiple inputs multiple outputs) method. These methods differ not only in the requirements of measurement data acquisition systems, but also in the requirements concerning the purpose of the research and the accuracy of the analysis results. Due to hardware limitations, the SISO procedure was used in this study. Figure 3 shows the test stand with the apparatus for conducting the EMA, which includes: (1) frame, (2) modal hammer, (3) base, (4) support, (5) mounting screws, (6) accelerometer (7) boring bar, (8) data acquisition module and (9) computer with software.

For measuring and data acquisition, the PULSE Lite system from Brüel & Kjær was used, which includes:

- Piezoelectric accelerometer 4514 53958, with basic parameters [22].
- Modal hammer 8206-003 54990, with basic parameters [23].
- Data acquisition system 3560-L.

In addition, the manufacturer provides PULSE LabShop software enabling the registration and processing of collected data. The soft-

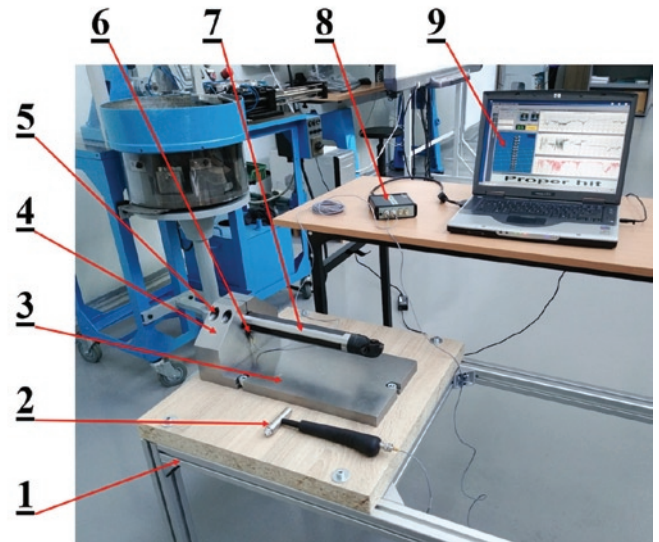


Fig. 3. The test stand: 1 – frame, 2 – modal hammer, 3 – base, 4 – support, 5 – mounting screws, 6 – accelerometer, 7 – boring bar, 8 – data acquisition module, 9 – computer

ware allows to make a Fast Fourier Transform (FFT) of collected data.

In the procedure of preparing the measurement set, a very important step is the appropriate setting of the impulse input signal level, weighting of the impulse input signal and weighting of the system response to the input excitation. The setting of the input signal level is performed by registering a certain number of repetitions of the impulse input, and then averaging the input force value. Forcing signal weighting consists in registering a specified number of excitation repetitions and passing through a “transient” window (close to rectangular) in order to register the excitation signal when the excitation reaches its maximum value. The system’s response signal to a given excitation is weighting by registering a specific number of repetitions of the system’s response to a given excitation and passing these signals through an exponential window, which allows the signal to be registered in its full length. The purpose of using correction windows is to minimize environmental noise and improve the quality of measurement signals.

After making the settings in the program, it was necessary to define the shape of the test object in the Pulse Lite program (Figure 4), as well as indicate the places where the excitations were performed (green and black hammers) and the place of mounting the accelerometer (red arrow). Due to the limited possibilities of spatial modeling in Pulse Lite, the shape of the boring bar was modeled approximately as a cylinder. There were 10 measuring points on the boring bar, which were excited to vibrations three times in succession. The accelerometer was placed on the boring bar as close as possible to the support. The boring bar was tested three times. The study was carried out in the frequency domain up to 3200 Hz. The sampling rate was 6400 Hz, while the time of the recording signal was 1 s.

## 2.3. Results and Discussion

As a result of the TMA in Autodesk Inventor, 8 modes of free vibrations were obtained in the tested frequency range. The frequency values for individual modes of free vibrations are presented in Table 2.

Only two of the generated modes of vibrations affected the boring bar directly (Mode 1 and 5). The remaining modes were associated with the vibrations of the other elements of the test stand.

As a result of the EMA, waveforms of the vibration amplitude as a function of frequency were obtained. An exemplary waveform with marked resonances is presented in Figure 5.

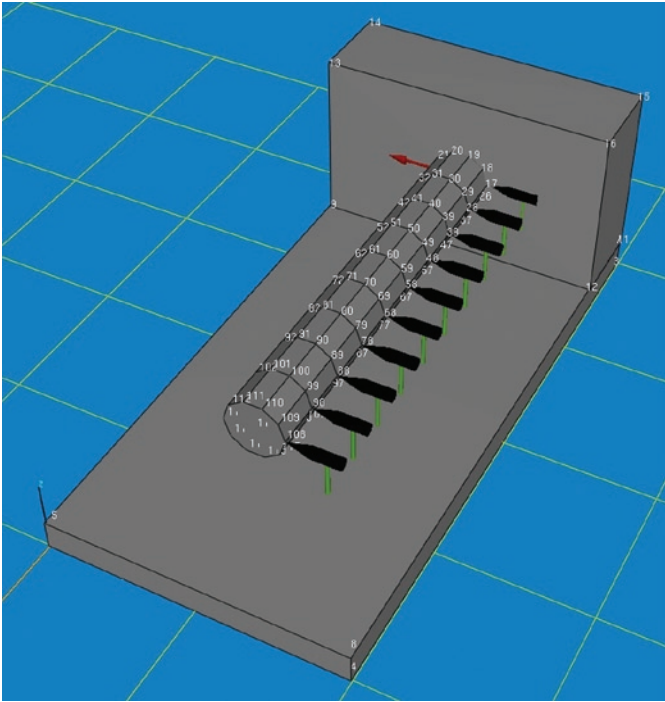


Fig. 4. The approximate geometric model of the test stand, points excited to vibrations, and the location of the displacement sensor in Pulse Lite

Table 2. Results of theoretical modal analysis

Mode	1	2	3	4	5	6	7	8
Frequency [Hz]	301	922	1685	1711	1769	2200	2473	2660

In the analyzed measuring range, the presence of three resonance peaks directly related to the vibrations of the boring bar can be clearly observed. The first resonance appeared at a frequency of 231 Hz. It was noticed that for this frequency all excited points were displaced in the same direction. This is called rigid body mode in which the entire tested object is displaced without deformation [27]. Due to the fact that the excited points were arranged in a straight line instead of in the shape of a characteristic wave, this frequency was rejected in further analysis. In the course of the amplitude of the transition function, for low frequency values, there are also smaller resonance peaks, which probably come from the vibrations of the table on which the station is mounted. In order to avoid frame vibrations, the stand should be attached, e.g. directly to the ground. In further analysis, the vibrations of the stand frame were considered insignificant and omitted.

The second resonance appeared at a frequency of 327 Hz. In this case, the excited points formed one characteristic half-wave. Figure 6 shows a comparison of the obtained forms during numerical and experimental analysis.

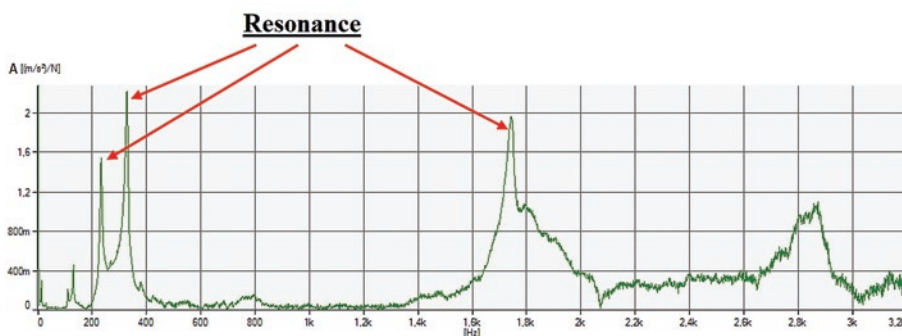


Fig. 5. An example of the amplitude waveform of the transition function

For the numerical analysis, the value of the natural frequency of the boring bar was 301 Hz, while for the experimental tests this value was 327 Hz. In the analyzed case, the divergence of the obtained results was 7.95%.

The third resonance occurred at a frequency of 1741 Hz. In this case, the excited points formed two characteristic half-waves. Figure 7 shows a comparison of the obtained forms during numerical and experimental analysis.

For numerical analysis, the value of the free vibration frequency of the boring bar was 1769 Hz, while for experimental tests this value was 1741 Hz. In the analyzed case, the divergence of the obtained results was 1.61%.

The obtained divergence of results below 8% allows to state that the model faithfully reflects the nature of the dynamic properties of the real boring bar.

### 3. Static properties

#### 3.1 Numerical analysis

Static tests were carried out to determine the theoretical static strength of the boring bar. The tests were carried out for three values of the static load with the masses 10 kg, 20 kg, 30 kg. The values of the forces were estimated as sufficient to determine the static stiffness. The actual weights were prepared for the tests, which are presented in Figure 9. Then each of the weights was weighed three times with a digital dynamometer to determine the mean force of gravity. Table 3 shows the results of converting the load into the value of gravity.

In order to carry out numerical tests of the static properties of the boring bar, the tool was placed in the opposite direction to that during dynamic tests, so that the acting force of gravity acted inside the actual test stand (Figure 11). Next, all the elements of the test stand were assigned appropriate material properties in ac-

Table 3. Converting the load value to the value of gravity

Mass [kg]	10	20	30
Measurement 1 [N]	98.7	194.8	294.0
Measurement 2 [N]	98.6	195.0	293.7
Measurement 3 [N]	98.6	195.1	293.8
Average force of gravity [N]	98.6	195.0	293.8

cordance with Table 1. In order to fix the boring bar in the test stand model, the degrees of freedom were removed in places where there were mounting holes in the stand base and a finite element mesh was imposed. The mesh was defined in such a way that 271271 nodes and 175984 elements were created. At the very end, appropriately calculated gravity forces were applied (Table 3). The force was applied to the cylindrical face of the boring bar head. Figure 8 shows the view of the numerical model with the finite element mesh and the loading force.

#### 3.2. Experimental research

In order to carry out experimental tests of the static properties of the boring bar on the test stand, three weights were prepared with which it was properly loaded. Figure 9 shows the view of the prepared weights.

Then the boring bar was mounted on the test stand and properly loaded. The displacement of the boring bar head was measured using two displacement sensors. One of the sensors was placed directly above the boring bar head,

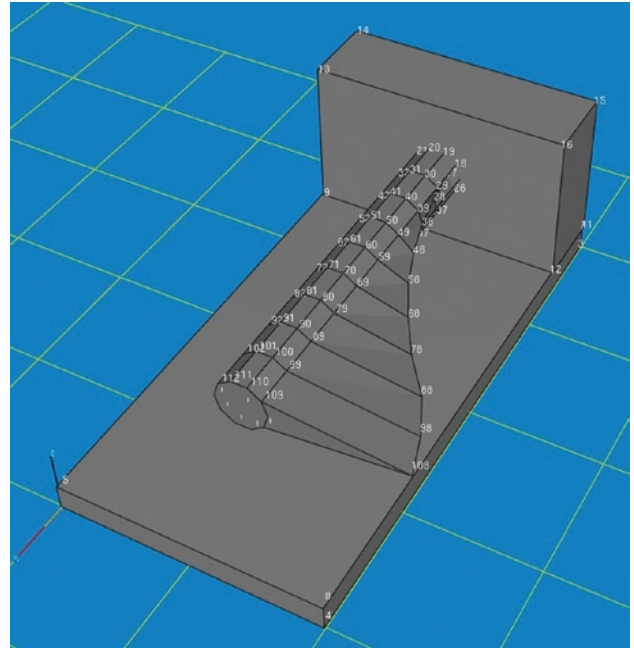
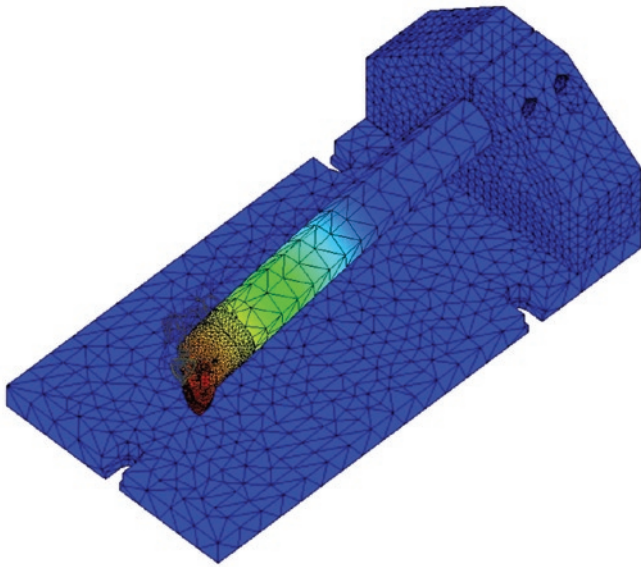


Fig. 6. Comparison of the first mode of free vibration of the boring bar obtained as a result of: a) numerical analysis for the frequency of 301 Hz, b) experimental modal analysis for the frequency of 327 Hz

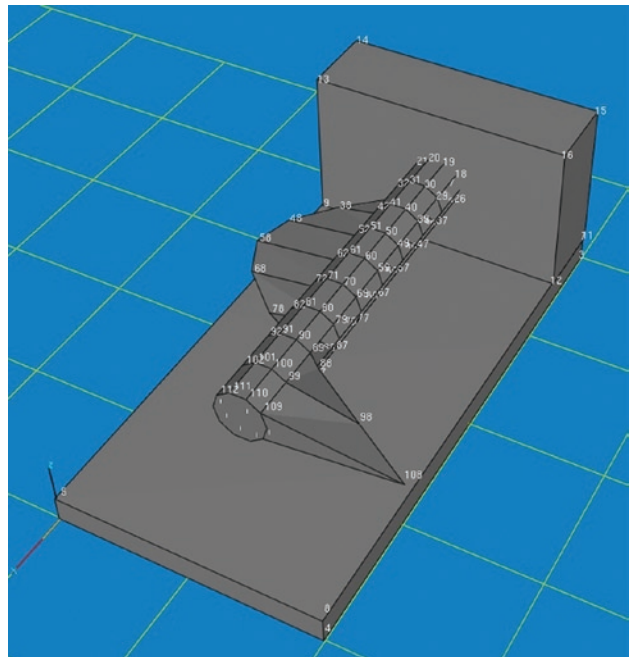
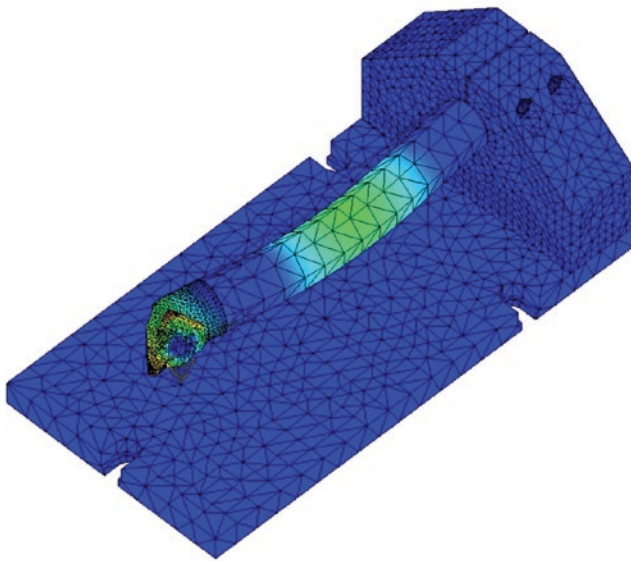


Fig. 7. Comparison of the second mode of free vibration of the boring bar obtained as a result of: a) numerical analysis for the frequency of 1769 Hz, b) experimental modal analysis for the frequency of 1741 Hz

while the other was placed under the frame of the test stand. Figure 10 shows the view of the test stand, which includes: (1) weight, (2) displacement sensor under the frame, (3) frame, (4) base, (5) support, (6) boring bar shank, (7) a boring bar head and (8) a boring bar head displacement sensor.

### 3.3. Results and Discussion

The experimental tests were carried out in three measurement series. The results obtained are presented in Table 4.

After subtracting the displacement indications of both sensors for individual measurements, the mean value of the displacement for the load with a force of 98.6 N was 141  $\mu\text{m}$ . For a load with a force of 195.0 N, this value was 275  $\mu\text{m}$ , while for a load with a force of 293.8 N, 423  $\mu\text{m}$  was obtained.

Table 5 shows a comparison of the results obtained numerically with the experimental results, while Figure 11 shows the displacement of the boring bar head under the load of 293.8 N during numerical tests.

For the numerical analysis, the displacement value of the boring bar head under the load of 98.6 N was 138  $\mu\text{m}$ , while for the experimental tests this value was 141  $\mu\text{m}$ . In the analyzed case, the divergence of the obtained results was 2.13%. For the numerical analysis, the displacement value of the boring bar head under the load of 195.0 N was 273  $\mu\text{m}$ , while for the experimental tests this value was 275  $\mu\text{m}$ . In the analyzed case, the divergence of the obtained results was 0.73%. For the numerical analysis, the displacement value of the boring bar head under the load of 293.8 N was 411  $\mu\text{m}$ , while for the experimental

Nodes: 271271  
Elements: 175984

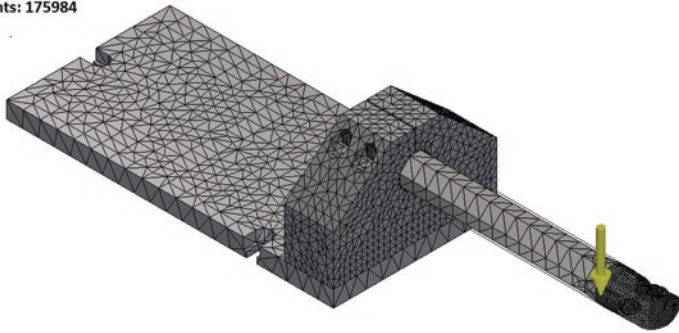


Fig. 8. View of the finite element mesh with the loading force of the test stand model with the boring bar

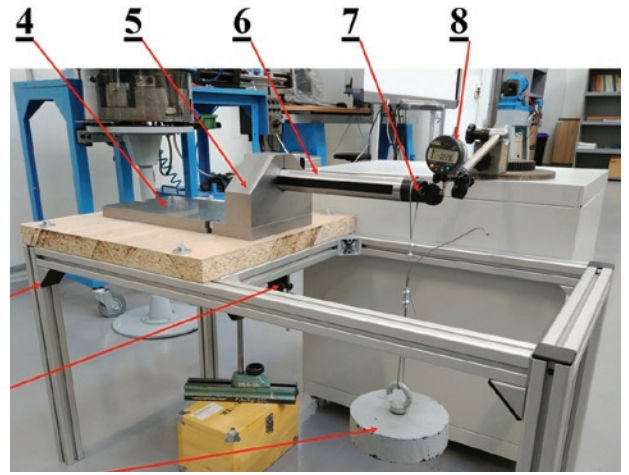


Fig. 10. View of the test stand during the loading process: 1 – weight, 2 – displacement sensor under the frame, 3 – frame, 4 – base, 5 – support, 6 – boring bar shank, 7 – boring bar head, 8 – boring bar head displacement sensor



Fig. 9. View of the prepared weights: 1 – 10 kg, 2 – 20 kg, 3 – 30 kg

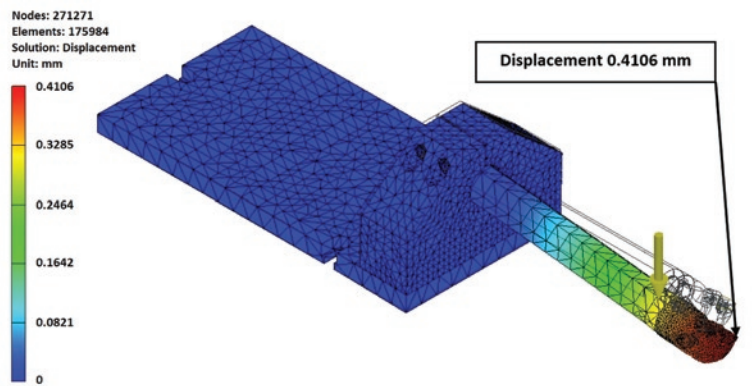


Fig. 11. The results of the numerical analysis for a force load of 293.8 N

tests this value was 423  $\mu\text{m}$ . In the analyzed case, the divergence of the obtained results was 2.84%.

The obtained divergence of results below 3% allows to state that the model faithfully reflects the static properties of the real boring bar.

#### 4. Summary and conclusions

During theoretical and experimental research, modal analyzes of deep hole boring tools were carried out. The analyzes were performed to compare the results of dynamic properties obtained by simulation

and experiment. The obtained results indicate high convergence. During the tests of the dynamic properties of the tested object for all the obtained modes of free vibrations of the boring bar, the relative error of the frequency value did not exceed 8%. During the tests of the static properties for all tested values of the boring bar loads, the relative error of the displacement value of the boring bar head did not exceed 3%. Such a high convergence of the results proves that the structure

Table 4. Results of measurements of the displacement of the boring bar and table frame under the influence of static loads

Sensor above the boring bar head										
Measurement	Load 98.6 N			Load 195.0 N			Load 293.8 N			
	Before	In process	Difference	Before	In process	Difference	Before	In process	Difference	
1	0 $\mu\text{m}$	-236 $\mu\text{m}$	236 $\mu\text{m}$	-7 $\mu\text{m}$	-445 $\mu\text{m}$	438 $\mu\text{m}$	0 $\mu\text{m}$	-691 $\mu\text{m}$	691 $\mu\text{m}$	
2	-1 $\mu\text{m}$	-218 $\mu\text{m}$	217 $\mu\text{m}$	2 $\mu\text{m}$	-443 $\mu\text{m}$	445 $\mu\text{m}$	3 $\mu\text{m}$	-679 $\mu\text{m}$	682 $\mu\text{m}$	
3	0 $\mu\text{m}$	-217 $\mu\text{m}$	217 $\mu\text{m}$	3 $\mu\text{m}$	-445 $\mu\text{m}$	448 $\mu\text{m}$	4 $\mu\text{m}$	-682 $\mu\text{m}$	686 $\mu\text{m}$	
Sensor under the frame										
Measurement	Load 98.6 N			Load 195.0 N			Load 293.8 N			
	Before	In process	Difference	Before	In process	Difference	Before	In process	Difference	
1	505 $\mu\text{m}$	590 $\mu\text{m}$	85 $\mu\text{m}$	583 $\mu\text{m}$	750 $\mu\text{m}$	167 $\mu\text{m}$	576 $\mu\text{m}$	841 $\mu\text{m}$	265 $\mu\text{m}$	
2	501 $\mu\text{m}$	585 $\mu\text{m}$	84 $\mu\text{m}$	580 $\mu\text{m}$	749 $\mu\text{m}$	169 $\mu\text{m}$	579 $\mu\text{m}$	839 $\mu\text{m}$	260 $\mu\text{m}$	
3	506 $\mu\text{m}$	585 $\mu\text{m}$	79 $\mu\text{m}$	581 $\mu\text{m}$	750 $\mu\text{m}$	169 $\mu\text{m}$	579 $\mu\text{m}$	842 $\mu\text{m}$	263 $\mu\text{m}$	
Mean value of the differences in the sensor's readings above the head and under the frame			141 $\mu\text{m}$	Mean value of the differences in the sensor's readings above the head and under the frame			275 $\mu\text{m}$	Mean value of the differences in the sensor's readings above the head and under the frame		
								423 $\mu\text{m}$		

Table 5. Comparison of the results of numerical and experimental analyzes

Load	98.6 N	195.0 N	293.8 N
Displacement – numerical analysis	138 $\mu\text{m}$	273 $\mu\text{m}$	411 $\mu\text{m}$
Displacement – experimental research	141 $\mu\text{m}$	275 $\mu\text{m}$	423 $\mu\text{m}$

of the real object was correctly represented in the numerical model and the properties of the materials used in the theoretical analysis were very precisely defined. The results also confirmed the correctly adopted method of dividing the 3D model into finite elements during meshing, as well as the method of receiving the degrees of freedom. Thanks to the experimental verification of the simulation model, it can be said that the obtained simulation results are reliable. Simula-

tions of the dynamic and static behavior of an object under different loading conditions can be successfully based on a digitally validated prototype and can be performed without further extensive experimental research.

It should also be mentioned that the mounting method also affects the dynamic and static properties of the boring bar [1]. In order to determine the dynamic and static properties of the boring bar during the machining process, additional tests should be carried out on the method of mounting the tool, e.g. in the tool holder of the machine tool.

The conducted tests are preliminary tests, the purpose of which is to design and manufacture a prototype of a deep hole boring bar tool with increased dynamic properties. To do this, it was necessary to verify the model of the real boring bar.

## References

- Akesson H, Smirnova T, Hakansson L. Analysis of dynamic properties of boring bars concerning different clamping conditions. *Mechanical Systems and Signal Processing* 2009; 23 (8): 2629-2647; [10.1016/j.ymssp.2009.05.012](https://doi.org/10.1016/j.ymssp.2009.05.012).
- Al-Regib E, Ni J, Lee S-H. Programming spindle speed variation for machine tool chatter suppression. *International Journal of Machine Tools and Manufacture* 2003; 43 (12): 1229-1240; [https://doi.org/10.1016/S0890-6955\(03\)00126-3](https://doi.org/10.1016/S0890-6955(03)00126-3).
- Bechcinski G, Ewad H, Tsiakoumis V, Pawlowski W, Kepczak N, McMillan A, Batako DL A. A Model and Application of Vibratory Surface Grinding, *Journal of Manufacturing Science and Engineering* 2018; 140 (10):101011-101011-9; <https://doi.org/10.1115/1.4040725>.
- Bień J, Krzyżanowski J, Poprawski W, Skoczynski W, Szymkowiak J. Experimental study of bridge structure dynamic characteristics using periodic excitation. *Proceedings of ISMA 2002*; 2: 555-562.
- Brecher Ch, Baumler S, Guralnik A. Experimental Modal Analysis Using a Tracking Interferometer. *CIRP Annals* 2014; 63 (1): 345-348; <https://doi.org/10.1016/j.cirp.2014.03.131>.
- Chen W H, Lu Z R, Lin W, Chen S H, Ni Y Q, Xia Y, Liao W Y. Theoretical and experimental modal analysis of the Guangzhou New TV Tower. *Engineering Structures* 2011; 33 (12): 3628-3646; <https://doi.org/10.1016/j.engstruct.2011.07.028>.
- Chomette B, Carrou J-L. Operational Modal Analysis Applied to the Concert Harp. *Mechanical Systems and Signal Processing* 2015; 56-57: 81-91; <https://doi.org/10.1016/j.ymssp.2014.10.011>.
- Ebrahimi R, Esfahanian M, Ziaei-Rad S. Vibration Modeling and Modification of Cutting Platform in a Harvest Combine by Means of Operational Modal Analysis (OMA). *Measurement* 2013; 46 (10): 3959-3967; <https://doi.org/10.1016/j.measurement.2013.07.037>.
- Falkowicz K, Ferdynus M, Dębski H. Numerical analysis of compressed plates with a cut-out operating in the geometrically nonlinear range. *Eksplotacja i Niezawodność – Maintenance and Reliability* 2015; 17 (2): 222–227; <http://dx.doi.org/10.17531/ein.2015.2.8>.
- Gagnol V, Le T P, Ray P. Modal Identification of Spindle-tool Unit in High-speed Machining. *Mechanical Systems and Signal Processing* 2011; 25 (7): 238-2398; <https://doi.org/10.1016/j.ymssp.2011.02.019>.
- Gurraj S, Guravtar Singh M, Swastik P. Improving the Surface roughness and Flank wear of the boring process using particle damped boring bars. *Materials Today: Proceedings* 2018; 5 (14): 28186-28194; <https://doi.org/10.1016/j.matpr.2018.10.062>.
- Karwat B, Rubacha P, Stańczyk E. Simulational and experimental determination of the exploitation parameters of a screw conveyor. *Eksplotacja i Niezawodność – Maintenance and Reliability* 2020; 22 (4): 741–747, <http://dx.doi.org/10.17531/ein.2020.4.18>.
- Kilikevičius A, Rimša V, Rucki M. Investigation of influence of aircraft propeller modal parameters on small airplane performance. *Eksplotacja i Niezawodność – Maintenance and Reliability* 2020; 22 (1): 1–5, <http://dx.doi.org/10.17531/ein.2020.1.1>.
- Kim N H, Won D, Ziegert J C. Numerical analysis and parameter study of a mechanical damper for use in long slender endmills. *International Journal of Machine Tools and Manufacture* 2006; 46 (5): 500-507; <https://doi.org/10.1016/j.ijmactools.2005.07.004>.
- Li B, Cai H, Mao X, Huang J, Luo B. Estimation of CNC Machine-tool Dynamic Parameters Based on Random Cutting Excitation Through Operational Modal Analysis. *International Journal of Machine Tools & Manufacture* 2013; 71: 26-40; <https://doi.org/10.1016/j.ijmactools.2013.04.001>.
- Matsuo M, Yasui T, Inamura T, Matsumura M. High-speed Test of Thermal Effects for a Machine-tool Structure Based on Modal Analysis. *Precision Engineering* 1986; 8 (2): 72-78; [https://doi.org/10.1016/0141-6359\(86\)90089-9](https://doi.org/10.1016/0141-6359(86)90089-9).
- Nangolo N F, Soukup J, Rychlikova L, Skocilas J. A Combined Numerical and Modal Analysis on Vertical Vibration Response of Railway Vehicle. *Procedia Engineering* 2014; 96: 310-319; <https://doi.org/10.1016/j.proeng.2014.12.136>.
- Ondra V, Titurus B. Theoretical and experimental modal analysis of a beam-tendon system. *Mechanical Systems and Signal Processing* 2019; 132: 55-71; <https://doi.org/10.1016/j.ymssp.2019.06.016>.
- Palanisamy P, Rajendran I, Shanmugasundaram S. Optimization of machining parameters using genetic algorithm and experimental validation for end-milling operations. *The International Journal of Advanced Manufacturing Technology* 2007; 32 (7-8): 644-655; [10.1007/s00170-005-0384-3](https://doi.org/10.1007/s00170-005-0384-3).
- Pawlowski W. Dynamic Model of Oscillation-Assisted Cylindrical Plunge Grinding With Chatter, *Journal of Manufacturing Science and Engineering* 2013; 135 (5): 051010-051010-6; <https://doi.org/10.1115/1.4024819>.
- Pawlowski W, Kaczmarek L, Louda P. Theoretical and experimental modal analysis of the cylinder unit filled with PUR foam. *Eksplotacja i Niezawodność – Maintenance and Reliability* 2016; 18 (3): 428–435; <http://dx.doi.org/10.17531/ein.2016.3.15>.
- Product data, 2006. DeltaTron® Accelerometers, Types 4514, 4514-001, 4514-002, 4514-004, 4514-B, 4514-B-001, 4514-B-002 and 4514-B-004, Brüel&Kjær.
- Product data, 2005. Impact Hammers - Types 8206, 8206-001, 8206-002 and 8206-003, Brüel&Kjær.

24. Rahmatalla S, Hudson K, Liu Y, Eun H Ch. Finite Element Modal Analysis and Vibration-waveforms in Health Inspection of Old Bridges. *Finite Elements in Analysis and Design* 2014; 78: 40-46; <https://doi.org/10.1016/j.finel.2013.09.006>.
25. Schmitz T L, Smith K S. *Machining Dynamics. Frequency Response to Improved Productivity*. Boston, MA: Springer US; 2009; 8-13.
26. Sorby K, Ostling D. Precision turning with instrumented vibration-damped boring bars. *Procedia CIRP* 2018; 77: 666-669; <https://doi.org/10.1016/j.procir.2018.08.181>.
27. <http://www.stressebook.com/rigid-body-mode/> Access: 10.04.2020.
28. Thomas M D, Knight W A, Sadek M M. The Impact Damper Boring Bar and Its Performance When Cutting. In: S. A. Tobias et al., editors. *Proceedings of the Thirteenth International Machine Tool Design and Research Conference*; 1973; 47-48; [https://doi.org/10.1007/978-1-349-01857-4\\_7](https://doi.org/10.1007/978-1-349-01857-4_7).
29. Thorenz B, Friedrich M, Westermann H-H, Dopfer F. Evaluation of the influence of different inner cores on the dynamic behavior of boring bars. *Procedia CIRP* 2019; 81: 1171-1176; <https://doi.org/10.1016/j.procir.2019.03.287>.
30. Vivo A, Brutti C, Leofanti J. Modal Shape Identification of Large Structure Exposed to Wind Excitation by Operational Modal Analysis Technique. *Mechanical Systems and Signal Processing* 2013; 39 (1-2): 195-206; <https://doi.org/10.1016/j.ymssp.2013.03.025>.
31. Wang M, Zan T, Yang Y, Fei R. Design and implementation of nonlinear TMD for chatter suppression: An application in turning processes. *International Journal of Machine Tools and Manufacture* 2010; 50 (5): 474-479; <https://doi.org/10.1016/j.ijmactools.2010.01.004>.
32. Weck M, Brecher C. *Werkzeugmaschinen 5. Messtechnische Untersuchung und Beurteilung, dynamische Stabilität*. 7th ed. Berlin, Heidelberg, New York: Springer; 2006; 201: 356-366.
33. Zaghbani I, Songmene V. Estimation of Machine-tool Dynamic Parameters During Machining Operation Through Operational Modal Analysis. *International Journal of Machine Tools & Manufacture* 2009; 49 (12-13): 947-957; <https://doi.org/10.1016/j.ijmactools.2009.06.010>.
34. Zhang G P, Huang Y M, Shi W H, Fu W P. Predicting Dynamic Behaviours of a Whole Machine Tool Structure Based on Computer-aided Engineering. *International Journal of Machine Tools & Manufacture* 2003; 43 (7): 699-706; [https://doi.org/10.1016/S0890-6955-\(03\)00026-9](https://doi.org/10.1016/S0890-6955-(03)00026-9).
35. Żywica G, Kaczmarczyk T Z. Experimental evaluation of the dynamic properties of an energy microturbine with defects in the rotating system. *Eksplotacja i Niezawodność – Maintenance and Reliability* 2019; 21 (4): 670–678, <http://dx.doi.org/10.17531/ein.2019.4.17>.





Article citation info:

Hlinka J, Kostial R, Horpatzka M. Application of enhanced methods for safety assessment of FADEC. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 63–73, <http://dx.doi.org/10.17531/ein.2021.1.7>.

Indexed by:



## Application of enhanced methods for safety assessment of FADEC

Jiri Hlinka<sup>a</sup>, Rostislav Kostial<sup>a</sup>, Michaela Horpatzka<sup>a</sup>

<sup>a</sup>Institute of Aerospace Engineering, Brno University of Technology, Technická 2896/2 616 69 Brno Czech Republic

### Highlights

- Safety assessment for FADECs in early development phases.
- A unified concept of criticality levels respecting FADEC use in different types of aircraft.
- Combination of Item and Functional FMEA used to reduce time-effort for safety assessment.
- New failure effect classification used to link effects to different aircraft categories.

### Abstract

The paper deals with safety and reliability assessment as an integral part of the development process for modern aviation products with potentially critical functions. Focus is on digital engine control units, their development process and tools offering potential savings in otherwise time demanding and expensive safety assessment processes. The paper shows application of several approaches, which together form an innovative way for safety assessment of aerospace products (otherwise strictly limited by regulation procedures). It is focused on practical ways towards reduction of development costs during safety assessment, which do not compromise its comprehensiveness. Described approaches are based on experience from development of numerous aerospace products in last nearly 20 years. As an addition, possibility to further enhance the proposed innovative effect classification by application of FMECA was shown. Possible methods for quantitative assessment using Fuzzy logic and/or multiple-criteria decision analysis were discussed.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>) Full Authority Digital Engine Control (FADEC), aircraft, safety assessment, reliability.

## 1. Introduction

Recent fast development in digital technologies in many fields enables replacement of “old-fashioned” analogue or even mechanical control systems by state of the art digital control solutions with many advantages. In aerospace, this process was slowed-down by the fact that replacement of control functions which are safety critical is a demanding and expensive process. Therefore, some aerospace applications still use abovementioned obsolete technologies. Especially in the area of small aircraft with limited resources for development and certification, this issue prevents faster development.

To tackle this issue, new effective approaches need to be developed. The paper deals with research of new approaches to effectively develop complex electronic systems for general aviation aircraft, in particular to ensure effective safety assessment for FADEC (Full Authority Digital Engine Control) under development.

## 2. Full Authority Digital Engine Control in General Aviation

Full authority digital engine control has been in wider use for large transport aircraft since the 1980's. Existing FADECs for large transport or military aircraft have a relatively “narrow” definition of application. In contrary, FADEC considered in this paper is for the general aviation field, where the range of applications can be wide. At

the same time, limited resources for development of new solutions in general aviation caused that FADEC technology is still not present in many engine types in current service.

General aviation category includes a variety of applications from small aircraft used for fun flying, through agricultural aircraft, up to aircraft used for transport of passengers. These applications may significantly differ with size/design of the aircraft (and requirements on safety), but also with type of flights performed by particular aircraft.

Although being initially designed for turbine engines, recently FADEC becomes increasingly popular for use among smaller aircraft with piston engines as well.

In this area, manufacturers of piston engines like Continental and Lycoming are increasingly using this technology for their engines. Lycoming uses its iE2 FADEC technology (TO-450, TIO-540-NXT, TSIO-550, TEO-540-A1A engines). Continental uses its PowerLink FADEC (IO-240, IO-360, IO-550, IOF-240, IOF-550, TSIOF-550 engines). Main advantages of FADEC in this area include simplicity of the engine control (pilot can focus more on situational awareness and less on the aircraft control), better diagnostics of problems, and improved performance and efficiency. Similar attention is given to FADECs for diesel reciprocating engines for aviation use. As claimed by Cox [12], prices of FADECs for this application were between \$2500 and \$7500.

E-mail addresses: J. Hlinka - [hlinka@fme.vutbr.cz](mailto:hlinka@fme.vutbr.cz), R. Kostial - [kostial@fme.vutbr.cz](mailto:kostial@fme.vutbr.cz), M. Horpatzka - [michaela.horpatzka@vutbr.cz](mailto:michaela.horpatzka@vutbr.cz)

Also today's small turbine engines in general aviation are often equipped with FADEC in all or some configuration options. As an example, Rolls-Royce M250 turboshaft engine developed from Allison Model 250 (produced and continuously improved from the 1960s) features dual-channel FADEC in the latest series. The engine is used for example in MD 530F from 2016 or in Bell 407GX from 2018 [26].

Pratt and Whitney has two variants of PT6 engine with electronic control. PT6C is a medium-class helicopter turboshaft engine with dual-channel FADEC with hydro-mechanical backup, used in AW139 [30]. The latest PT6E offers a dual-channel integrated electronic propeller and the engine control system also with single-lever power control. Single-lever power control (for both engine and propeller) reduces pilot workload. In 2019 PT6E was announced as the power plant for Pilatus PC-12 NGX [29].

Czech company PBS Velka Bites developed small turbine engines. FADEC is a standard equipment of models TJ100 and TJ150. Model TJ100 is aimed at light sport aircraft, gliders, UAVs and micro jets, while bigger TJ150 was designed directly for unmanned applications [1].

The paper deals with FADEC for small turbine engines for general aviation. All works are part of a wider research project done in cooperation with UNIS company. Research is dedicated to advanced technology of modular control and diagnostics systems for small turbine engines with thrust around 1500 N and small turboprop engines with power around 180kW. Such engine size has a wide range of aircraft applications and flight profiles. This may include for example UAS (unmanned aerial systems), small sport aircraft (2 or 4 seaters), or auxiliary power for gliders, see Tab. 1. Each application has specific performance demands, engine operation duration, or frequency of power level changes.

### 3. System Safety Assessment Process in Aerospace

Although safety assessment is an integral part of design and certification of new equipment for aerospace use, current common methods have significant space for improvement. Regulation requirements (both FAA 14 CFR, and EASA CS) prescribe during development/design safety assessment combination of standard methods like FHA (Functional Hazard Assessment), FMEA/FMECA (Failure Modes, Effects/and Criticality Analysis), FTA (Fault Tree Analysis), RBD (Reliability Block Diagrams). Recommended practices for aerospace are summarized in SAE APR 4761 [14]. Critical review of prediction techniques is the subject of several papers, for example [27], [35]. However, for the development of complex product, like FADEC, also new sophisticated methods can be applied.

FHA is a systematic, comprehensive examination of functions to identify and classify conditions of those functions according to their severity [14]. FHA is usually used as a preliminary analysis during the early design phase, when exact components of the system are not yet known. Since it evaluates functions of the system and effects of their loss, it helps to find critical systems/parts already in early design phases.

FMEA is a qualitative method of analysis that involves the study of possible failure modes and faults in sub items, and their effects at various indenture levels [17]. For aerospace use it provides a systematic, bottom up method of identifying the failure modes of systems, components or functions and determining the effect on the aircraft. It is a key method to prove the requirement "no catastrophic event should result from the failure of a single component". Typically is used Functional or Item (Piece parts) FMEA [14].

Both FHA and FMEA are qualitative assessment methods. For quantitative assessment are used FTA, RBD or Markov analysis. These are top-down methods which proceed down through more detailed levels of design. After qualitative analysis, when failure condi-

Table 1. Possible range of use for general aviation FADEC (intended use and limitations)

<b>Auxiliary Propulsion</b>	
<b>Single Engine Aircraft Glider</b>	<ul style="list-style-type: none"> <li>a) Engine is not used for <b>critical phases</b> of flight (take off, landing etc.)</li> <li>b) Electrical power is not generated by engine. Independent source (battery) is used.</li> <li>c) Independent fuel cut offs.</li> <li>d) Regulation <b>CS-22 +CRI + CS-23, CS-E</b></li> </ul>
<b>Primary Propulsion</b>	
<b>Single Engine Aircraft</b>	<ul style="list-style-type: none"> <li>a) Electrical power backup (batteries) for at least 30 minutes flight in case of generator failure.</li> <li>b) Independent fuel cut offs.</li> <li>c) Aircraft without anti-icing system are limited to IMC without icing.</li> <li>d) Regulation <b>CS-23, CS-E (CS-VLA, CS-LSA, L-2)</b></li> </ul>
<b>Multi Engine Aircraft</b>	<ul style="list-style-type: none"> <li>a) Electrical power backup (batteries) for at least 30 minutes flight in case of generators failure.</li> <li>b) Independent fuel cut offs.</li> <li>c) Each engine is controlled by independent own FADEC unit.</li> <li>d) Aircraft without anti-icing system are limited to IMC without icing.</li> <li>e) Regulation <b>CS-23, CS-E</b></li> </ul>
<b>UAV/UAS Primary Propulsion</b>	
<b>Single Engine Aircraft</b>	<ul style="list-style-type: none"> <li>a) Electrical power backup (batteries) for at least 30 minutes flight in case of generator failure (including remote control, and communication with operator.)</li> <li>b) Independent remotely controlled fuel cut offs.</li> <li>c) Aircraft without anti-icing system are limited to IMC without icing.</li> <li>d) Regulation <b>STANAG 4671, CS-LUAS</b></li> </ul>
<b>Multi Engine Aircraft</b>	<ul style="list-style-type: none"> <li>a) Electrical power backup (batteries) for at least 30 minutes flight in case of generators failure (including remote control, and communication with operator.)</li> <li>b) Independent remotely controlled fuel cut offs.</li> <li>c) Each engine is controlled by independent own FADEC unit.</li> <li>d) Aircraft without anti-icing system are limited to IMC without icing.</li> <li>e) Regulation <b>STANAG 4671, CS-LUAS</b></li> </ul>

tions are identified, quantitative analysis can be applied to find what single failure or combinations of failures exist at lower levels that might cause each failure condition [14].

In addition, for software development in the aerospace industry, recommendations of RTCA DO-178 [11] are applied. More information on safety assessment of software for aerospace use can be found in a number of papers. For example, in [33] is an overview of the RTCA DO-178C and its impacts on Certification of Safety-Critical avionic systems. Another overview and certification of the safety critical computer systems using RTCA DO-178 is presented in [19]. More practical use of RTCA DO-178 for condition monitoring system is presented in [13].

Although new progressive methods offering some advantages can be found in several research works (for example [20]), aerospace industry relies on above-described well proven methods which are also established in aerospace regulation requirements. Therefore, the work presented in the paper is based on FHA and FMEA. In addition, the paper focuses on new ways to reduce time effort and costs for safety assessment using these methods, which are acceptable for the aerospace certification process. In fact, modifications proposed in the paper are so extensive, that they form an innovative approach to both FHA and FMEA which was not to such extent applied in aviation before. For example, presented enhanced FHA uses a totally new definition of criticality levels allowing rapid application of results on different aircraft classes (with different applications and failure effects). Proposed hybrid FMEA approach (although may be seen in similar applications for other industrial sectors), in this paper is interlinked to classification from enhanced FHA (for greater flexibility for different aircraft types), and optimized for aerospace application (respecting its typical segmentation/functional zoning). Since the aerospace industry is facing escalation of development costs with every new aircraft generation (additional development costs related to more strict requirements and more complex systems), reduction of effort and costs in every aspect of the development process is extremely important.

### 3.1. System Safety Assessment of FADEC

For safety assessments of FADEC are usually used methods, which allow simulation - for example Markov analysis (Markov chains) which was used for prototypes of FADEC for JAS 39 Gripen [15], or where the Markov process [24] and Monte Carlo simulation [25] based time limited dispatch analysis for FADEC was used. Another option is an analysis based on Bayesian networks. Research [21] used improved BN analysis for commercial aircrafts FADEC. Simulations

require detailed model of the FADEC system and thus are suitable for later design phases, where the system structure is mostly established and it is assumed there will be no more major changes. Since the detailed modelling of complex systems is time consuming, there are efforts to simplify these models or methods [9].

However in the described case, the assessment was done for the FADEC in pre-prototype and prototype phase of development, where many parts were subject to change. Therefore putting an effort into creation of a detailed system model was impractical.

To adapt to early design conditions, adjustment was done to the traditional safety assessment methods especially FHA and FMEA to minimize the need to rework analysis every time the change occurs. For quantitative analysis FTA (Fault Tree Analysis) was used, which is not described in detail in the paper since its standard form described in SAE ARP 4761 [14] was used. The use of FTA in combination with Markov analysis for FADEC reliability assessment is described in detail in [22] for example. System safety for FADEC from a software perspective was addressed in [28].

## 4. Enhanced safety assessment concept for complex electronic systems

Current general safety assessment process for aviation in a simple form is shown on Fig. 1. This scheme was derived from recommendations of EASA CS AMC 25.1309 (Acceptable Means of Compliance) System design and analysis [3].

The team of authors was from the beginning facing strict submission of FADEC with a wide range of use, see Tab. 1. Similar needs can be expected for different engine control units producers in a given power range. To enable efficient and precise safety assessment of such complex electronic system, some new techniques were adopted.

These include:

- Functions criticality level analysis – a unified concept of criticality levels respecting FADEC use in different types of aircraft, see chapter 4.1;
- Enhanced FHA – used to identify critical functions of FADEC, see chapter 4.2;
- Hybrid Block FMEA – used to reduce time-effort for safety assessment, see chapter 4.3;
- Two-phase failure effect classification – used to link FADEC failure effects to different aircraft categories (applications), see chapter 4.3.1.

These techniques can be applied on any complex electronic system (in general). For aviation, in addition, all safety assessment techniques must comply with main airworthiness requirements for aircraft design and certification, typically EASA CS-23 [7] or FAA 14 CFR Part 23 [2] (for aircraft with fixed wing and propulsion unit), EASA CS-22 [6] (for gliders with auxiliary power unit), or other similar requirements. Depending on the country of origin, also Chinese or Russian equivalent airworthiness requirements can be applied. However, most of the regulation requirements link to the same industrial standards and practices. For example, CS-23 and 14 CFR Part 23 requirements link to safety assessment procedures described in ASTM F3230-17 [31]. Detailed guidelines for safety assessment including a list of assessment methods are also available in SAE ARP 4761 [14]. A list of basic assessment methods is shortly mentioned in chapter 3, further information on safety assessment procedures is for example in [26]. The paper is focused on practical ways towards reduction of develop-

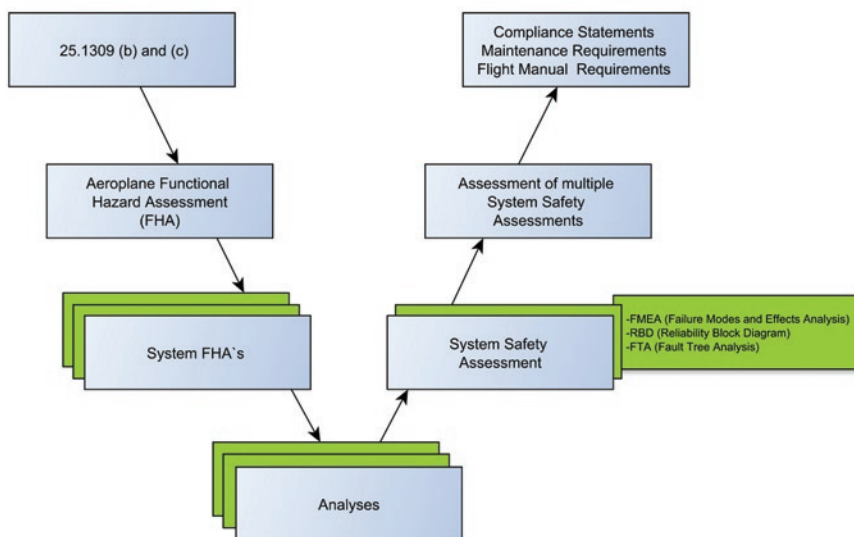


Fig. 1. Current general safety assessment process for aviation

Table 2. Functions criticality level

Functions Criticality level	Description
Essential	Functionality of particular part of the system or the system itself is directly influenced by a function behaviour. It essentially provides intended function of the system or part of the system (Thrust control, fuel flow control etc.)
Moderate	The function indirectly influences essential functions. However, moderate function does not provide intended function itself (Overspeed protection, Turbine temperatures, Oil pressure etc.).
Marginal	Functions that supports essential or moderate functions in the proposed envelope.
Insignificant	Functions without influence on essential or moderate functions.

ment costs during safety assessment, which do not compromise its comprehensiveness.

#### 4.1. Functions Criticality Level Analysis

As can be seen from Tab. 1, FADEC for general aviation engine can perform a wide variety of missions. In addition, it can also have a wide range of critical functions, i.e. engine control, electric power generation control, etc. Therefore, it is necessary to decompose all FADEC functions and link them to categories. New applied method divides all FADEC functions into four categories, see Tab. 2. These functions are linked to Enhanced FHA analysis and complement the Failure Classification.

#### 4.2. Enhanced FHA

Every FADEC function is analysed using FHA (Functional Hazard Assessment). The main goal of FHA results is to provide a list of potentially risky functions (and shortcomings), which should be further analysed, and/or corrective action should be taken (design change, added redundancy, etc.). Standard FHA is generally used for one specific application. This may be a problem if intended use of the product is for different specifications (CS-22, CS-23) or applications (manned/ unmanned). Enhanced FHA uniquely solved this problem, since it was developed to effectively cover all intended applications of analysed product. Enhanced FHA shown in Fig. 2 for the first time ever identifies effects for all FADEC applications in the single table.

According to aviation regulation requirements, all functions with HAZARDOUS or CATASTROPHIC consequences have to be further analysed using a prescribed set of analysis methods. Also functions from categories ESSENTIAL or MODERATE (according to Tab. 2) should have detailed safety assessment. These steps were done as normal engineering procedure out of the scope of the paper.

See chapter 5.2 for more details on example FADEC results.

#### 4.3. Hybrid Block FMEA

Classic safety assessment process defined in SAE ARP4761 recognizes Part FMEA or Functional FMEA. Since developed FADEC was a complex electronic device, with a significant number of electronic parts, standard part FMEA would be time consuming. On the other hand, functional FMEA would not respect fully hardware “block composition” of FADEC. Therefore, “hybrid FMEA” was proposed and applied, combining advantages of both, part and functional FMEA. Goal of the application of hybrid method was to reduce the number of analysed components, and ability to quickly integrate design changes into safety assessment.

Decomposition of analysed FADEC led to functional blocks. Each block is a set of components performing defined functions. Two block types can be recognized:

- Simple block – performs single function (i.e. temperature measurement, el. current filtration, ...)

(PRELIMINARY) FUNCTIONAL HAZARD ANALYSIS					
PROJECT:		SYSTEM:		PAGE	
Function ID	Function	Failure condition/ Hazard Description	Flight Phase	Failure classification	Function Criticality classification
EC.01	Engine power control	Loss of thrust/ engine stop. Failure mode results in immediate single engine stop. Flight crew follows flight manual procedures in the case of single engine loss.	ENR		ESSENTIAL(4)

(PRELIMINARY) FUNCTIONAL HAZARD ANALYSIS FAILURE CLASSIFICATION					
AUXILIARY PROPULSION SYSTEM CS-22 (MANNED)		PRIMARY PROPULSION SYSTEM (MANNED)		UAV/UAS (UNMANNED)	
Single engine		Single engine	Multi engine	Single engine	Multi engine
MINOR		MAJOR (Up to CATASTROPHIC in IFR and IMC condition)	MINOR	MAJOR (Up to CATASTROPHIC in IFR and IMC condition)	MINOR

Proposed unique Failure classification interconnects “function”, “function criticality classification” and “failure effects for different aircraft types” (respecting aviation regulation requirements).

Fig. 2. Example FHA applied on the FADEC

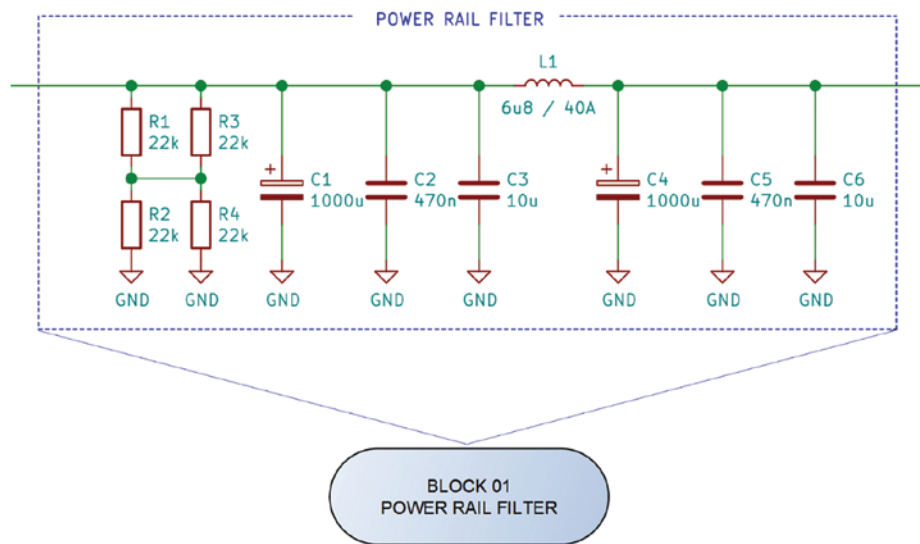


Fig. 3. Example of simple filter block for el. energy filtration

- Node block – performs more than one function (i.e. MCU-Micro Controller Unit with sensors data processing and fuel flow control).

Each function must be analysed in FMEA for Node blocks. Fig. 3 shows an example of a functional block (composed of filters in the circuit of integrated electric power generator) which consist of capacitors (C1, C2 etc.) resistors (R1, R2 etc.) and coils (L1) with connections to ground (GND).

#### 4.3.1. Two-phase failure effect classification for FMEA

Same failures can have different effects on different aircraft types/categories. Shown methodology takes this fact into account, it has two-phase process.

Table 3. FADEC system criticality level

FADEC Criticality Level	Description
<b>Safety - Critical</b>	Failures which directly affect the system ability to perform primary and/ or essential functions. System can no longer perform its primary functions. Emergency shutdown or switch to backup system required.
<b>Serious</b>	Failures which directly affect the system ability to perform primary and/ or essential functions. System is able perform its primary functions for limited amount of time, before shutdown or switching to backup system.
<b>Degraded</b>	Failures which can affect performance of the system. Primary functions are preserved with limited performance for an unlimited time.
<b>Not Critical</b>	Failures without an effect on primary or essential functions.

Table 4. Example of different failures with same effects on aircraft

BLOCK/ Failure Effect	Effect on System	Effect on Aircraft / Failure Class
<b>Power Rail Filter/ Short</b>	In case of a failure, ECU (Electronic Control Unit) overvoltage protection is activated. In worst case scenario, ECU is switched-off and back-up system is used for engine control. Crew has visual indication on activation of back-up system.	Airplane continues in flight using back-up system for engine control. Some features are deactivated, i.e. overspeed protection, automatic engine start, etc. ECU cannot be re-activated during flight.  Failure Class: <b>FAILURE.2</b>
<b>Control Microprocessor/ Analog to Digital Converter failure</b>	In case of a failure, analogue segment of MCU (Microprocessor) and analogue inputs fail, resulting in ECU emergency shut down and use of back-up system for engine control. Crew has visual indication on activation of back-up system.	
<b>Overspeed Protection/ Loss of Sensing</b>	In case of a failure, MCU has no information on turbine rpm. This results in ECU emergency shut down and use of back-up system for engine control. Crew has visual indication on activation of back-up system. Crew is able to monitor turbine rpm using on-board instruments (indication is independent on overspeed protection).	

### Phase One: Failure effect classification on a system level

Phase one classifies failure effects on FADEC system itself. Every safety assessment classification is done just for FADEC system, no effect on aircraft or engine is considered. Goal of the Phase One is to predict critical elements of FADEC (without link to particular application). Therefore, different assessment classification is used compared to aviation standards [3], see Tab. 3.

### Phase Two: Failure effect classification on an airplane level

Based on previous long-time experience with safety assessment of different types of engine control systems, we can state that failures of engine control system components can lead only to a limited number of failure effects on the airplane level. Tab. 4 demonstrates several different MCU (Main Control Unit) failures with the same effect on the aircraft.

Above mentioned failure effects are summarized and classified for different aircraft. An example of “Hybrid Block FMEA” with two-phase failure effect classification is shown on Tab. 5 and Tab. 6.

#### 4.4. Extension to Hybrid Block FMECA

To further enhance proposed innovative effect classification, it is possible to use FMECA (Failure Modes, Effects and Criticality Analysis) instead of FMEA. This will move proposed solution towards quantitative assessment, adding also criticality level (for example in the form of CN – Criticality number). Authors did consider this extension for future enhancement of the presented method. Many models for criticality definition exist, including models based on risk matrix,

Table 5. Example of FMEA format

Failure mode and effect analysis (FMEA)							
Project:	System:	NO. DW:		Page:			
Block ID	Function	Failure Mode	Failure description	Failure effects			Note
				Flight phase See. Table 7	Effects on system	Effects on aircraft	
<b>Block 1</b>	Power-Rail filter provides el. filtration and connectivity to aircraft on-board electrical grid.	Short circuit	Short circuit on one or more of the block items.	APR	In case of a failure, ECU over-voltage protection is activated. In worst case scenario, ECU is switched-off and back-up system is used for engine control. <b>CLASS: Critical</b>	In case of failure, the aircraft continue in flight on back-up system. Some functions are deactivated (overspeed protection, automated start etc.). ECU cannot be reactivated during the flight. <b>CLASS: Failure.02</b>	-

Table 6. List of failures and classification on aircraft

List of failures - Failure mode and effect analysis (FMEA)			
Project:	System:	Failure No.:	FAILURE DESCRIPTION:
		FAILURE.01	In case of failure, the engine losses thrust due to the emergency shutdown. <b>Change in aircraft behaviour</b> <b>Master warning – red light and sound</b> <b>(except: jet engine glider)</b>
Plane Category	Flight phase See. Table 7	Failure Classification	Note
Jet engine Glider	STD	-	-
	TOF, APR, ICL, LND	-	Engine is not used for critical phases of flight
	ENR	Max. MINOR	Failure can limit the flight range or lead to controlled landing to the terrain.
Ultralight plane (Max. 400 kg)	STD (START)	-	-
	TOF, APR, ICL, LND	MAJOR	Flight crew follows flight emergency procedures for loss of thrust. Forced emergency landing.
	ENR	MINOR	Flight crew follows flight emergency procedures for loss of thrust. Can lead to emergency landing to the terrain.
Single engine CS-23 plane (Max. 8618 kg)	STD	-	-
	TOF, APR, ICL, LND	HAZARDOUS to CATASTROPHIC	Potentially leads to the catastrophic result in the case of improper (or limit) take off, initial climb, approach and landing. Extensive crew workload (in limited time) could impair ability to perform task.
	ENR	MAJOR to CATASTROPHIC	Failure mode results in forced emergency landing according to the flight manual emergency procedures. It potentially leads to the catastrophic results in IMC conditions.
Multi engine CS-23 plane (Max. 8618 kg)	STD (START)	-	-
	TOF, APR, ICL, LND	MAJOR to HAZARDOUS	Failure mode results in immediate single engine stop and significant crew workload. Flight crew follows flight manual procedures in the case of single engine loss. Flight crew uses fully functional engine and flight controls to eliminate negative effects of loss of engine control. Potentially hazardous condition in case of inappropriate reaction to asymmetrical thrust.
	ENR	MINOR	Failure mode results in immediate single engine stop. Flight crew follows flight manual procedures in the case of single engine loss.

Table 7. ICAO flight phases [10]

ICAO FLIGHT PHASES		
Phase	Abbreviation	Description
STANDING	STD	Prior to pushback or taxi, or after arrival, at the gate, ramp, or parking area, while the aircraft is stationary.
TAXI	TXI	The aircraft is moving on the aerodrome surface under its own power prior to take off or after landing.
TAKEOFF	TOF	From the application of take off power, through rotation and to an altitude of 35 feet above runway elevation.
INITIAL CLIMB	ICL	From the end of the Take-off sub-phase to the first prescribed power reduction, or until reaching 1,000 feet above runway elevation or the VFR pattern, whichever comes first.
EN ROUTE	ENR	Instrument Flight Rules (IFR): From completion of Initial Climb through cruise altitude and completion of controlled descent to the Initial Approach Fix (IAF). Visual Flight Rules (VFR): From completion of Initial Climb through cruise and controlled descent to the VFR pattern altitude or 1,000 feet above runway elevation, whichever comes first
MANEUVERING	MNV	Low altitude/aerobatic flight operations.
APPROACH	APR	Instrument Flight Rules (IFR): From the Initial Approach Fix (IAF) to the beginning of the landing flare. Visual Flight Rules (VFR): From the point of VFR pattern entry, or 1,000 feet above the runway elevation, to the beginning of the landing flare.
LANDING	LDG	From the beginning of the landing flare until aircraft exits the landing runway, comes to a stop on the runway, or when power is applied for take off in the case of a touch-and-go landing.
EMERGENCY DESCENT	EMG	A controlled descent during any airborne phase in response to a perceived emergency situation.
UNCONTROLLED DESCENT	UND	A descent during any airborne phase in which the aircraft does not sustain controlled flight.

which are most widely used in aerospace. Several selected models for CN definition are listed below.

4.4.1. CN based on criticality factors ([16])

$$C_{KR} = \{\pi_1, \pi_2, \pi_3, \dots, \pi_N\} \frac{1}{N} \quad (1)$$

where factors  $\pi_{(1..N)}$  are weighting factors that express influences on failure effects. These factors can for example represent influence of:

- failure classes,
- effect of the part failure on the system,
- failure probability of one part in a set of all analysed parts,
- ease of failure detection,
- speed of response on failure.

All these factors are based on expert judgement which leads to certain subjectivity of assessment. Therefore this method is suitable primarily for assessments, where there is no reliable source of information on failure probability.

4.4.2. CN defined using generic base failure rate with influencing factors ([16])

$$C_{KRi} = \sum_1^N (\beta \cdot \alpha \cdot K_E \cdot K_A \cdot \lambda_G \cdot t \cdot 10^6)_i \quad (2)$$

where:

- $C_{KRi}$  criticality factor of the part,
- $i$  ID number of the part,
- $N$  total number of parts,
- $\beta$  conditional probability that the failure will lead to a critical failure of the system,

- $\alpha$  relative ratio between failure rate of the given type to total failure rate for given part,
- $\lambda_G$  failure rate of a part with influence of all possible failure modes. The usual form is: failure rate/ $10^6$ ,
- $t$  operating time that each part accumulates during whole operating time of the system,
- $K_E$  corrective factor, incorporates effects of different operating conditions against conditions, for which was  $\lambda_G$  determined,
- $K_A$  corrective factor, incorporates effects of different operating loads against loads, for which was  $\lambda_G$  determined.

Failure mode (modal) criticality number ([32])

$$C_m = (\beta \cdot \alpha \cdot \lambda_p \cdot t) \quad (3)$$

where:

- $C_m$  Failure mode criticality number
- $\beta$  Conditional probability of the current failure mode's failure effect
- $\alpha$  Failure mode ratio
- $\lambda_p$  Item failure rate
- $t$  duration of applicable mission phase (expressed in hours or operating cycles)

4.4.4. Risk priority number (RPN) method in FMECA

RPN method reviews the risk level of failure modes using assessment of *probability of failure mode occurrence* (O), *effects severity* (S) and the *probability of detecting the failure* (D). It ranks O,S and D on 1 – 10 [34]. Risk assessment is calculated by multiplying the ranking values of O, S and D [5].

Although in aerospace mode occurrence (O) can usually be defined with high degree of confidence (thanks to the previous experience and operational data), effects severity (S) and probability of detecting the failure (D) may sometimes involve high degree of subjective judge-

# DETECTABILITY

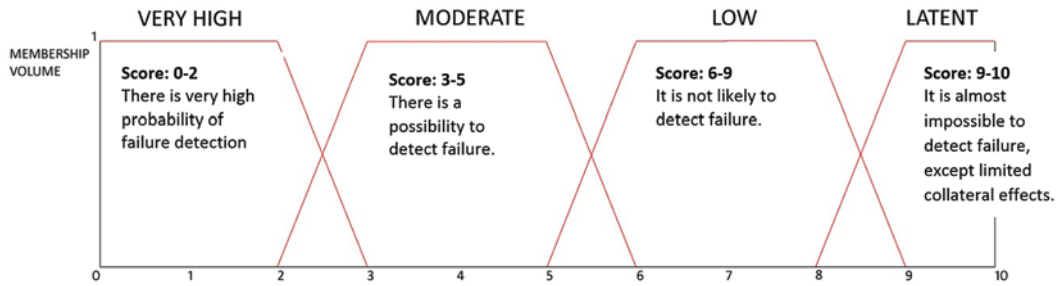


Fig. 4. Fuzzy membership function for linguistic variable- detectability [18]

Table 8. Detectability scoring table, ref. [18] with highlighted capacitor C1 failure modes (green – short, blue – value change, red – open)

	Direct detectability (indication)			Indirect detectability (results of non-function)			Pre-flight inspection/ procedures		Score	
	YES (multiple levels)	YES (single level)	NO	YES	Partially	NO	YES	NO		
<b>Direct detectability:</b> Failure is indicated. Flight crew is able to respond to the failure effects and proceed according to the flight manual	x			x			x		1	HIGH ↑ ↓ LATENT
		x		x			x		2	
		x			x		x		3	
		x			x			x	4	
		x				x	x		5	
<b>Indirect detectability:</b> Failure is indirectly indicated to the crew by its collateral effects. Flight crew is able to identify occurring failure without significant workload.			x	x			x		6	
			x		x		x		7	
			x			x	x		8	
<b>Pre-flight inspection test:</b> Item malfunction is possible to detect during pre-flight test, according to the flight manual.			x		x			x	9	
			x			x		x	10	

ment. To reduce this significant shortage, which is not only linked to RPN method (but also other mentioned methods), Fuzzy logic may further be applied as a supporting tool. This can be especially meaningful in case of probability of detecting the failure (D). An alternative approach on optimization of parameters is for example in [34].

Although a number of papers can be found, where RPM is criticized for some shortcomings, which were summarized by Liu in [23], it is most suitable method for Fuzzy logic application (and reduction of degree of subjective judgement). Most significant shortcomings usually mentioned include RPN values and their varying sensitivity to small changes, or that parameters O, S and D are equally weighted.

Choice of specific criticality analysis method primarily depends on available input information and secondarily on specific conditions and goals of the analysis. For example case described in the paper, most suitable seems to be RPN method, which will be considered in further text. Authors do consider extension of methods presented in Chapter 4 towards partially quantitative assessment using FMECA with application of Fuzzy logic or Multiple-criteria decision analysis. Chapters 4.5 and 4.6 are showing potential of this extension, and should be considered as an introduction to future work.

#### 4.4.5. Fuzzy extended criticality inputs

To evaluate a given item probability of detecting failure (D), for example scoring Tab. 8 can be used (based on [18]). Detectability

scoring interval is  $\langle 0,10 \rangle$ . Lower score corresponds with higher probability of failure detection. High score corresponds with lower probability of failure detection resulting in a latent failure. Detectability fuzzy membership is established in Fig. 4. The trapezoidal membership function is used.

Tab. 9 shows an example component from simple filter block shown on Fig. 3. In this case the component is ceramic capacitor C1, and its failure modes detectability. As can be seen, there is no direct possibility to detect any of the three failures modes shown. Short circuit can be detected based on collateral effects (activation of overload protection and automatic switch to HBM mode), and during pre-flight inspection. Change in operational parameters is practically undetectable and latent until more capacitors degrade, or until another failure mode occurs. Open circuit of the capacitor can cause filtration degradation which can influence some of very sensitive parts of the system. The detectability of the failure is very complicated.

#### Fuzzy interface process

As can be seen in Fig. 5, input values O, S, and D are starting point for Fuzzy inference process. Fuzzy procedures described many times in the literature can be applied. The most used inference technique is Mamdani, developed by Professor Ebrahim Mamdani of London University in 1975. Detailed description of fuzzy inference process is out of the scope of the paper. It uses several process steps, including



Fuzzification, Rule evaluation (using Fuzzy inference rules) and De-fuzzification.

The last step, De-fuzzification, is done in the order to gain the fuzzy process single scalar quantity output. Ranking represents the extended criticality level of the failure mode.

Table 9. Detecability list of C1 capacitor failure modes

DETECABILITY LIST OF C1 CAPACITOR FAILURE MODES		
Short circuit	6	LATENT
Open circuit	9	LATENT
Value Change	10	LATENT

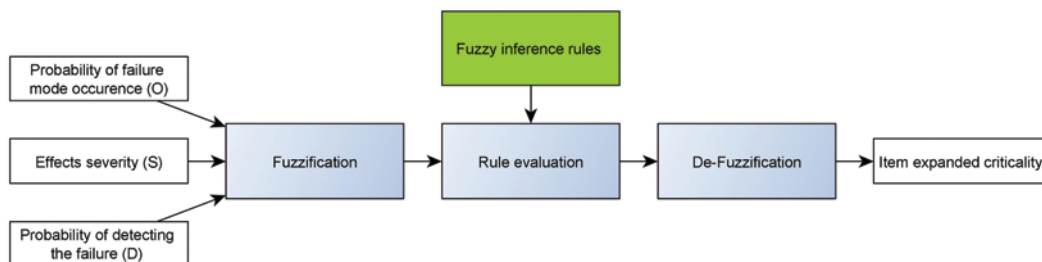


Fig. 5. Fuzzy inference process

The last step, De-fuzzification, is done in the order to gain the fuzzy process single scalar quantity output. Ranking represents the extended criticality level of the failure mode. For De-fuzzification, probably the most used defuzzification technique is centroid technique. It finds where vertical line would slice the aggregate set on final Fuzzy scoring into two equal masses. Mathematically this center of gravity (COG) can be expressed as follow:

$$COG = \frac{\int_a^b \mu_x(x) \cdot x \cdot dx}{\int_a^b \mu_x(x) \cdot dx} \quad (4)$$

where  $\mu_x$  is membership function on final scoring.

Risk assessment methodology using fuzzification for RPN methodology was discussed in ref. [8]

#### 4.4.6. Multiple-criteria decision analysis

Second method considered for future evaluation of criticality is multiple-criteria decision analysis. If applied during FADEC proto-

type design, it has an advantage of different weighting for O, S, and D criteria. It also has small sensitivity for changes of non-critical criteria. On the other hand, it is sensitive for changes of critical criteria used for decision making. There was some work which uses multiple-criteria decision before, for example [4], where was used TOPSIS method maritime risk evaluation.

Authors plan to make comprehensive evaluation of both methods on the case of one functional FADEC block. Method with best results will be than recommended for application on whole FADEC.

## 5. Result and discussion - Enhanced safety assessment concept applied on FADEC

### 5.1. Analysed system description

All methods described in chapter 4 were applied to the engine digital control unit for small turbine engine with 1500 N thrust and integrated electric generator. Control unit was composed of 4 main modules with the total 1168 components. Control unit general composition is shown on Fig. 6.

### 5.2. System analysis results

For the particular FADEC system, enhanced FHA was performed (as described in chapter 4.2) for aircraft categories presented in Tab. 1. The goal was to identify effects resulting from the failure of the analysed function. In total 21 functions were defined and analysed covering complete FADEC functionality with respect to higher aircraft levels. In total 6 critical functions were selected for more detailed analysis. In addition, for less critical functions, corrective actions were proposed (often new procedures for flight manual).

System components were divided into functional blocks, complete FADEC was divided into 78 functional blocks. Blocks were analysed using hybrid block FMEA with Failure Effect Classification. Total of

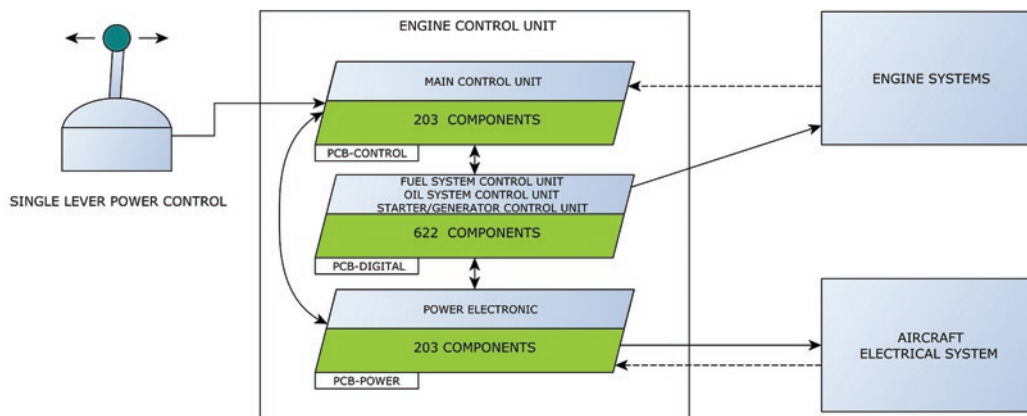


Fig. 6. General composition / scheme of analysed FADEC

## HYBRID BLOCK FMEA TIME SAVING CHART



Fig.7. Time savings received by application of presented methods

Table 10. Comparison between part FMEA and Hybrid block FMEA

PCB	Parts	Functional blocks	Duration - Hours Part FMEA (10 min per part)	Duration - Hours HYBRID BLOCK FMEA (30 min per BLOCK)	% of time (Hybrid FMEA/ Part FMEA)
PCB-1	203	17	2030 h	510 h	25%
PCB-2	622	41	6220 h	1230 h	20%
PCB-3	343	20	3430 h	600 h	17%

7 main effects on aircraft level (effects with significant impact on the function of the whole aircraft and safety of flight) were identified.

During the FADEC development countless minor changes were made, such as replacements of some elements, or changes in the power supply or filtration parts of FADEC. Thanks to the use of Enhanced safety assessment and especially the Hybrid block FMEA, it was not necessary to analyse these minor changes at the level of individual components. Only failure rates for the given block were corrected and so the FMEA evaluation did not change.

There were a few major changes during the development which had influence on FADEC functions or number and layout of PCBs (Printed Circuit Boards). These changes had to be revised in Hybrid block FMEA but for blocks affected by design changes only.

If we consider minimum time necessary for single part analysis in part FMEA to be in average 10 min (taking into account great number of repeating parts, which speed-up the assessment process), and compare it with average 30 min for the analysis of a single block in hybrid block FMEA, we can estimate time savings for safety assessment of FADEC like device, see Tab. 10 and Fig 7.

Calculated times are based on long-time experience from aerospace safety assessment process (item analysis time, influence of connection into the system and effects on other system items/elements are considered). The concept of safety assessment has been proven to be suitable for safety assessment in early phases of FADEC development. It can be assumed that methods mentioned in this paper will be suitable for other complex electronic systems.

### 6. Conclusion

More strict regulation requirements and more complex aircraft systems are the main reason for increasing development costs for recent aerospace projects. Reduction of development effort and costs in every aspect of the development process is therefore extremely important. At the same time, it is not possible to omit any function on the

aircraft and its detailed analysis. Unique solutions presented in the paper were strictly driven by a requirement to ensure the same extent of analysed functions as traditional methods, without possibility to omit any important information (i.e. function or component failure). Structure and outputs were continuously compared with previous works on other aviation products. At the same time, developed solutions enable quick adoption of safety assessment on different aircraft types (different FADEC applications) without the need to repeat a complete set of analyses from the beginning for each aircraft type. Thanks to a combination of specially defined function criticality levels with enhanced FHA, any future application in different aircraft category can be quickly analysed without demanding modifications of complete safety assessment.

Enhanced safety assessment done on an example FADEC confirmed, that small design changes inside blocks (with small/no functional effects) do not require comprehensive and time demanding revision of complete safety assessment (as in case of classic Part FMEA application). Small design change applied in this example design was, among others, integration of filtration capacitors. At the same time, reliability of major hardware blocks is available.

Larger design modifications (like change of number of PCBs) require major revision of safety assessment. However, this revision can be easily applied only to modified blocks. Functional blocks proved useful also for later fault tree analysis.

At every moment of performed works, comparison was done to monitor, if new method analyses all functions / component failures like in the case of traditional methods prescribed by regulation requirements, tracking if comprehensiveness and reliability correctness of the process is ensured. There was no evidence of any shortcoming as a result of developed procedures.

As an addition, possibility to further enhance the proposed innovative effect classification by application of FMECA was shown. Possible methods for quantitative assessment using Fuzzy logic and/or

multiple-criteria decision analysis were discussed. Authors do consider this extension for future enhancement of the presented method.

### Acknowledgement

*Work published in the paper was funded by the Ministry of Industry and Trade of the Czech Republic within the TRIO program, through the research project FV20043 Advanced Technology of Modular Control and Diagnostics Systems for Aircraft Engines.*

### References

1. Aircraft Engines. Aircraft Engines - PBS. [<https://www.pbs.cz/en/our-business/aerospace/aircraftengines>].
2. Airworthiness Standards: Normal, Utility, Acrobatic, and Commuter Category Airplanes. FAA 14 CFR Part 23
3. AMC 25.1309 System design and analysis. EASA CS-25 Amendment 21. 2018.
4. Bařhan V, Demirel H, Gul M. An FMEA-based TOPSIS approach under single valued neutrosophic sets for maritime risk evaluation: the case of ship navigation safety. *Soft Comput* 2020, <https://doi.org/10.1007/s00500-020-05108-y>
5. Bluvband Z, Grabov P. Failure Analysis of FMEA. 2009 Annual Reliability and Maintainability Symposium 2009; 344-347, <https://doi.org/10.1109/RAMS.2009.4914700>.
6. Certification Specification for Sailplanes and Powered Sailplanes. EASA CS-22 Amendment 2, 2009.
7. Certification Specification for Normal-Category Aeroplanes. EASA CS-23 Amendment 5, 2017.
8. Chang K H, Cheng C H. A risk assessment methodology using intuitionistic fuzzy set in FMEA. *International Journal of Systems Science* 2010; 41; 1457-1471, <https://doi.org/10.1080/00207720903353633>.
9. Chen B, Li C, Li Y, Wang A. Reliability analysis method of an aircraft engine FADEC system. 8th International Conference on Reliability, Maintainability and Safety 2009; 8: 289-292, <https://doi.org/10.1109/ICRMS.2009.5270188>.
10. Data Definition Standard - English- Attribute Values. ICAO ECCAIRS Aviation 1.3.0.12. 2013.
11. DO-178C Software Considerations in Airborne Systems and Equipment Certification. RTCA
12. FADEC Comes Of Age. [<https://www.planeandpilotmag.com/article/fadec-comes-of-age/?start=1>].
13. Gerdes M, Galar D, Scholz D. Decision trees and the effects of feature extraction parameters for robust sensor network design. *Eksploatacja i Niezawodnosć - Maintenance and Reliability* 2017; 19 (1): 31-42, <https://doi.org/10.17531/ein.2017.1.5>.
14. Guidelines and methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment. SAE ARP 4761, 1996.
15. Hjelmgren K, Svensson S, Hannius O. Reliability analysis of a single-engine aircraft FADEC. *International Symposium on Product Quality and Integrity*; 1998 Jan 19-22; Anaheim., <https://ieeexplore.ieee.org/document/653811/references#references>.
16. Holub R, Vintr Z. Spolehlivost letadlove techniky (Dependability of aircraft). Brno University of Technology, electronic textbook, 2001.
17. IEC 60050-192:2015 International Electrotechnical Vocabulary (IEV) - Part 192: Dependability. International Electrotechnical Commission.
18. Janhuba L. The Integrated Method Utilizing Graph Theory and Fuzzy Logic for Safety and Reliability Assessment of Airborne Systems. Brno University of Technology; 2018, <https://doi.org/10.13164/conf.read.2018.4>.
19. Kornecki A, Zalewski J. Software certification for safety-critical systems: A status report. 2008 International Multiconference on Computer Science and Information Technology Wisia, 2008, <https://doi.org/10.1109/IMCSIT.2008.4747314>.
20. Li J, Wang Z, Ren Y, Yang D, Lv X. A novel reliability estimation method of multi-state system based on structure learning algorithm. *Eksploatacja i Niezawodnosć - Maintenance and Reliability* 2020; 22 (1): 170-178, <https://doi.org/10.17531/ein.2020.1.20>.
21. Li N, Lu Z, Zhou J. Reliability assessment based on Bayesian networks for full authority digital engine control systems. 11th International Conference on Reliability, Maintainability and Safety (ICRMS) 2016; 11, <https://doi.org/10.1109/ICRMS.2016.8050158>.
22. Liang H, Zhang S, Wei Z, Shao N. System Safety Analysis of a Full Authority Digital Engine Control System. *International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC) 2017*, <https://doi.org/10.1109/SDPC.2017.109>.
23. Liu H C. FMEA using uncertainty theories and MCDM methods. *FMEA Using Uncertainty Theories and MCDM Methods*. Springer, 2016; 13-27, [https://doi.org/10.1007/978-981-10-1466-6\\_2](https://doi.org/10.1007/978-981-10-1466-6_2).
24. Lu Z, Liang X, Zuo M J. Markov process based time limited dispatch analysis with constraints of both dispatch reliability and average safety levels. *Reliability Engineering & System Safety* 2017, 167: 84 - 94, <https://doi.org/10.1016/j.res.2017.05.031>.
25. Lu Z, Zhuo J, Li X. Monte Carlo simulation based time limited dispatch analysis with the constraint of dispatch reliability for electronic engine control systems. *Aerospace Science and Technology* 2018, 72: 397 - 408, <https://doi.org/10.1016/j.ast.2017.11.023>.
26. M250 turboshaft - RollsRoyce. [<https://www.rolls-royce.com/products-and-services/civil-aerospace/helicopters/m250-turboshaft.aspx#/>].
27. Pandian G, Das D, Li Ch, Zio E, Pecht M. A critique of reliability prediction techniques for avionics applications. *Chinese Journal of Aeronautics* 2018; 31(1): 10-20, <https://doi.org/10.1016/j.cja.2017.11.004>.
28. Prabhu S S, Kapil H, Lakshmaiah S H. Safety Critical Embedded Software: Significance and Approach to Reliability. 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, <https://doi.org/10.1109/ICACCI.2018.8554566>.
29. PT6 E-Series Engine - Pratt & Whitney. [<https://www.pwc.ca/en/products-and-services/products/helicopter-engines/pt6c>].
30. PT6C - Pratt and Whitney. [<https://www.pwc.ca/en/products-and-services/products/helicopter-engines/pt6c>].
31. Standard Practice for Safety Assessment of Systems and Equipment in Small Aircraft. ASTM F3230-17. 2017.
32. TM 5-698-4, Failure Modes, Effects and Criticality Analyses (FMECA) for Command, Control, Communications, Computer, Intelligence, Surveillance, and Reconnaissance (C4ISR) Facilities. Department of the Army. 2006, [https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB\\_ID=83559](https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=83559).
33. Youn W K, Hong S B, Oh K R, Ahn O S. Software certification of safety-critical avionic systems: DO-178C and its impacts. *IEEE Aerospace and Electronic Systems Magazine* 2015; 30(4):4-13, <https://doi.org/10.1109/MAES.2014.140109>.
34. Yu M N, Yu Z H, Jian H L, Xiao J Z. The optimization of RPN criticality analysis method in FMECA. 2009 International Conference on Apeerceiving Computing and Intelligence Analysis 2009; 166-170, <https://ieeexplore.ieee.org/document/5361125>.
35. Zeng Z, Kang R, Chen Y. Using PoF models to predict system reliability considering failure collaboration, *Chinese Journal of Aeronautics* 2016; 29(5): 1294-1301, <https://doi.org/10.1016/j.cja.2016.08.014>.
36. Zio E, Fan M, Zeng Z, Kang R. Application of reliability technologies in civil aviation: Lessons learnt and perspectives. *Chinese Journal of Aeronautics* 2019; 32(1): 143-158, <https://doi.org/10.1016/j.cja.2018.05.014>.

Article citation info:

Fan L, Wang K, Fan D. A combined universal generating function and physics of failure Reliability Prediction Method for an LED driver. *Eksplloatacja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 74–83, <http://dx.doi.org/10.17531/ein.2021.1.8>.

## A combined universal generating function and physics of failure Reliability Prediction Method for an LED driver

Liming Fan<sup>a</sup>, Kunsheng Wang<sup>a</sup>, Dongming Fan<sup>b\*</sup>

<sup>a</sup>China Aerospace Academy of Systems Science and Engineering, Beijing, 100048, P. R. China

<sup>b</sup>School of Reliability and Systems Engineering, Beihang University, Beijing, PR China

Indexed by:



### Highlights

- The gap is bridged between the component level and the circuit system level.
- A multi-phase model is established according to function processes of circuit.
- A universal numerical approach is provided for predicting the lifetime of LED drivers.

### Abstract

The accurate and effective reliability prediction of light emitting diode (LED) drivers has emerged as a key issue in LED applications. However, previous studies have mainly focused on the reliability of electrolytic capacitors or other single components while ignoring circuit topology. In this study, universal generating function (UGF) and physics of failure (PoF) are integrated to predict the reliability of LED drivers. Utilizing PoF, lifetime data for each component are obtained. A system reliability model with multi-phase is established, and system reliability can be predicted using UGF. Illustrated by a two-channel LED driver, the beneficial effects of capacitors and MOSFETs for the reliability of LED drivers is verified. This study (i) provides a universal numerical approach to predict the lifetime of LED drivers considering circuit topology, (ii) enhances the modelling and reliability evaluation of circuits, and (iii) bridges the gap between component and circuit system levels.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

reliability prediction, LED driver, reliability modelling, physics of failure, universal generating function.

### Notation

$x_1, \dots, x_n$	Independent variables in system	$A$	The coefficient
$u_{x_i}(z)$	Probability distribution of each variable	$k$	The Boltzmann constant
$P_{ij}$	Probability of variable $i$ in the possible state $j$	$E_a$	The activation energy
$k_{ij}$	The number of states of variable $x_i$	$T$	Absolute temperature
$t_A$	The failure time of component $A$	$ESR_T$	The initial ESR
$F_A$	The cumulative distribution of component $A$	$C$	The temperature-dependent degradation rates of capacitors
$f_A(t)$	The probability density function of component $A$	$C_0$	The base degradation rate of capacitors
$Q_1, Q_2$	The MOSFET	$A_{cap}$	The surface area of the electrolytic capacitor
$N_1, N_2$	The windings	$d$	The average distance of the electrolytic capacitor
$C_{1-4}$	The capacity	$\varpi$	The conductivity of the electrolytic capacitor
$LS_1, LS_2$	The LED strings	$RH$	The relative humidity
$D_1, D_2$	The diodes	$V_{in}$	The power supply
$L_{lk1}, L_{lk2}$	The inductors	$X_{1-3}$	The description of the structure in circuit

(\*) Corresponding author.

E-mail addresses: L. Fan - [liming\\_fan2020@126.com](mailto:liming_fan2020@126.com), K. Wang - [kunsheng\\_wang@163.com](mailto:kunsheng_wang@163.com), D. Fan - [fdm314314@126.com](mailto:fdm314314@126.com)

## Abbreviations

SSL	Solid-State Lighting
LED	Light-Emitting Diode
LCD	Liquid Crystal Display
PF	Particle Filter
RUL	Remaining Useful Time

HCI	Hot Carrier Injection
ESR	Equivalent Series Resistance
UGF	Universal Generating Function
TTF	Time and Time to Failure
PoF	Physics of Failure

## 1. Introduction

As a new type of solid-state lighting (SSL), LED applications are rapidly increasing in many domains, including industrial/decorative illumination, liquid crystal display (LCD) backlighting, and automobile lamps [16] [22]. In addition to their wide range of applications, LEDs provide longer operating times and higher levels of reliability. However, compared with the lifetime of LED chips, LED drivers are found to have shorter lifetimes, as reported by the U.S. Department of Energy [33]. This discrepancy can contribute to up to 52% of total system failures [25] [29], as the operating time and reliability of LED systems are largely dependent on LED drivers [36].

In general, researchers have focused on two types of LED drivers: linear mode drivers and switch mode drivers [32] [39]. By studying LED driver schematics, many researchers have developed various approaches to predict the reliability of LED drivers. In terms of linear mode drivers, Song Lan et al. [17] presented a black box method to locate the weak point and study the reliability of a regulated LED driver circuit. A particle filter (PF) method [14] was implemented and combined with nonlinear least squares (NLS) and nonlinear mixed-effect estimation (NLME) to predict the remaining useful time (RUL). Lan et al. [15] [16] studied the degradation of a linear mode high-power LED driver, and hot carrier injection (HCI) has been revealed as the main degradation mechanism. In terms of switch mode drivers, Koosuke et al. [8] presented a diagnosis method for an output smoothing capacitor, and the lifecycle ageing rate was projected for the active circuit over the system life thanks to equivalent series resistance (ESR). Based on the PoF, Zhou [40] predicted the RUL by utilizing an accelerated ageing test.

However, several issues exist in current studies: (i) researchers have only focused on the electrolytic capacitors or other single types of components in LED drivers; (ii) the circuit topology and multiphase status of LED drivers have been ignored in previous analyses; and (iii) there has been no assurance that degradation laws and other particular factors will not change during acceleration tests. These drawbacks can lead to inaccurate prediction results, especially for complex LED driver systems with multiple phases.

To address these issues, this study aims to establish a reliability prediction model based on the operation profile of a circuit with multiple phases and bridge the gap between the component level and circuit system level. First, utilizing a PoF model, various parameters are discretized based on the failure mechanisms and the distribution model. Then, the failure probability curve can be fitted in terms of the correlation between the time and time to failure (TTF). Third, a specific two-channel switch mode LED driver is illustrated to establish the reliability model. Finally, system reliability prediction results are estimated using a UGF. The main contributions are as follows: (1) The proposed approach can provide a way to bridge the gap between the component level and circuit system level using a combined UGF and PoF approach. (2) This universal multiphase modelling approach, based on the function process of the circuit, can predict the lifetime of an LED driver.

The remainder of this paper is organized as follows. Some basic concepts, including the UGF and TTF, are introduced in Section 2. Based on the framework of the novel reliability prediction approach, the detailed modelling, analysis, and prediction of two-channel LED drivers are proposed in Section 3. The experimental procedure and re-

sults are proposed in Section 4. Finally, a discussion and conclusions are provided in Sections 5 and 6.

## 2. Basic conception

### 2.1. Universal generating function (UGF)

Generally, the reliability prediction of a circuit follows competition rules [12]. The process regards the circuit as a cascaded structure and takes the minimum lifetime of components as the lifetime. However, a competition model is clearly inaccurate in reliability prediction, especially for circuit systems with high complexity such as series-parallel and backup structures. To bridge the gap between the component level and circuit system level, the UGF approach is utilized in this paper.

The UGF approach [2] [26] [34] is typically used to describe the relationship of reliability between components and a system. It was first introduced by Ushakov [34] and generally applied in reliability assessment with multi-state components [26]. Assume that there are  $n$  independent components  $x_1, \dots, x_n$ ,  $k_i$  is the possible state of component  $x_i$ , and  $p_{ij}$  is the probability of component  $x_i$  at state  $j$ . The probability distribution of each variable is discretely represented by a u-function ( $u_{x_i}(e^t)$ ) and then replaces  $e^t$  as  $z$ , as shown in (1). In an evaluation of system reliability, the u-function combines the reliability and performance to describe the component state distribution:

$$u_{x_i}(z) = \sum_{j=1}^{k_i} p_{ij} \cdot z^j \quad (1)$$

If the component is a binary unit, state 0 is the failure state, and state 1 is the success state. The u-function of component  $x_i$  can be denoted as follows:

$$u_{x_i}(z) = (1 - p_i)z^0 + p_i z^1 \quad (2)$$

According to a series or parallel structure, the u-function of the structure can be described as follows [2] [18]:

$$U_{series}(z) = \min \{x_1, \dots, x_n\} = u_{x_1}(z) \otimes_{\times} u_{x_2}(z) \cdots \otimes_{\times} u_{x_n}(z) \quad (3)$$

$$U_{parallel}(z) = \max \{x_1, \dots, x_n\} = u_{x_1}(z) \otimes_{\max} u_{x_2}(z) \cdots \otimes_{\max} u_{x_n}(z) \quad (4)$$

The reliability of the system can be described as shown in (4), in which  $z$  represents the no-fault status:

$$R = U'(z) \quad (5)$$

### 2.2. Time-to-failure distribution for multiple phases

An LED driver circuit has the following characteristics: (i) the circuit system can be divided into  $i$  phases based on the direction of the current and the operation of the circuit; (ii) the components have

various operation times; (iii) the TTF of components obeys various distributions; and (iv)  $i$  phases form a cycle, and the multiple cycles are serially structured [23] [35].

In accordance with the terminology introduced in [6] [7] [28], if the random variable  $A$  is in state  $i(1 \leq i \leq n)$ , or  $A=i$  it simply means that  $A$  has failed in the  $t_A \in [(i-1)\Delta, i\Delta]$  interval, where  $t_A$  is the failure time of component  $A$ ,  $F_A$  is the cumulative distribution,  $\Delta$  is the interval length  $\Delta=T/n$ , and  $n$  is the time granularity:

$$P(A=i) = P((i-1)\Delta < t_A \leq i\Delta) = \int_{(i-1)\Delta}^{i\Delta} f_A(t) dt \quad (6)$$

$$= F_A(i\Delta) - F_A((i-1)\Delta)$$

Similarly, if  $A$  is said to be in state  $(n+1)$ , then  $A$  has survived the mission time  $T$ :

$$P(A=n+1) = P(t_A > T) = \int_T^{\infty} f_A(t) dt = 1 - F_A(T) \quad (7)$$

### 3. Reliability Prediction Approach for LED Driver

#### 3.1. Framework of reliability prediction approach

In practice, the normal degenerative process of a specific LED driver system can take years to display a conspicuous change. To address this restriction, a novel reliability prediction approach based on the PoF is proposed, which is shown in Fig. 1. Integrated with the failure mechanisms and topological structure of the system, this ap-

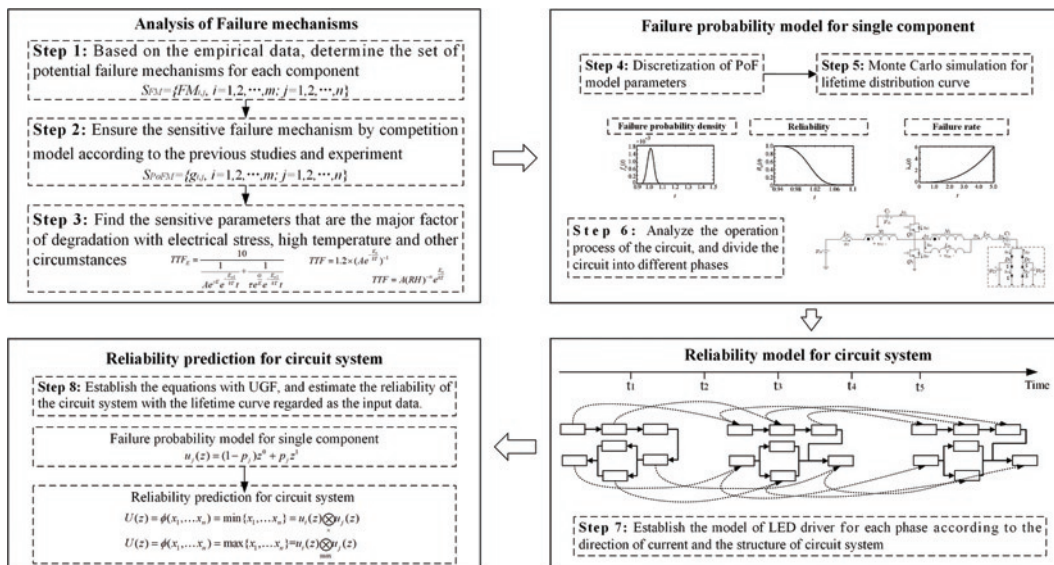


Fig. 1. The process of reliability prediction approach

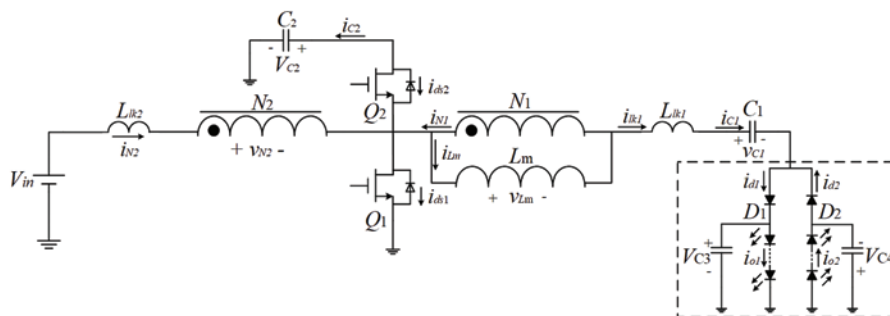


Fig. 2. The structure of the two-channel LED driver

proach provides a universal numerical and simulation approach for predicting the lifetime of LED drivers. With the analysis of failure mechanisms, the major factors of degradation of components are determined (Step 1-Step 3). By utilizing the PoF method, the lifetime data for each component can be obtained by random sampling simulations (Step 4-Step 5). Then, the operation process of the circuit can be analysed and divided into multiple phases (Step 6). A reliability model can be established for the LED driver at each phase (Step 7). Finally, the reliability of the circuit system can be estimated with the lifetime curve above as the input data according to the topological structure of the multiphase circuit utilizing a UGF (Step 8).

#### 3.2. Failure probability model for a single component

##### 3.2.1. Topology of LED driver

It is well known that the balance of current is a vital function for LED drivers [32]. A switch mode LED driver has the dominant advantage in balancing the current with full bridge and half-bridge topologies. This mode features a simple structure and precise current balance. Thus, switch mode LED drivers have received increasing attention [19] [37]. Compared with the traditional topology of a switch mode LED driver, a two-channel non-isolated LED driver has the following advantages [3] [4] [9]: (i) owing to a DC-blocking capacitor, LED currents can be balanced automatically; (ii) owing to the combination of the DC-blocking capacitor and a coupled inductor, a high-output voltage gain can be achieved; (iii) the voltage spike from the leakage inductance of the coupled inductor can be recycled to the capacitor  $C_2$  so that low-voltage MOSFETs can be used; (iv) the proposed two-channel LED driver can be extended to a multi-channel LED driver to satisfy industry requirements.

In terms of the switch mode LED driver [10] [40], a two-channel non-isolated LED driver is regarded as an example for this study, whose topology is shown in Fig. 2. It contains two MOSFETs  $Q_1$  and  $Q_2$ , a coupled inductor composed of a primary winding  $N_1$  and a secondary winding  $N_2$ , two energy-electrolytic capacitors  $C_1$  and  $C_2$ , two output diodes  $D_1$  and  $D_2$ , two output capacitors  $C_3$  and  $C_4$ , and two LED strings  $LS_1$  and  $LS_2$ . Moreover, the coupled inductor can be modelled as a primary winding  $N_1$  and a secondary winding  $N_2$  connected in series with a leakage inductor  $L_{lk2}$ , a magnetizing inductor  $L_m$  connected in parallel with  $N_1$ , and a leakage inductor  $L_{lk1}$ . In addition, the current flows and the anode and cathode labels are shown in Fig. 2 [3].

addition, the current flows and the anode and cathode labels are shown in Fig. 2 [3].

##### 3.2.2. PoF models

For different types of LED drivers, the topology of the circuit and component constitution are disparate, and the functional components are associated with different failure modes and mechanisms [38]. Previous reports [8] [14] [15] have discussed the failure mechanisms of amplifiers, diodes, resistors and capacitors in linear mode LED

Table 1. Mechanism models and parameter description

Component	Mechanism and formulation	Parameter description	Distribution	Feature parameter	Other parameters
<b>MOSFET</b> $Q_1/Q_2$ [15] [32][40]	Time-dependent dielectric breakdown $TTF = (Ae^{V^m} e^{-\frac{E_a}{kT}})^{-1}$	$A, E_a, k, T$ are the coefficient, activation energy, Boltzmann's constant and absolute temperature, respectively; $V$ represents voltage and $m$ is a constant; $m = 2.9$ [40]	$A: N(\mu, \sigma)$ $T: N(\mu, \sigma)$ $V: U(a, b)$	$A: \mu = 6.98 \times 10^{-5}$ $\sigma = 4.2 \times 10^{-6}$ $T: \mu = 328, \sigma = 20$ $V: \mu = 0.6\text{mV}$ [24]	$E_a = 1.3\text{eV}$ $k = 8.62 \times 10^5$ [24]
<b>Coupled inductor</b> $(L_{lk1} + N_1^2) / (L_{lk2} + N_2^2)$ [1][5]	Corrosion $TTF = (A(RH)^{-n} e^{-\frac{E_a}{kT}})^{-1}$	$A$ is the constant related to the corrosion area; $RH$ is the relative humidity; $n$ is the empirical constant; $E_a$ is the activation energy (eV); $k$ is the Boltzmann constant (J/K); $T$ is the absolute temperature (K)	$RH: U(a, b)$ $A: N(\mu, \sigma)$ $T: N(\mu, \sigma)$	$RH: \mu = 0.6$ [5] $A: \mu = 1.785 \times 10^{-5}$ $\sigma = 1 \times 10^{-6}$ $T: \mu = 328, \sigma = 20$	$E_a = 0.771\text{eV}$ $n = 3$ $k = 8.62 \times 10^5$ [30]
<b>Diode</b> [40]	Arrhenius degeneration $TTF = (Ae^{\frac{E_a}{kT}})^{-1}$	$A$ is the coefficient; $k$ is the Boltzmann constant (J/K); $E_a$ is the activation energy (eV)	$A: N(\mu, \sigma)$ $T: N(\mu, \sigma)$	$A: \mu = 8.4063 \times 10^{-5}$ $\sigma = 8 \times 10^{-7}$ $T: \mu = 328, \sigma = 20$	$E_a = 1.2\text{eV}$ $k = 8.62 \times 10^5$ [30]
<b>Capacitor</b> $C_1 / C_2$ [30]	Decrease of the effective surface area $ESR(t) = ESR_T \cdot e^{C \cdot t}$ $C(t) = C_0 \cdot e^{-E_a/(k \cdot T)}$	$ESR_T$ is the initial ESR; $C$ describes temperature-dependent degradation rates; $C_0$ is the base degradation rate; $E_a$ and $k$ are the activation energies and Boltzmann constant; $A_{cap}, d, \varpi$ are the surface area of the electrolytic capacitor, average distance, and conductivity [40].	$C: N(\mu, \sigma)$ $T: N(\mu, \sigma)$	$C: \mu = 3.46 \times 10^{-5}$ $\sigma = 2 \times 10^{-6}$ $T: \mu = 328, \sigma = 20$ [30]	$E_a = 0.7\text{eV}$ $k = 8.62 \times 10^5$ [30]
<b>Capacitor</b> $C_3 / C_4$ [31]	$ESR_T = \frac{d}{A_{cap} \cdot \varpi}$				

drivers. The reliability of a switch mode LED driver was analyzed in [30], and detailed degradation mechanisms for electrolytic capacitors, MOSFETs, inductors, diodes and sampling resistors were given.

The degradation process of a two-channel LED driver is extremely complicated and is usually caused by the gradual deterioration of components over a long period of time due to electrical stress, temperature effects, material characteristics and other circumstances. Generally, the degradation of various components can be related to specific performance parameters, and the various mechanisms can have competitive relations. Hence, according to previous research [5] [13] [20] [24] [30] [31], we focus attention on the greatest effect of each component, and the mechanism models are shown in Table 1.

To obtain the parameters of the mechanism models, the least square method is applied to fit the experimental data under various conditions. Table 1 proposes the mean values and the standard deviations, and the details of experimental data analysis, test procedures, and the fitting analysis are referred to [21].

Taking the diode as an example, we first sample the parameter based on the distribution, and the TTF data of the component can be obtained according to the number of samples. Then, we organize all the TTF data from small to large and divide the data into several groups utilizing an appropriate class. By calculating the frequency of each group, the cumulative probability can be obtained, and finally, the fitted curve between the TTF and failure probability is achieved.

In Fig. 3 (a), the x-axis represents time, and the y-axis represents the failure probability, namely, the unreliability of the component diode. The relationship between time and failure probability of the diode is shown in Fig. 3 (a). Moreover, as the various parameters of distributions change, the life curve can be different, as shown in Fig. 3 (b).

As shown in Fig. 3, the failure probability curve of the diode is proposed using the simulation method. As time goes on, the failure probability gradually increases and tends toward 1. Utilizing a temperature parameter that obeys the normal distribu-

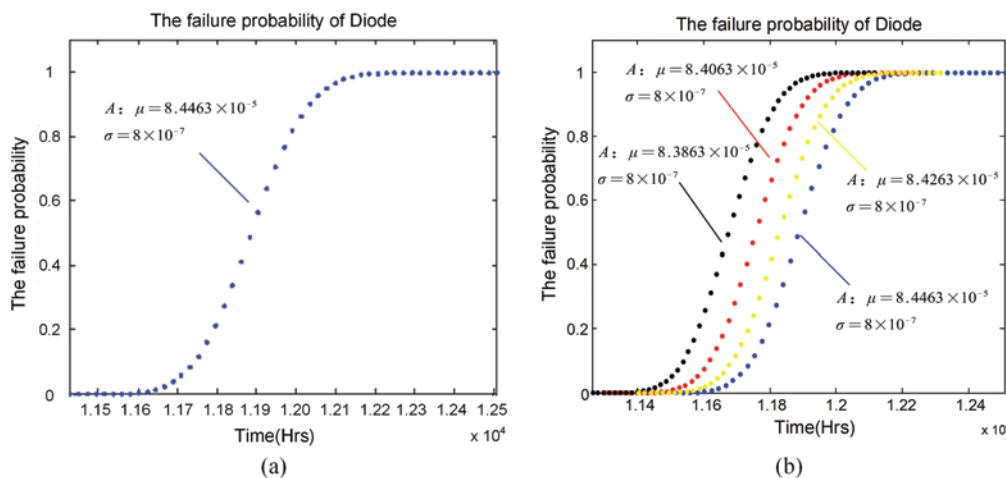


Fig. 3. The failure probability of diode: (a) failure probability curve; (b) the failure probability curve with various parameters 'A'

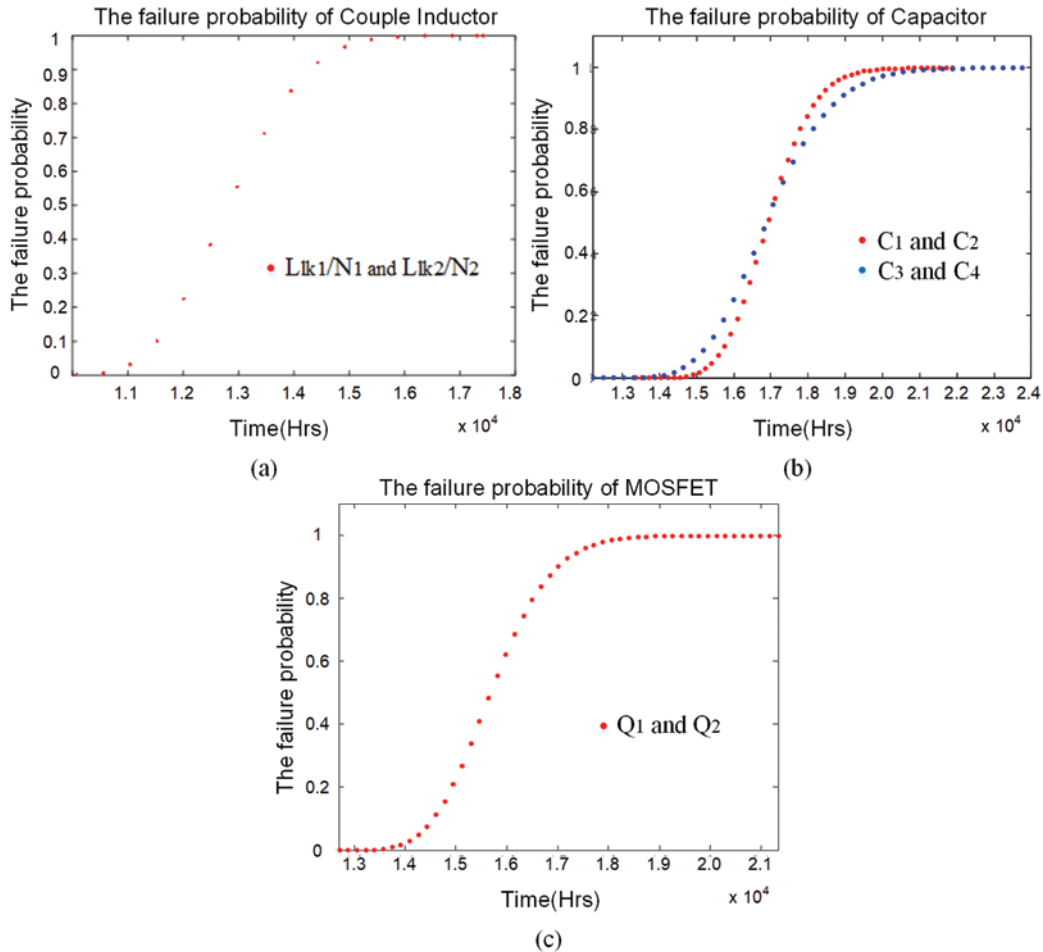


Fig. 4. The failure probability of other components: (a) the failure probability of the inductor; (b) the failure probability of the capacitor; (c) the failure probability of the MOSFET

tion, the statistical life fitting curve can be regarded as a comprehensive/general result for the components considering all the prevailing situations. Hence, the probability data at any time can be obtained based on the curve and regarded as an input prior probability to the integrated model in the next section. To simplify the analysis procedure, we assume that coupled inductors  $(L_{lk1} + N_1)/(L_{lk2} + N_2)$ ,  $C_1$  and  $C_2$ ,  $C_3$  and  $C_4$ ,  $Q_1$  and  $Q_2$  are the same as each other. The failure probability is shown in Fig. 4.

### 3.3. Reliability prediction for the two-channel LED driver

#### 3.3.1. Modelling rules for LED driver circuit

The failure probability of the components and their corresponding lifetime curves are mentioned above. To bridge the gap between the component level and circuit system level, modelling rules and procedures are proposed [11] [27]:

- (i) Define the circuit system. In this step, the scope of the system and reliability index are defined. Furthermore, its function and topology, including the components, are determined.
- (ii) Determine the boundary conditions. The input, output and interface of the system are defined. Certain inductance and energy storage components can also be the input signals when they are at high potential.
- (iii) Establish success criteria. Clearly define the success state and the output being focused on.
- (iv) Analyze the circle process. Dismantle the operational process of the circuit based on the direction of the current.

#### 3.3.2. Operation process of the two-channel LED driver

According to the topology structure, the operation process of the LED driver can be divided into six phases [3] [27], which are shown in Fig. 5. Switches  $Q_1$  and  $Q_2$  are turned on at different times, and their capacities are discharged or charged according to the electric potential. In addition, components in red represent the “OFF” state, and the “OFF” components will not deteriorate during the phase.

Model (a)-phase 1: As shown in Fig. 2 and Fig. 5 (a),  $Q_1$  is turned ON, and  $Q_2$  is OFF. During this state, the voltage is imposed on  $L_m$ . Meanwhile,  $C_1$  is discharged, so  $i_{C1}$  or  $i_{lk1}$  is increasing in the opposite direction. Moreover,  $D_1$  is reverse biased, but  $D_2$  is forward biased. Hence, the LED string  $L_{S1}$  is supplied by  $C_3$ , whereas the LED string  $L_{S2}$  is supplied by  $i_{lk}$  in the opposite direction. This state ends when  $Q_1$  is turned OFF [3].

Model (b)-phase 2: as shown in Fig. 2 and Fig. 5 (b), the switch  $Q_1$  switches from ON to OFF, and the switch  $Q_2$  still remains OFF. During this state, the voltage  $V_m - V_{C2}$ , which is a negative voltage, is imposed on  $L_{lk2}$  and  $L_m$ . During this blanking time, the body diode of switch  $Q_2$  is forward biased. Because of the short time of model (b)-phase 2, we hypothesize that the body diode of switch  $Q_2$  will have the same degeneration as switch  $Q_2$ . Moreover,  $D_1$  is forward biased, but  $D_2$  is reverse biased. Hence, the LED string  $L_{S1}$  is supplied by  $i_{lk}$ , whereas the LED string  $L_{S2}$  is supplied by  $C_4$ . This state ends when  $Q_2$  is turned ON.

Model (c)-phase 3: before state 3 begins, the body diode of the switch  $Q_2$  conducts. Thus,  $Q_2$  is turned ON with zero-voltage switching, as shown in Fig. 2 and Fig. 5 (c), but  $Q_1$  still remains turned OFF. In addition, the behavior of the two LED strings is the



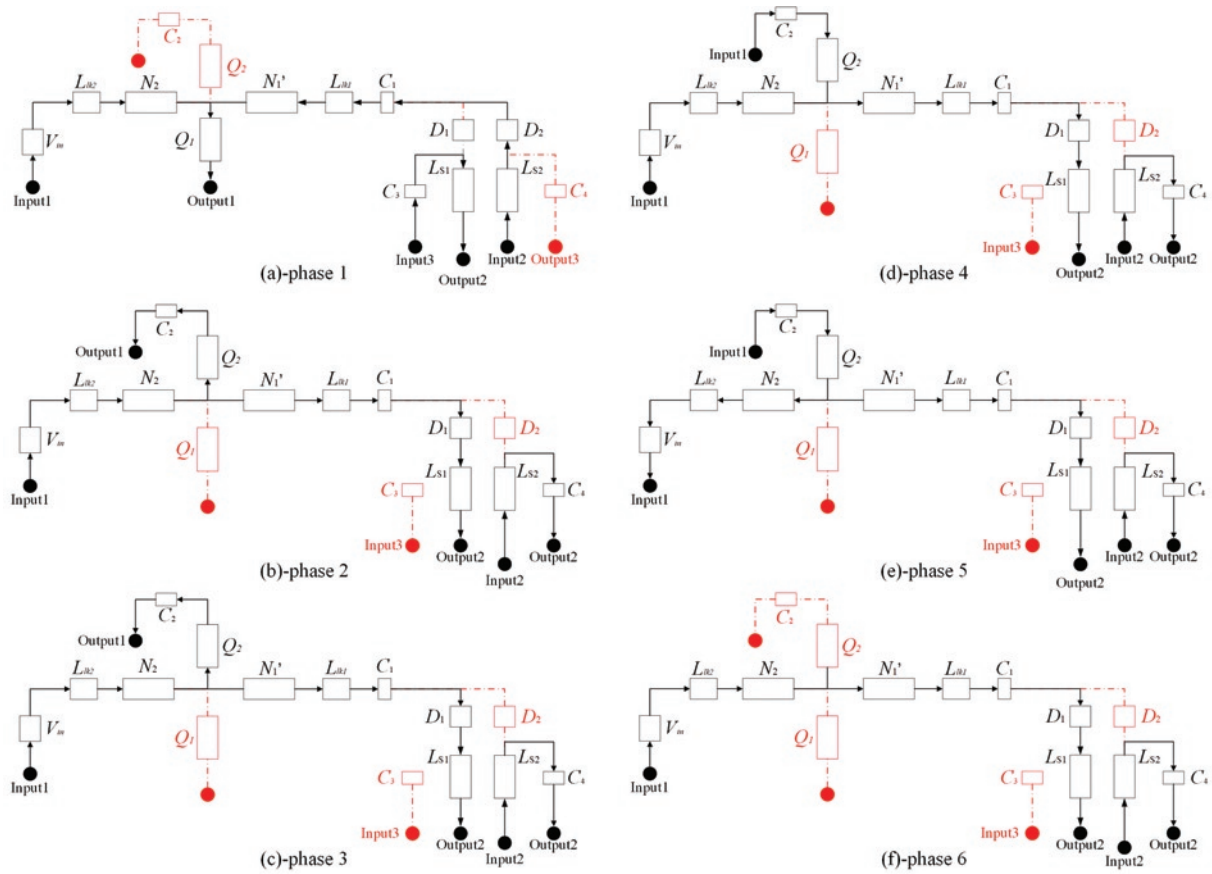


Fig. 5. The operational model of the circuit with various phases

same as that in state 2. Once  $i_{N2}$  is equal to  $i_{lk1}$ ,  $i_{ds2}$  reaches zero, and this state ends.

Model (d)-phase 4: as shown in Fig. 2 and Fig. 5 (d), the switch  $Q_1$  is still OFF, and the switch  $Q_2$  is still ON. During this state, the capacitor  $C_2$  is discharged, and the current  $i_{C2}$  begins to increase in the opposite direction. In addition, the behavior of the two LED strings is the same as that in model (b). This state ends when  $i_{N2}$  drops to zero.

Model (e)-phase 5: as shown in Fig. 2 and Fig. 5 (e), the switch  $Q_1$  is still OFF, and the switch  $Q_2$  is still ON. The direction of the current is opposite model (d)-phase 4. In addition, the behavior of the two LED strings is the same as that in model (b). This state ends when  $Q_2$  is turned OFF.

Model (f)-phase 6: as shown in Fig. 2 and Fig. 5 (f), switch  $Q_1$  is still OFF, and switch  $Q_2$  turns OFF. The direction of the current is the same as model (d)-phase 4. This state ends when  $Q_1$  is turned ON, and the next cycle is repeated.

Additionally, with the time-varying of current, the oscillations of electrical waveforms will create oscillations on the components, especially for the coupled inductors. Nonetheless, the problem oscillation is not the focus of our research, and thus, the effects on reliability are ignored in this study. This will be discussed in our future study.

### 3.3.3. System reliability evaluation utilizing UGF

We take model (a) as an example: the power supply  $V_m$  and coupled inductor ( $L_{lk2}$  and  $N_2$ ) are linked as a cascade architecture, denoted by  $X_1$ ; strings  $L_{s2}$ , diode  $D_2$ , capacitor  $C_1$  and coupled inductor ( $L_{lk1}$  and equivalent  $N_1'$ ) are also linked as a cascade architecture, denoted by  $X_2$ . Then,  $X_2$  is linked with coupled inductor  $L_{lk2}/N_2$  and MOSFET  $Q_1$  (denoted by  $X_4$ ). Capacitor  $C_3$  is in series with strings  $L_{s1}$  (denoted by  $X_3$ ), as shown in Fig. 6.

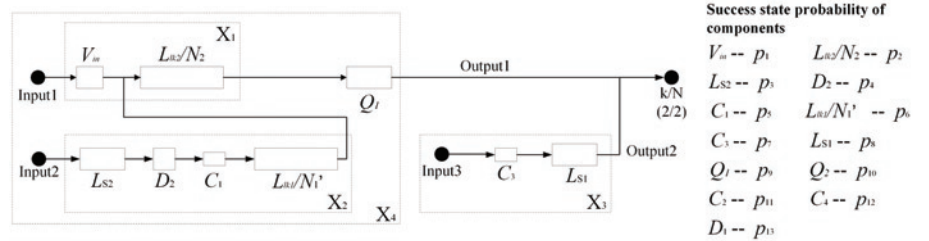


Fig. 6. The equivalent circuit of model (a)-phase 1

We assume that all the components are binary, and a logic gate of  $k$ -out-of- $N$  is used to describe the output of the system. The structure function of the circuit can be described as follows:

$$\begin{aligned}
 X &= \phi(V_m, L_{lk2}/N_2, L_{s2}, D_2, C_1, L_{lk1}/N_1', C_3, L_{s1}, Q_1) \\
 &= k/N \left\{ (\min(\max(V_m \cdot L_{lk2}/N_2 \cdot L_{s2} \cdot D_2 \cdot C_1 \cdot L_{lk1}/N_1') \cdot Q_1)), (C_3 \cdot L_{s1}) \right\} \quad (8) \\
 &= k/N \left\{ \min((\max(X_1, X_2) \cdot Q_1)), X_3 \right\}
 \end{aligned}$$

$$\begin{aligned}
 U_{X_1}(z) &= \left[ (1 - p_1)z^0 + p_1z^1 \right] \left[ (1 - p_2)z^0 + p_2z^1 \right] \\
 &= p_{1-2}z^1 + (1 - p_{1-2})z^0 \quad (9)
 \end{aligned}$$

$$U_{X_2}(z) = p_{3-6}z^1 + (1 - p_{3-6})z^0 \quad (10)$$

$$U_{X_3}(z) = p_7p_8z^1 + (1 - p_7p_8)z^0 \quad (11)$$

$$U_{X_4}(z) = (p_{1-2}p_9 + p_{2-6}p_9 - p_{1-6}p_9)z^1 + (1 - p_{1-2}p_9 - p_{2-6}p_9 + p_{1-6}p_9)z^0 \quad (12)$$

$$U_{X_5}(z) = U_{X_4}(z) \otimes_{k/N} U_{X_3}(z) \\ = (p_{1-2}p_{7-9} + p_{2-9} - p_{1-9})z^1 + (1 - p_{1-2}p_{7-9} - p_{2-9} + p_{1-9})z^0 \quad (13)$$

where  $p_1 - p_{13}$  represent the probability of components in the success state, as listed in Fig. 6. The reliability of the circuit system model (a)-phase 1 in Fig. 5 can be expressed as:

$$R = U'_{X_5}(1) \quad (14)$$

Other phases in Fig. 5 can be calculated using the same process above. Focusing on analyzing the impact of component failure on the LED driver, some assumptions are made to make the analysis clear and easy. As the input of the LED driver,  $V_{in}$  has a minimal failure probability, and it can be neglected. Likewise,  $L_{s1}$  and  $L_{s2}$ , the light-emitting units that are the output of the system, are also ignored, namely,  $p_1 = p_3 = p_8 = 1$ . In addition, according to the failure logic of the system, the system is working properly only if all phases are in the normal state. Therefore, the various phases are regarded as a series-mode frame to evaluate the reliability of the system in this study.

According to the approach above, the reliability prediction curve for the two-channel LED driver is displayed in Fig. 7 (a). The simulation results clearly show that the reliability of the circuit gradually degrades as a function of time due to the ageing process. It also appears that the failure rate increases sharply before 18000 h and then begins to flatten gradually in Fig. 7 (b).

#### 4. Experiments and Validation

To validate the proposed method, an experiment was designed to predict the reliability of the LED driver. The proposed type of LED drivers was selected, and 60 samples were made by Everlight Elec-

tronics, Ltd., which were used for testing. The real product is shown in Fig. 8 (a). In Fig. 8 (a), the two electrolytic capacitors are the components  $C_1$  and  $C_2$ . The output capacitors  $C_3$  and  $C_4$  are parallel with light emitting diodes and thus external to the board. Owing to the necessity of the diodes to glow normally, as shown in Fig. 2, the output capacitors  $C_3$  and  $C_4$  should be considered in the system reliability model.

According to the sample, if the nominal constant current flowing through the LED is 350 mA, the corresponding nominal forward voltage is approximately 3.5 V. The loads of the proposed drivers were kept constant by using five 3-W white LEDs for each string. In addition, the variations in the LED forward voltage are  $\pm 10\%$  of the nominal value, and the output power directly reflects the degradation of the LED driver, which can be conveniently measured. The main specifications of the LED driver are shown in Table 2.

Fig. 8 (b) shows the experimental setup. A local oven was used to contain the LED driver and to provide local ambient temperature control. The input voltage is controlled by the programmable power source, and the output voltage can be measured periodically by the voltage meter, thus, the state of the LED driver can be obtained. This procedure was repeated until all of the samples were in the failure state.

During the validation test, a total of 60 samples were aged individually at an ambient temperature of 328 K. With a constant input voltage, the working hours of each sample were recorded by the acceleration test until the LED driver was in a state of failure. Based on empirical data, the failure criterion is defined as within 10% of the rated current increase of the LED channel [30].

Fig. 9 shows the predicted reliability of the LED driver versus the experimental results. Due to the time limit and lab constraints, the samples are not sufficient for extensive experiments, and groups of various ambient temperatures were not considered in the experiment. Nevertheless, the prediction results are in good agreement with the

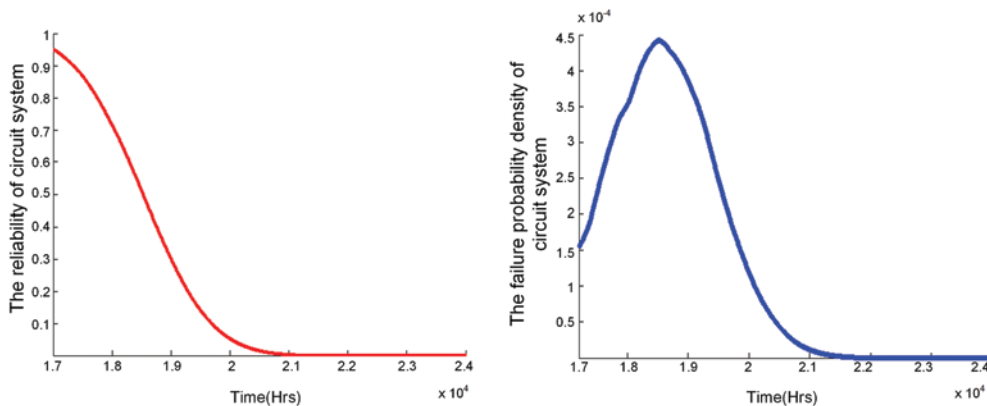


Fig. 7. Reliability prediction curve of circuit system: (a) reliability curve of system; (b) the failure probability density of system.

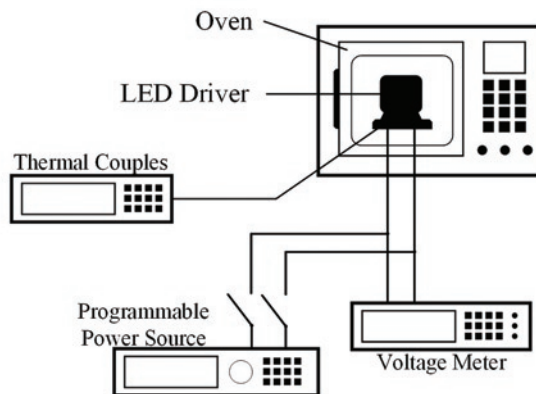
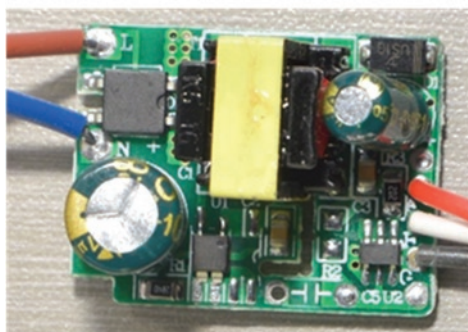


Fig. 8. (a) The layout of the specified LED driver; (b) experimental setup diagram.

Table 2. System specifications of the LED driver

System parameters	Specifications
Input voltage	12V ± 10%
Nominal LED channel voltage	3.5V × 5 = 17.5V ± 10%
Nominal LED channel current	350 mA
Switching frequency	100 kHz

test results. The maximum error between the test and simulation is less than 10%.

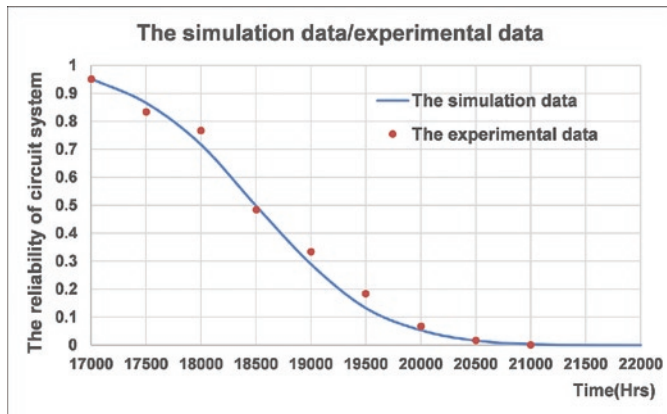


Fig. 9. Predicted reliability versus experimental data.

## 5. Discussions

### 5.1. Relevant analysis

One of the most important assumptions in modelling is that the components remain independent in the operation condition. Fig. 10 shows the reliability simulation results of the circuit system. The red line describes the degradation curve without considering the relation among the components, and the blue line displays the degradation curve with the dependence included. The results show that the reliability without considering the dependence on degrades more signifi-

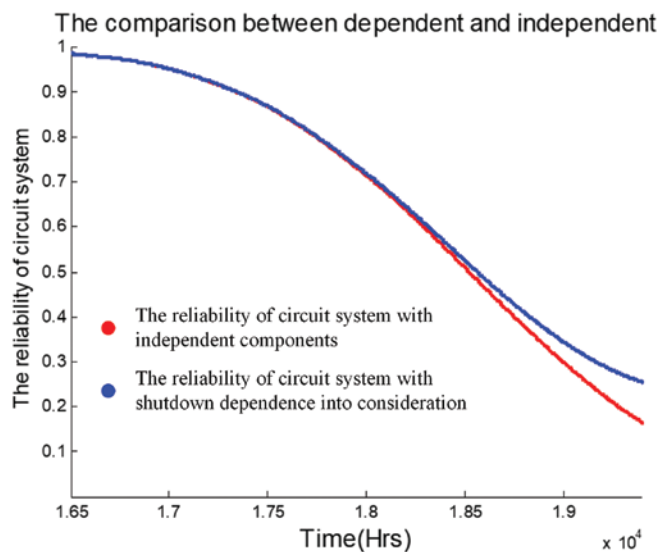


Fig. 10. Comparisons of degradation between the independent and dependent components.

cantly. When the time is approximately 19500 h, the reliability that considers the relation is 0.2459, which is nearly double the value of the scenario that does not. As shown in Fig. 10, the assessment of the circuit system reliability could be more accurate with the dynamic relation between components.

### 5.2. Structure importance analysis of components

Fig. 11 displays the structure importance of various components. The attribute of each component to the output of the circuit system can be described quantitatively. Fig. 11 clearly shows that the higher the attribute of a component is, the greater the importance for the output of the circuit system. In Fig. 11,  $p_7$  ( $C_3$ ),  $p_9$  ( $Q_1$ ), and  $p_{12}$  ( $C_4$ ) are vital components according to the entire circuit system. These components are the key objects in the structure of the circuit system.

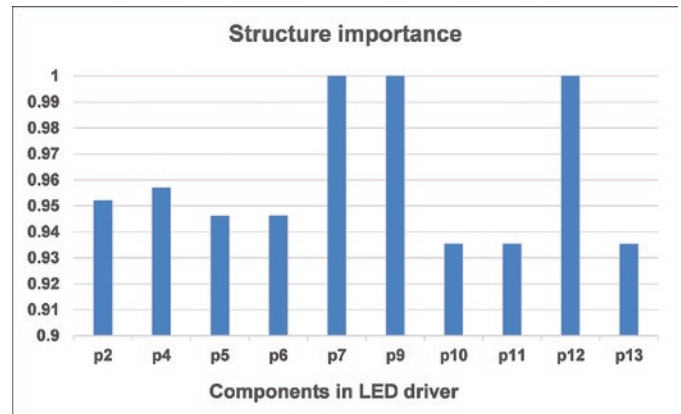


Fig. 11. The structure importance of various components

### 5.3. Sensitivity analysis of components

A sensitivity analysis of various components is proposed in this section. As shown in Fig. 12,  $p_{12}$  ( $C_4$ ) is the most sensitive among the vital components mentioned above. This means that a slight improvement acting on the component can result in a vital effect being obtained on the LED driver. Although  $p_7$  ( $C_3$ ) and  $p_9$  ( $Q_1$ ) are vital components in the analysis of structural importance, the two components work in the first phase of the operation process. Compared with  $p_{12}$  ( $C_4$ ), the improvement of the system reliability is nearly the same, as shown in Fig. 12.

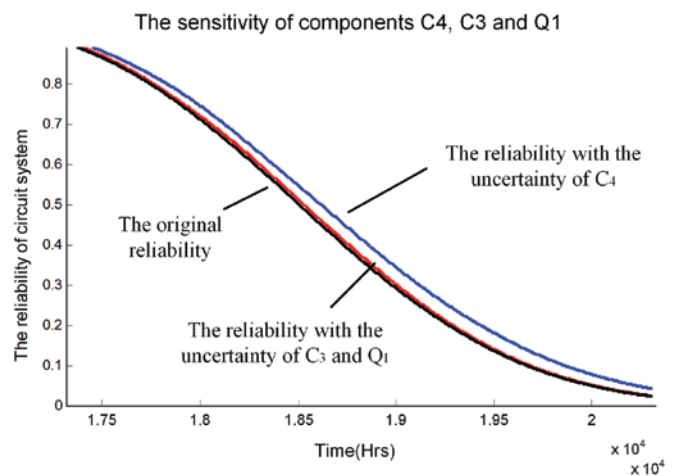


Fig. 12. Sensitivity analysis of components.

## 5.4. Sensitivity analysis of temperature

Fig. 13 shows the reliability sensitivity of the system to the temperature of components within a given ambient temperature. Owing to the time-varying thermal fields, the temperature of the components is unknown and uncertain. The system reliability is simulated in this study under various component temperatures (308 K, 328 K and 348 K). As shown in Fig. 13, the simulation of system reliability under 328 K is proposed, and the data with 308 K and 348 K are proposed in Fig. 13. Based on the various temperatures of the components, the failure probabilities of the system are 0.942 (348 K), 0.868 (328 K) and 0.753 (308 K) at 19500 h.

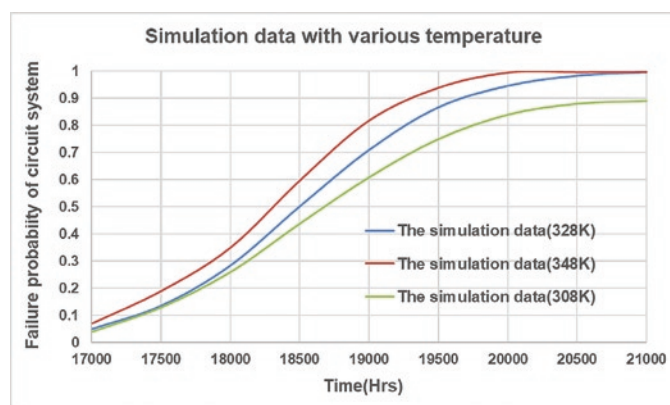


Fig. 13. Sensitivity of temperature for components

## 6. Conclusion

This paper proposed a graphical modelled aggregation approach to estimate the reliability and predict the lifetime of two-channel LED drivers. With the basis of the PoF of all the components in the circuit system, the simulation curves of the lifetimes for components could be obtained. Bridging the gap between the component level and circuit system level, the circuit was mapped into a reliability model accord-

ing to the operation process of the circuit and based on the modelling rules. Integrated with the UGF, the reliability and lifetime prediction of LED drivers can be obtained using a formalized algorithm. This study offers the following contributions: (i) the multiphase modelling of the circuit and the quantitative calculation are enhanced by the reliability model and the UGF; (ii) the gap is bridged between the component level and the circuit system level; and (iii) a universal numerical and emulation approach is provided for predicting the lifetime of LED drivers. The following conclusions are based on the above algorithm:

(i) The reliability of the circuit gradually degrades as a function of time, and the failure rate increases sharply before 18000 h and then begins to flatten gradually.

(ii) The reliability with independent components will degrade more significantly, at nearly double the reliability of the dependent components at approximately 19500 h.

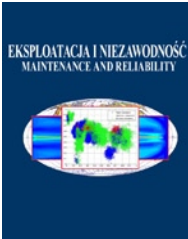
(iii) The capacitors and MOSFETs are identified as vital components of the circuit structure. Component is the most sensitive component.

The approach proposed in this study provides a universal approach, and it is applicable for all types of LED drivers that have the characteristics of multiphase and complex topological structures. However, a few open issues remain, such as the accuracy and precision of the fitting interval and the multiple states of components, which are to be studied in future work. For example, the failure analysis should confirm whether the failure modes used in the prediction approach are the same as the observed failure modes. The interconnection and multi-physical field coupling among various failure mechanisms are the main points of further research. Additionally, the effect of the operation process on practical circuits should be considered, such as the various internal and ambient temperatures, and oscillations on components remain challenging issues in the reliability prediction of LED drivers. Furthermore, the approach should also be extended to fault diagnosis by using the authentic data of the components.

## References

1. Alonso JM, Viña J, Vaquero DG, MartíNez G. Analysis and design of the integrated double buck-boost converter as a high-power-factor driver for power-LED lamps. *IEEE Transactions on Power Electronics* 2012; 59(4): 1689-1697, <https://doi.org/10.1109/TIE.2011.2109342>.
2. Bao ML, Ding Y, Singh C, Shao CZ. A multi-state model for reliability assessment of integrated gas and power systems utilizing universal generating function techniques. *IEEE Transactions on Smart Grid* 2019; 10(6): 6271-6283, <https://doi.org/10.1109/TSG.2019.2900796>.
3. Chen XB, Huang DC, Li Q, Lee FC. Multichannel LED driver with CLL resonant converter. *IEEE Transactions on Power Electronics* 2015; 3(3): 589-598, <https://doi.org/10.1109/JESTPE.2015.2417219>.
4. Chen Y, Nan YR, Kong QG. A loss-adaptive self-oscillating buck converter for LED driving. *IEEE Transactions on Power Electronics* 2012; 27(10): 4321-4329, <https://doi.org/10.1109/TPEL.2012.2190755>.
5. Cirmirakis D, Demosthenous A, Saeidi N, Donaldson N. Humidity-to-frequency sensor in CMOS technology with wireless readout. *IEEE Sensors Journal* 2013; 13(3): 900-908, <https://doi.org/10.1109/JSEN.2012.2217376>.
6. Fan D, Wang Z, Liu L, Ren Y. A modified GO-FLOW methodology with common cause failure based on discrete time bayesian network. *Nuclear Engineering & Design* 2016; 305: 476-488, <https://doi.org/10.1016/j.nucengdes.2016.06.010>.
7. Feng Q, Zhao XJ, Fan DM, Cai BP, Liu YQ, Ren Y. Resilience design method based on meta-structure: A case study of offshore wind farm. *Reliability Engineering & System Safety* 2019; 186: 232-244, <https://doi.org/10.1016/j.res.2019.02.024>.
8. Harada K, Katsuki A, Fujiwara M. Use of ESR for deterioration diagnosis of electrolytic capacitor. *IEEE Transactions on Power Electronics* 1993; 8(4): 355-361, <https://doi.org/10.1109/63.261004>.
9. Hwu KI, Jiang WZ. Expandable two-channel LED driver with galvanic isolation and automatic current balance. *IET Power Electronics* 2018; 11(5): 825-833, <https://doi.org/10.1049/iet-pel.2017.0508>.
10. Hwu KI, Jiang WZ. Non-isolated two-channel LED driver with automatic current balance and zero-voltage switching. *IEEE Transactions on Power Electronics* 2016; 31(12): 8359-8370, <https://doi.org/10.1109/TPEL.2016.2515088>.
11. Kasprzyk L. Modelling and analysis of dynamic states of the lead-acid batteries in electric vehicles. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2017; 19 (2): 229-236, <http://dx.doi.org/10.17531/ein.2017.2.10>.
12. Kozłowski E, Mazurkiewicz D, Żabiński T, Prucnal S, Sęp J. Assessment model of cutting tool condition for real-time supervision system. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2019; 21(4): 679-685, <https://doi.org/10.17531/ein.2019.4.18>.
13. Kozłowski E, Mazurkiewicz D, Żabiński T, Prucnal S, Sęp J. Machining sensor data management for operation-level predictive model. *Expert Systems with Applications* 2020; 159: 1-22, <https://doi.org/10.1016/j.eswa.2020.113600>.

14. Lan S, Tan CM. Application of particle filter technique for lifetime determination of a LED driver. *IEEE Transactions on Power Electronics* 2015; 15(2): 163-172, <https://doi.org/10.1109/TDMR.2015.2407410>.
15. Lan S, Tan CM. Degradation model of a linear-mode LED driver and its application in lifetime prediction. *IEEE Transactions Device Mater. Reliability* 2014; 14(3): 904-913, <https://doi.org/10.1109/TDMR.2014.2343253>.
16. Lan S, Tan CM, Wu K. Methodology of reliability enhancement for high power LED driver. *Microelectronics Reliability* 2014; 54(6-7): 1150-1159, <https://doi.org/10.1016/j.microrel.2014.01.027>.
17. Lan S, Tan CM, Wu K. Reliability study of LED driver-a case study of black box testing. *Microelectronics Reliability* 2012; 52(9-10): 1940-1944, <https://doi.org/10.1016/j.microrel.2012.06.023>.
18. Levitin G, Xing L, Dai Y. Minimum mission cost cold-standby sequencing in non-repairable multi-phase systems. *IEEE Transactions on Reliability* 2014; 63(1): 251-258, <https://doi.org/10.1109/TR.2014.2299192>.
19. Liu HW, Guo K, Zhang ZY. High-power LED photoelectrothermal analysis based on backpropagation artificial neural networks. *IEEE Transactions on Electron Devices* 2017; 65(7): 2867-2873, <https://doi.org/10.1109/TED.2017.2701346>.
20. Liu D, Wang SP, Tomovic MM. Degradation modeling method for rotary lip seal based on failure mechanism analysis and stochastic process. *Eksploracja i Niezawodność - Maintenance and Reliability* 2020; 22(3):381-390, <https://doi.org/10.17531/ein.2020.3.1>.
21. Lin RL, Liu SY, Lee CC, Chang YC. Taylor-series-expression-based equivalent circuit models of LED for analysis of LED driver system. *IEEE Transactions on Industry Applications* 2013; 49(4): 1854-1862, <https://doi.org/10.1109/TIA.2013.2258313>.
22. Li SN, Tan SC, Lee CK, Waffenschmidt E, Hui R, Tse CK. A survey, classification, and critical review of light-emitting diode drivers. *IEEE Transactions on Power Electronics* 2016; 31(2): 1503-1516, <https://doi.org/10.1109/TPEL.2015.2417563>.
23. Li ZF, Wang ZL, Ren Y, Yang DZ, Lv X. A novel reliability estimation method of multi-state based on structure learning algorithm. *Eksploracja i Niezawodność - Maintenance and Reliability* 2020; 22(1): 170-178, <https://doi.org/10.17531/ein.2020.1.20>.
24. Marioli D, Sardini E, Serpelloni M. An inductive telemetric measurement system for humidity sensing. *Measurement Science & Technology* 2008; 19(11): 1054-1057, <https://doi.org/10.1088/0957-0233/19/11/115204>.
25. Meneghini M, Lago MD, Trivellin N, Meneghesso G, Zanoni E. Degradation mechanisms of high-power LEDs for lighting applications: an overview. *IEEE Transactions on Industry Applications* 2014; 50(1): 78-85, <https://doi.org/10.1109/TIA.2013.2268049>.
26. Ren Y, Fan DM, Feng Q, Wang ZL, Sun B, Yang DZ. Agent-based restoration approach for reliability with load balancing on smart grids. *Applied Energy* 2019; 249: 46-57, <https://doi.org/10.1016/j.apenergy.2019.04.119>.
27. Ren Y, Fan DM, Ma XR, Wang ZL, Feng Q, Yang DZ. A GO-FLOW and Dynamic Bayesian Network Combination Approach for Reliability Evaluation with Uncertainty: A Case Study on a Nuclear Power Plant. *IEEE Access* 2017; 6: 7177-7189, <https://doi.org/10.1109/ACCESS.2017.2775743>.
28. Ren Y, Fan DM, Wang ZL, Yang DZ, Feng Q, Sun B, Liu LL. System Dynamic Behavior Modeling Based on Extended GO Methodology. *IEEE Access* 2018; 6: 22513-22523, <https://doi.org/10.1109/ACCESS.2018.2816165>.
29. Sun B, Fan XJ, Li L, Ye HY, Van DW, Zhang GQ. A reliability prediction for integrated LED lamp with electrolytic capacitor-free driver. *IEEE Transactions on Components Packaging and Manufacturing Technology* 2017; 7(7): 1081-1088, <https://doi.org/10.1109/TCPMT.2017.2698468>.
30. Sun B, Fan X, Qian X, Zhang G. PoF-simulation-assisted reliability prediction for electrolytic capacitor in LED drivers. *IEEE Transactions on Industrial Electronics* 2016; 63(11): 6726-6735, <https://doi.org/10.1109/TIE.2016.2581156>.
31. Sun B, Fan XJ, Van DM, Cui CQ, Zhang GQ. A stochastic process based reliability prediction method for LED driver. *Reliability Engineering & System Safety* 2018; 178: 140-146, <https://doi.org/10.1016/j.res.2018.06.001>.
32. Sun B, Jiang XP, Yung KC, Fan JJ, Pecht MG. A review of prognostics techniques for high-power white LEDs. *IEEE Transactions on Power Electronics* 2017; 32(8): 6338-6362, <https://doi.org/10.1109/TPEL.2016.2618422>.
33. U.S. Department of Energy (DOE). LED luminaire lifetime: Recommendations for testing and reporting," 3rd ed. 2014. [Online]. Available: <http://energy.gov/eere/ssl/led-lighting-facts>
34. Ushakov IA. Universal generating function. *Soviet Journal of Computer and Systems Sciences* 1986; 24(5):118-129.
35. Więclawski K, Mącza J, Szezurowski K. Electric current as a source of information about control parameters of indirect injection fuel injector. *Eksploracja i Niezawodność - Maintenance and Reliability* 2020; 22(3): 449-454, <https://doi.org/10.17531/ein.2020.3.7>.
36. Williams RK, Darwish MN, Blanchard RA, Siemieniec R, Rutter P, Kawaguchi Y. The trench power MOSFET-Part II: application specific VDMOS, LDMOS, packaging, and reliability. *IEEE Transactions on Electron Devices* 2017; 63(3): 692-712, <https://doi.org/10.1109/TED.2017.2655149>.
37. Wu H, Wong SC, Tse CK, Chen QH. A PFC single-coupled-inductor multiple-output LED driver without electrolytic capacitor. *IEEE Transactions on Power Electronics* 2019; 34(2): 1709-1725, <https://doi.org/10.1109/TPEL.2018.2829203>.
38. Xia Q, Yang DZ, Wang ZL, Ren Y, Sun B, Feng Q, Qian C. Multiphysical modeling for life analysis of lithium-ion battery pack in electric vehicles. *Renewable & Sustainable Energy Reviews* 2020; 131, <https://doi.org/10.1016/j.rser.2020.109993>.
39. Yadlapalli RT, Narasipuram RP, Kotapati A. An overview of energy efficient solid state LED driver topologies. *International Journal of Energy Research* 2020; 44(2): 612-630, <https://doi.org/10.1002/er.4924>.
40. Zhang H. A viable nontesting method to predict the lifetime of LED drivers. *IEEE Journal of Emerging and Selected Topics in Power Electronics* 2018; 6(3): 1246-1251, <https://doi.org/10.1109/JESTPE.2018.2826364>.
41. Zhang J, Wang J, Wu X. A Capacitor-isolated LED driver with inherent current balance capability. *IEEE Transactions on Power Electronics* 2012; 59(4): 1708-1716, <https://doi.org/10.1109/TIE.2011.2138111>.



Article citation info:

Podulka P. Fast Fourier Transform detection and reduction of high-frequency errors from the results of surface topography profile measurements of honed textures. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 84–93, <http://dx.doi.org/10.17531/ein.2021.1.9>.

## Fast Fourier Transform detection and reduction of high-frequency errors from the results of surface topography profile measurements of honed textures

Przemysław Podulka

Faculty of Mechanical Engineering and Aeronautics, Rzeszow University of Technology, ul. Powstancow Warszawy 8, 35-959 Rzeszów, Poland

Indexed by:



### Highlights

- The approach for detection and reduction of selected measurement errors is presented.
- Power spectral density and autocorrelation function are valuable in noise analysis.
- For minimisation of effect of measurement errors the ‘noise surface’ is defined.
- Profile analysis might be more valuable than areal when measurement noise is assessed.
- Fast Fourier Transform is useful in definition of high-frequency measurement noise.

### Abstract

In this paper, various type of noise detection procedures with surface topography profile analysis were proposed, compared (studied) and suggested. The honed cylinder liner surface textures with additionally burnished oil pockets were measured with a stylus or optical approaches. Measurement errors, defined as high-frequency measurement noise, were taken into sufficient consideration. It was proposed to select the noise detection methods more with profile (2D) than areal (3D) assessments; some-frequency noise was much easier to observe in profile than surface analysis. Moreover, applications of various type of regular filtration methods, mostly based on Gaussian functions, were compared with Fast Fourier Transform filtration for detection or reduction of some (high) frequency-defined measurement errors.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

surface topography, measurement, measurement errors, measurement noise, cylinder liner, oil pocket, dimple, valley, texture features.

## 1. Introduction

Measurement and analysis of surface topography (ST) are highly advantageous in assessments of tribological performance of ‘engineering surfaces’ [17], especially wear resistant [6], lubricant retention, sealing, friction, material contact in general [25]. ST is created in the last stage of the machining process; besides, medium-scale components of ST, waviness, in particular, arise as a result of imperfection in the manufacturing process [21].

Detailed information about ST can be obtained by measurements and analysis of the received raw data. Results in the assessment of wear behaviour of selected surface textures are fraught with many factors that can affect the accuracy of the research carried out. In general, errors in the analysis of ST, dedicated mainly for tribological elements, can be classified as errors in measuring equipment and environment [2], the measured object errors, software [30] or measuring method errors [37]. While considering the measurement uncertainty, it can be grouped in errors: typical for measuring methods [24], those caused by digitisation process [23], another received when data processing is accomplished [26] and other errors [20, 29, 44].

There can arise a few problems in the processing of raw measured data. Usually, the ST parameters of car engine parts are calculated after form (errors of shape and waviness) removal [32]. Errors of this type of data processing can be especially visible when edge-area of

two-process textures are considered [28]. Improper selection of form removal procedure can cause classification of properly made parts as a lack and their rejections [31].

ST measurement equipment is often roughly divided into the stylus and non-contact methods [40]. In defiance of significant development of optical techniques, a stylus profilometer is still the most common roughness measuring device in the mechanical industry [22]. Contrary to stylus (contact, in general) methods which are robust but slow optical methods are fast but deeply sensitive to external effects. In general, optical methods require the isolation of instruments from the external environment, which is a complicated task to be accomplished. Therefore, the results of measurement of a rough surface can be fraught by errors caused by the noise occurrence. The noise has different sources, including those generated internally, and those defined as external sources received from the environment.

There are many types of noise in surface metrology, e.g. scattering [43], background [14], measurement/instrument [4], outlier [41], static [9] and other noise-like [1] errors. The measurement noise can be briefly described as the noise added to the output signal [12] occurring when the normal use of measuring instrument is accomplished. The effect of noise occurrence was minimised with various methods, e.g. correlogram correlation [16], by proposals of limitation and matching of bandwidth for various instruments [3], some optimisation approaches for selected (coherence scanning interferometry) measuring

E-mail addresses: P. Podulka - [p.podulka@prz.edu.pl](mailto:p.podulka@prz.edu.pl)

methods [7], the frequency de-noising technique based on the wavelet decompositions [5], Fourier reduction or random phase exclusion schemes [8], and other plenty methods [27].

Widely used in surface texture characterisation is the Fast Fourier Transform (FFT) [18]. The FFT technique is likely to play an important role in the analysis of dry contact of the rough surface [35], contact assessments in general [19]. For example, the power values of power spectral density (PSD) for the atomic force microscopy (AFM) digital data were determined by the FFT algorithms instead of the root-mean-square (rms) and peak-to-valley value when a surface morphology of pentacene thin films and their substrates were considered [15]. Moreover, peak detecting algorithm that combines the white-light phase-shifting interferometry (WLPSI) method and FFT coherence-peak-sensing technique was proposed, which could accurately determine the local fringe peak and improve the vertical resolution of the measurement [39]. Application of this type of algorithm could effectively reduce the batwing effects at the edges and solve the problem of positioning error in the maximum modulation.

FFT can also be associated with other filters, e.g. Gaussian [38], for measurement of surface roughness of optical thin films; the surface profile was obtained by the fringe pattern analysis using the FFT method. Besides, the asperity radius of curvature and asperity density, which are generally derived from a rough surface simulation with FFT, are the two essential parameters for statistical contact model [11]. The simulated rough surface with desired parameters was widely used as an input for the numerical simulation of tribological behaviour such as the asperity contact, lubrication, and wear. Gaussian rough surfaces with desired autocorrelation function (ACF) and spatial statistical parameters, including skewness ( $Ssk$ ) and kurtosis ( $Sku$ ), was developed by combining the FFT, translation process theory, and Johnson translator system [42]. An FFT algorithm can also reduce computational time in the calculation of universal spline filter. Furthermore, for practical use, an improved method prior to the FFT algorithm can be proposed to suppress the end effect [45].

In some cases, the FFT was considered to be one of the essential tools of digital signal processing for surface profiles filtration and was compared with the existing methods [33]. Moreover, a fast white-light interference method for measuring surface depth profiles at nanometer scales was also provided with FFT appliances. White-light profilers have relied either on path difference scanning or on spectral analysis of the reflection from a fixed interferometer. It was shown that, by performing this spectral analysis with an imaging Fourier transform spectrometer, the high speed of spectral techniques might be successfully combined with the simple data interpretation characteristic of a used scanning method [10]. A two-dimensional FFT technique was also proposed for accelerating the computation of scattering characteristics of features on surfaces by using the discrete-dipole approximation; FFT reduced the execution time of data processing as well [36].

In this paper, the FFT filtration method was proposed for detection and reduction of high-frequency measurement errors. It was suggested to analyse the ST with the profile (2D) instead of areal (3D) assessments. A number of profiling techniques have been developed for measuring the topography of rough surfaces [34]. Nonetheless, there is still a grave problem with the selection of procedure for detection of measurement noise. The influence of noise-separation methods on areal and profile ST analysis was not comprehensively studied in previous researches as well.

## 2. Analysed materials, measurement process and applied methods

In this paper, the honed cylinder liner texture with dimples created by the burnishing techniques was taken into account. They were measured by the stylus instrument, Talyscan 150. The nominal tip radius was about 2  $\mu\text{m}$ , height resolution 10 nm and the measured area

5 mm by 5 mm (1000 x 1000 measured points), sampling interval was 5  $\mu\text{m}$ .

Detection of noise, especially with high-frequency, from the results of raw measured data, was performed with FFT application. The FFT approach was provided for areal (3D) details but in many cases the results were presented for profiles (2D) only. It was caused by the more direct (visible) results in profile detection and reduction of the high-frequency measurement errors by analysis of the PSD and AFC graphs. For each type of surface texture (whereas topography contains the deep/wide dimple or not), the noise-detection procedure was modified to provide the best fits.

The effect of ST feature distribution on noise detection and reduction was carefully studied. Results were compared with those received by regular filtering methods, e.g. Gaussian (GF), robust Gaussian (RGF) or Spline filters (SF). Furthermore, analysis of the power spectral density and the autocorrelation function graphs were accomplished. The PSD was defined with no other functions, e.g. Zoom Factor equal to 1, Loops (number of iterations) also, the smoothing or windows functions were not provided as well. PSDs were received by the support of a commercial software. The noise removal results were also defined as the results of S-operator filtering, described in details previously in surface measurement (ISO) standards.

## 3. Results and discussions

### 3.1. Areal and profile analysis of measurement noise

Considering the measurement errors, especially the high-frequency measurement noise (HFMN), it is necessary to describe it along with the associated measurement bandwidth. When HFMN was scrutinised, the PSD of measured detail has been carefully examined, explicitly taking into account the existence of the high-frequencies (frequencies of small scales) on the analysed graph. Moreover, visual detection of HFMN of both contour map plots and PSD graphs from the results of the measurement process was dependent on the width and depth of the ST features.

It was assumed that the size of dimples (valleys) additionally burnished on the honed cylinder liner texture has a significant impact on the noise detection process. In Figure 1, contour map plot of cylinder liner, measured with various velocity (a – 0.2 mm/s and b – 1 mm/s), was presented. It was found that if the width (depth) of the considered feature was greater than 0.1 mm (0.010 mm) the HFMN detection was exceedingly difficult or it was infeasible with PSD graph analysis, and there were no required (high) noise-frequencies on the diagrams. The occurrence of deep/wide features on the analysed detail made the visual (eye-view) noise detection enormously complicated.

It was also proposed to compare optical disclosure of HFMN for detail before (raw measured data) and after form removal. ‘Flat’ surface analysis might give more convincing results than details with higher amplitude (containing form and waviness) when measurement noise was detected with visual PSD plot assessment. The form was eliminated with the application of least-square fitted polynomial plane of 3<sup>rd</sup> degree enhanced by the valley excluding method. This form separation approach was investigated and substantially confirmed by the previous studies, the usefulness of the applied scheme for characterisation of two-process cylinder liner topographies containing (relatively) large dimples was noticeably improved.

When the mentioned PSD diagrams were visually compared (c and d example against the a and b, from Figure 1), there was no discrepancy (or they were negligible) between the results obtained. Consequently, it was reasonably assumed that detection of HFMN by the analysis of the PSD graph of honed cylinder liner topographies including oil reservoirs, was not reasonably precise regardless of areal form removal process was completed or not. Notwithstanding the previous, the effect of form removal on noise identification (and vice

versa) was not thoroughly scrutinised; nevertheless, this was not one of the underlying assumptions of the conducted research.

Similar problems in optical detecting of the noise from the results of surface texture measurement were obtained with profile (2D) analysis. In Figure 2, profiles (and respectively their PSD diagrams) extracted from honed cylinder liner topography, measured with different speed, were presented. Differences in measurement results were mainly (easy) noticeable with visual analysis of profile plots. Nevertheless, the variations on the PSD graphs were trifling or did not exist. In the considered profile, the valley depth and width were around 30  $\mu\text{m}$  and 0.5 mm, respectively.

It was also assumed that when the number (density) of valleys in 2D (3D) analysis increased, the PSD-detection of HFVN became

utterly impracticable. Subsequently, the process of noise disclosure, based on the PSD diagram assessment, appeared to be not precisely stated when honed cylinder liner topographies were studied with valley assays.

The effect of noise amplitude, not analysed in these studies, was also extraordinarily difficult to determine. In further research, where defining of error amplitude would be the primary task, the feature (e.g. valley, dimple) size (and density) and its distribution (considering as well the distance between each of them) might be of great importance. Moreover, the noise identification with PSD assessment was even more complicated that the noise was not visible for each measurement condition (velocity). Additionally, HFVN was discernible with visual profile analysis in contrast to the PSD graph assays for each surface texture measurement velocity.

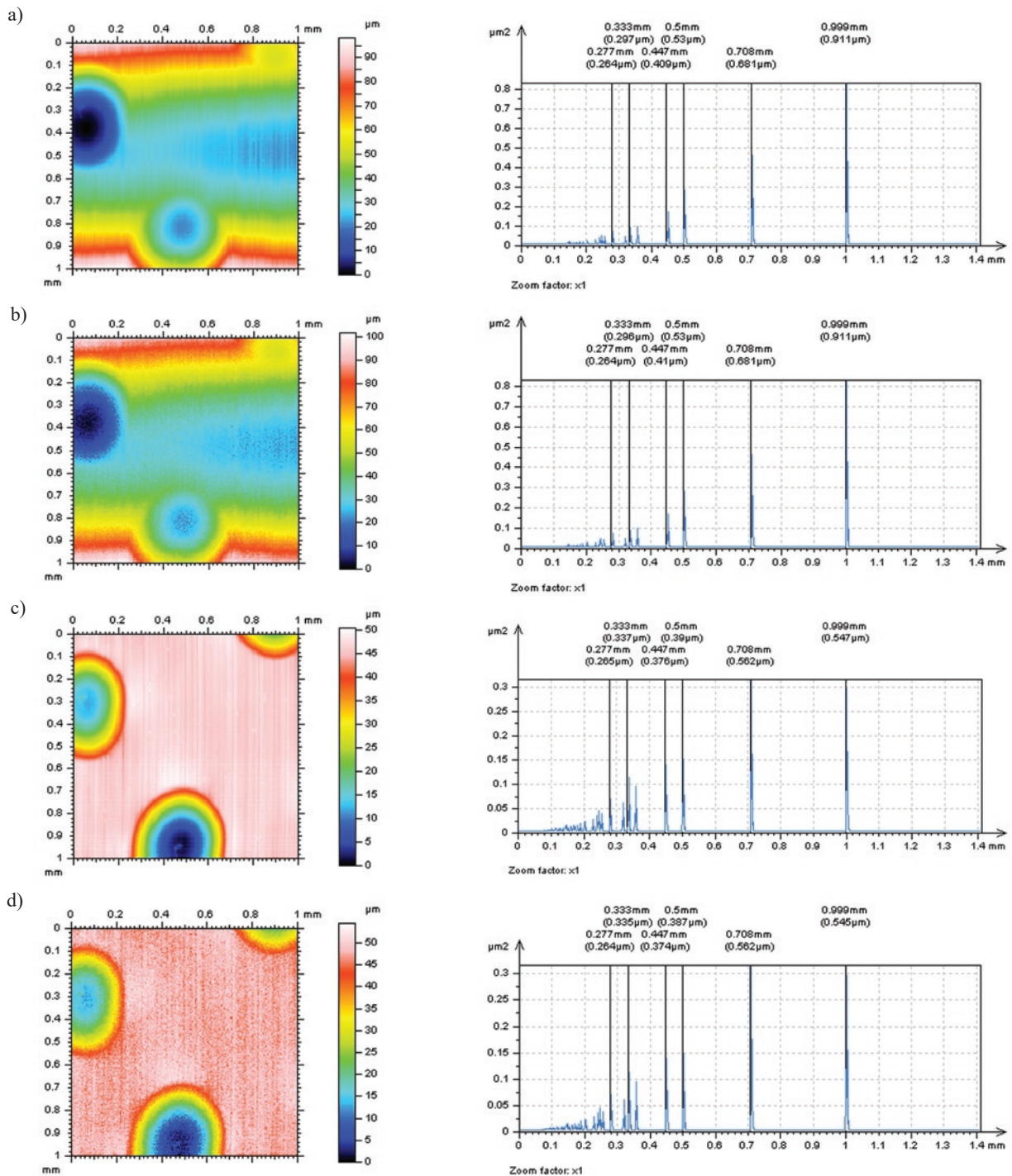


Fig. 1. Contour map plots and PSDs of cylinder liner texture containing wide dimples, measured by stylus instrument with a velocity equal to 0.2 mm/s (a, c) and 1 mm/s (b, d); raw measured data after levelling process (a, b) and texture after form removal by a polynomial of 3rd degree with valley excluding method approach (c, d)



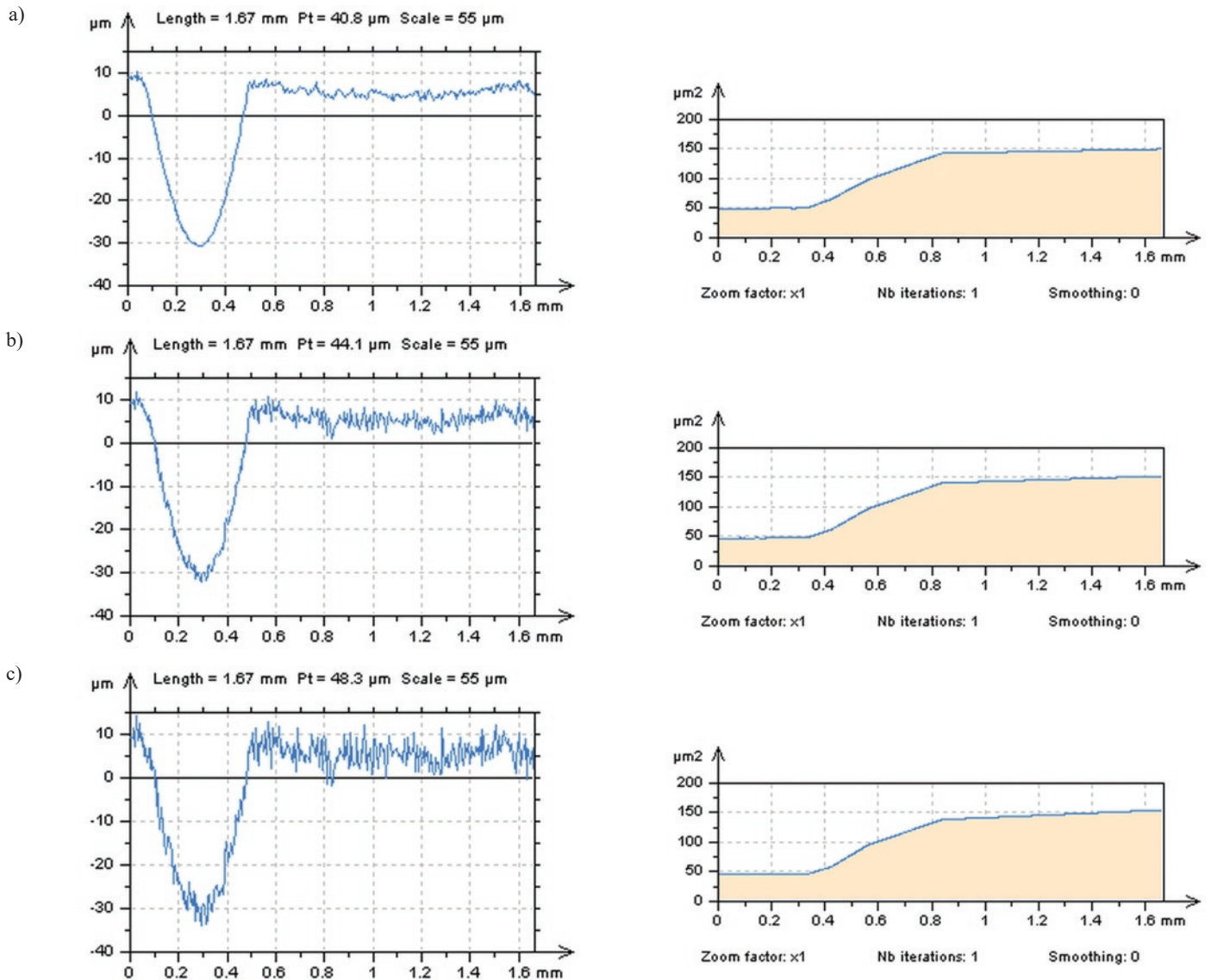


Fig. 2. Profiles and their PSDs respectively, extracted from cylinder liner texture containing one wide dimple, received for stylus measurement with speed equal to (a) 0.2 mm/s, (b) 0.5 mm/s and (c) 1 mm/s

### 3.2. The effect of feature (dimple) size on the detection of high-frequency measurement errors

The impact of feature (dimples, valley) size (width, depth) on detection of HFMN from the results of raw measured data was taken into consideration; the number (density) of valleys was not studied. It was assumed that for honed cylinder line textures containing dimples with a depth greater than 10  $\mu\text{m}$ , the detection of HFMN with PSD analysis was severely impeded. The smallest valley depth was found the more accurate PSD-identification of measurement noise was observed. In Figure 3, three different profiles (containing dimples with various depth) extracted from honed cylinder liner topography are introduced. As it was assumed in the previous section, in condition when the valley depth was greater than 10  $\mu\text{m}$ , the detection of HFMN with PSD analysis became unattainable (a, b). However, when the dimple depth was smaller than the previously specified value, the noise (high) frequencies were found in the PSD diagram (c). This valley-depth value should be identified for  $Sq$  of both the surface and the noise (HFMN specifically). Despite that, the arbitrarily selected values should not be decisive; it is suggested to identify the noise occurs in the measurement results by PSD analysis with out-of-deep-feature (out-of-deep-valley/out-of-deep-dimple to be more precise) characteristics. This finding can be entirely appropriate if the noise amplitude is relatively

small, concerning the surface height (for both methods – out-of-dimple scheme and approach based on excluding the valleys).

Not only the depth of the surface features affects the PSD-accuracy of noise detection. It was found that if the depth of valleys was relatively small (usually smaller than  $Sq/Rq$  value of non-dimple detail/profile), the HFMN was also challenging to be discerned. The direct impact of dimple width can be clearly designated by the consideration of feature width (A) against the width of out-of-feature detail/profile (B); the description was appropriately introduced in Figure 4. Each profile was shortened to reduce the A/B coefficient. The size of out-of-feature (dimple) profile detail was equal to 0.85 mm, and correspondingly, the width of the valley was decreased from 0.15 mm (a) to 0.05 mm (c), or its value was negligible (d). Consequently, when the value of A/B factor decreased, the PSD-identification of HFMN occurrence increased. In general, when the width of the feature (A) was smaller than 10% of profile width (A + B), then the noise PSD-disclosure increased; visual detection of HFMN was facilitated.

### 3.3. Reduction of noise from the results of surface topography measurements

Once the procedure of (high-frequency) noise detection is applied, the selection of approach for reduction (removal) of measurement errors should be chosen very attentively. The final selection of noise

reduction procedure has basically the considerable influence on the calculation (values) of areal surface topography parameters. Properly accomplished measurement process might not provide desired (adequate) results when the processing of received raw measured data is not performed competently. The method of reduction of HFMN is, in addition to the form removal procedure, one of the main factors that can cause a false estimation of detail properties and its classification as a lack and rejection.

For reduction of the effect of measurement errors on values of areal surface texture parameters, the FFT filtering was compared with commonly-used (mostly available in commercial software of measuring equipment) algorithms. Several approaches for noise reduction issue were proposed separately. Basically, the HFMN can be defined by analysis of the obtained results with the S-filtering process [13], e.g. regular Gaussian, spline or FFT methods. In the previous research, it was found that the procedure of reduction of HFMN can be selected by assessment of received data from the filtering approach. The results of S-filtering process can be defined as a noise surface (NS) or noise profile (NP), in this specific instance, when the 3D or, cor-

respondingly, 2D measured data were considered. It was also found that NS (NP) received by the S-operation should consist of only the high-frequencies or those (high) frequencies should be described as an overwhelmingly 'dominant' frequencies. Simultaneously, the frequencies that are required to be removed should be those dominant in the results of noise reduction (filtering as a considered case) procedure. In general, the more dominant frequencies are those with high-frequency spectrum the better results of HFMN reduction process are received.

From the analysis of PSDs of profiles from honed surface texture after reduction of HFMN, it was assumed that the most dominating high-frequency was found for the results obtained after S-filtering by the FFT method. In Figure 5, profiles and their PSDs were presented for profiles after various type of filtering approaches. From Gaussian schemes (GF or RGF), robust filtering caused a greater amplitude of received NP than the regular filter. When the FFT approach was applied, the high-frequencies were those 'dominant'; the amplitude of those frequencies (on PSD graph) was greater than for profiles after noise reduction by other methods (GF, RGF and SF). The amplitude

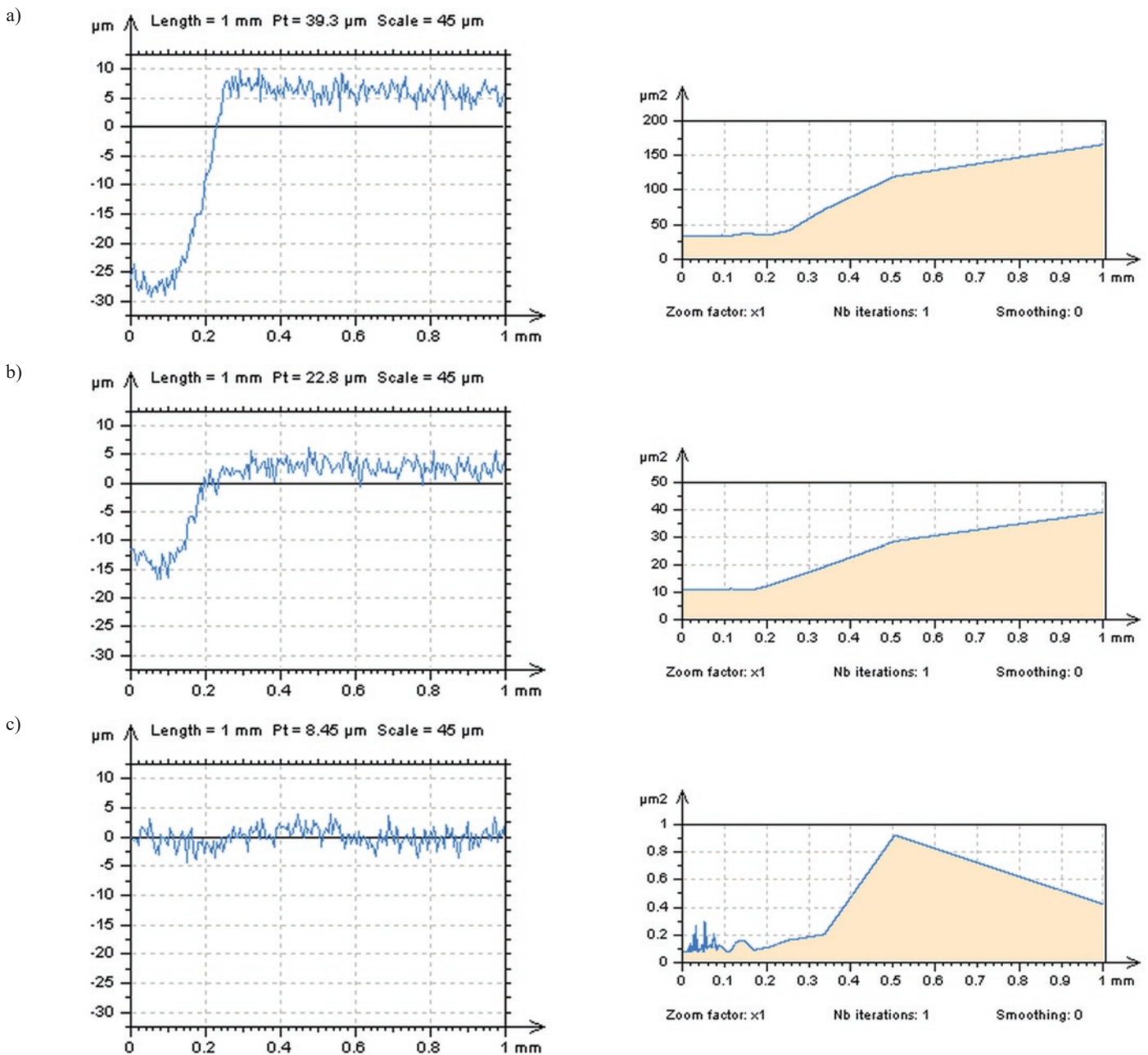


Fig. 3. Profiles and their PSD graphs correspondingly, extracted from plateau-honed cylinder liner texture measured by stylus method with a velocity equal to 0.6 mm/s; profile with deep (depth equal to 35 μm approximately) valley (a), profile with a relatively not-deep (20 μm) valley (b) and non-dimple profile (c)

(maximum height) of received NP was the greatest when RGF was used. Nonetheless, the amplitude of HFMN (NS) was not considered in this paper.

Some features located on the ACF graph (Figure 6) occurred in periodic sentence when commonly available (Gaussian or spline)

filters were applied. Received frequencies were more 'periodic' than those after noise reduction by the FFT approach. Moreover, the peak density of ACF also increased when FFT filtering was accomplished instead of generally used filters. The process of noise reduction in profile analysis can be found by definition and assessment of dominant frequency. Additionally, when HFMN was considered to be removed

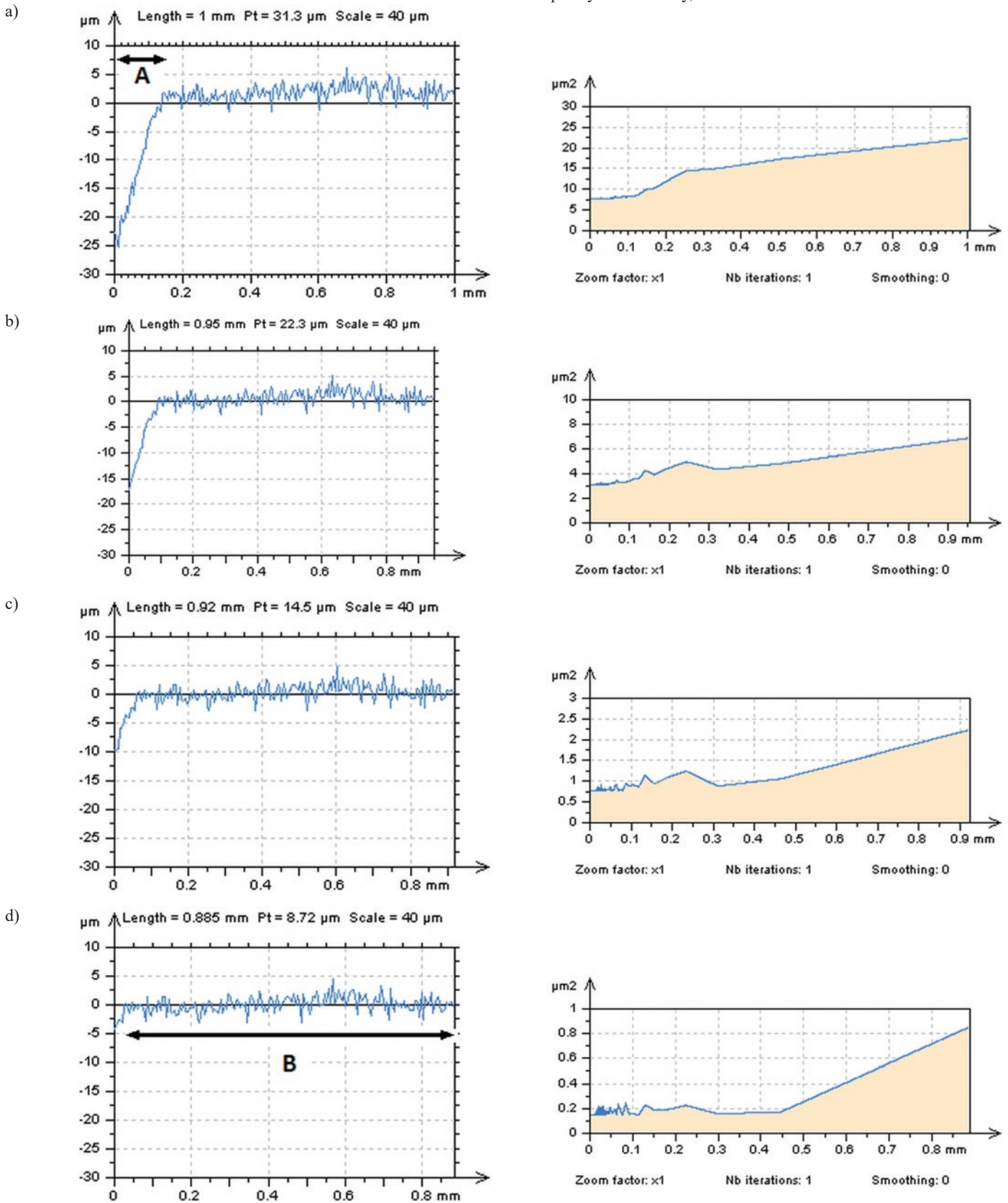


Fig. 4. Profiles, and their PSD graphs correspondingly, extracted from plateau-honed cylinder liner texture measured by stylus method with a velocity equal to 0.7 mm/s; profile with A-size (width) of dimple equal to 0.15 mm (a), 0.1 mm (b), 0.05 mm (c), and non-dimple profile (d)

(minimised), the dominant frequency (high- in particular) should not be found in the received post-processing (after S-operator filtering) data. The smallest amplitude of dominant (removed) frequency was established in the received de-noised data the highly advantageous of the applied algorithm was clearly indicated.

HFMN was removed by application of various filter with a bandwidth equal to 0.015 mm, 0.025 mm and 0.035 mm to improve the

utility of FFT approach; nevertheless, studies can be extended to the other values depending on the type of surface texture analysed. For all of the considered cut-off values, the amplitude of dominant (high-) frequency was maximised (minimised) according to the regular filter, when the PSDs of received (NP) results were observed (Figure 7).

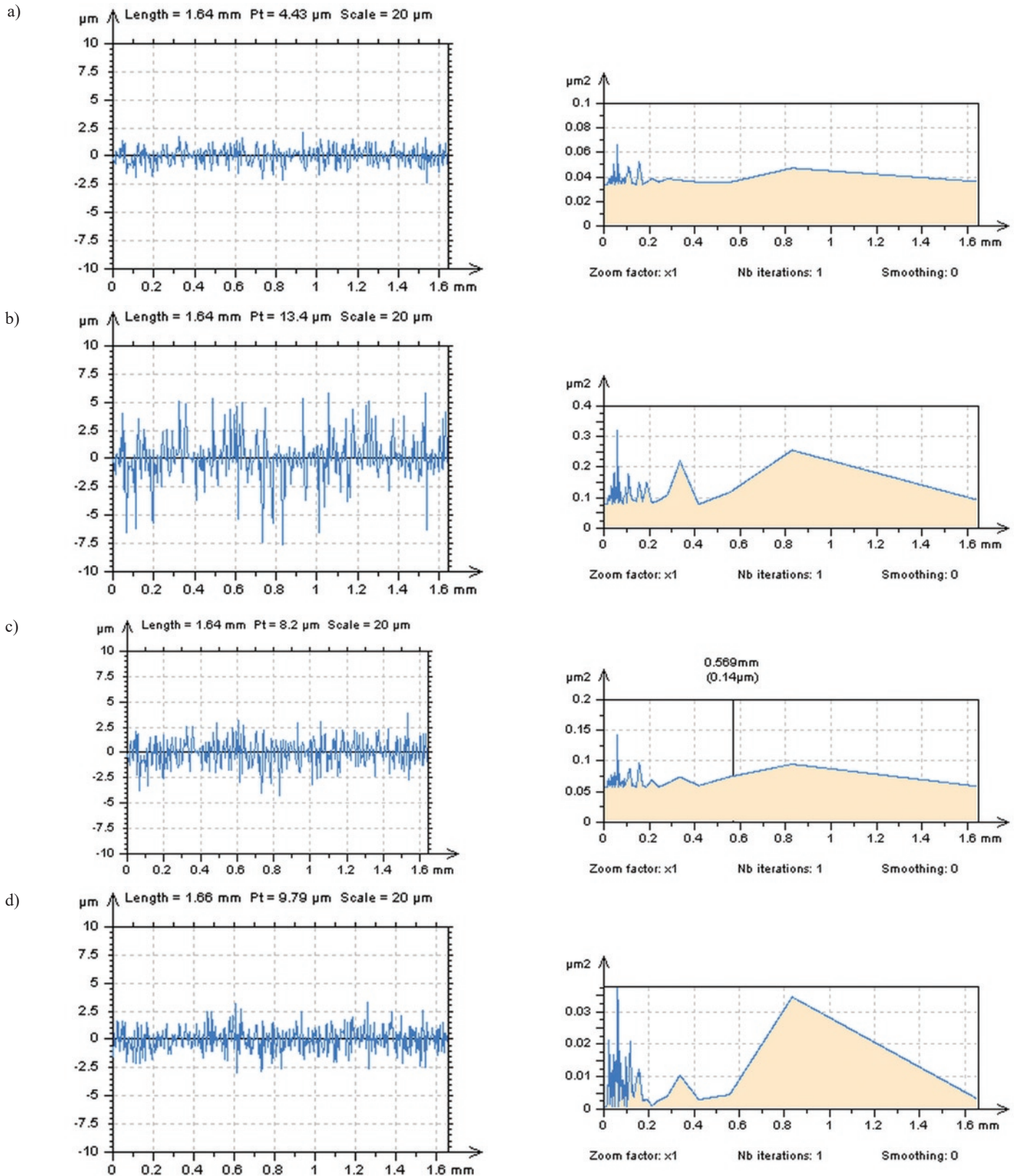


Fig. 5. NPs and its PSDs correspondingly, measured with speed equal to 0.8 mm/s from plateau-honed cylinder liner texture, and after reduction of HFMN by GF (a), RGF (b), SF (c) and FFT (d), cut-off = 0.015 mm

#### 4. Conclusions

The detection and then reduction of high-frequency measurement noise is a highly demanding task that makes the analysis of surface topography reasonably intuitive. Nonetheless, some applications might be accepted:

1. For honed cylinder liner textures (plateau-honed in particular) containing additionally burnished oil-reservoirs (dimple, valley, wholes in general), the high-frequency measurement noise was exceedingly difficult to detect for areal analysis with power spectral or autocorrelation function applications.

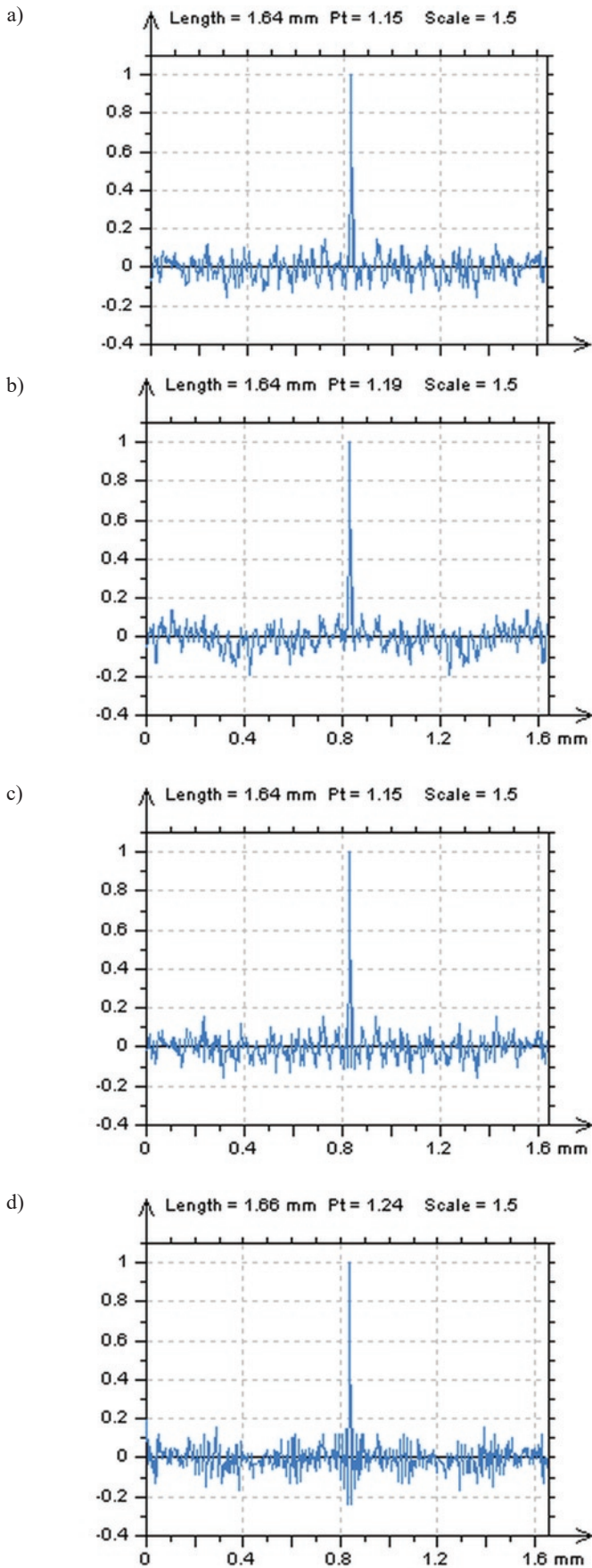


Fig. 6. ACFs of NP, received from measured profile (speed 0.8 mm/s) of honed surface texture obtained after removal of HF MN by GF (a), RGF (b), SF (c) and FFT (d) method with cut-off = 0.015 mm

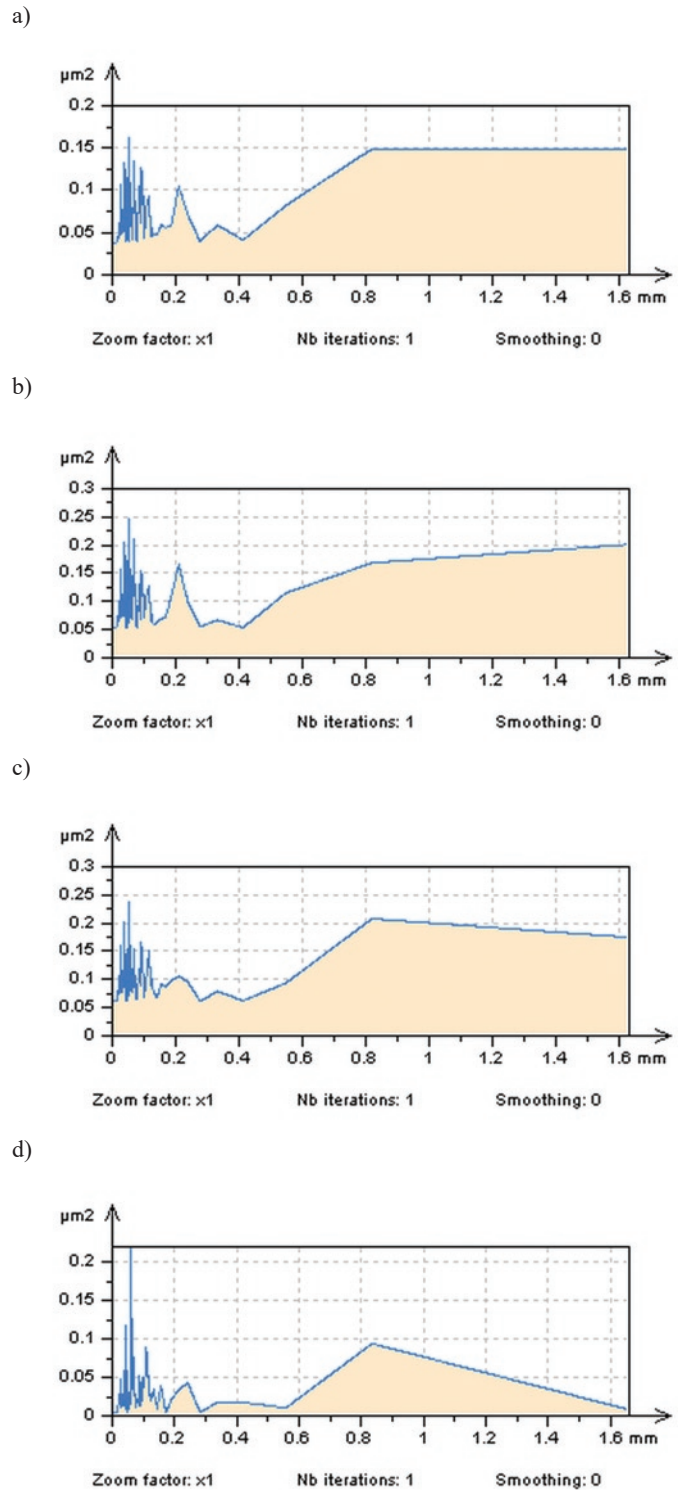


Fig. 7. PSDs of measured profile (speed 0.8 mm/s) from honed surface texture obtained after removal of HF MN by GF (a), RGF (b), SF (c) and FFT (d) method with cut-off = 0.035 mm

The eye view assessment of the occurrence of high-frequency errors in the results of surface topography measurements was increasingly complicated when texture contains some deep features.

2. The form (e.g. cylindrical) of analysed detail has no impact on the process of 'power spectral density' (PSD) detection of high-frequency measurement noise. The differences in PSD graphs, calculated for both raw measured data and data after application of the process of areal form removal (e.g. least-square fitted polynomial plane of 3<sup>rd</sup> order improved by valley excluding method approach), were imperceptible.
3. It is suggested to identify the noise in the results of surface texture measurement by power spectral density analysis of out-of-feature (deep valleys in particular) details. This might be entirely justified when the noise amplitude is relatively small, according to the surface amplitude (height). Moreover, when the number (density) of valleys in profile (areal) analysis increased, the detection of high-frequency measurement noise, with the analysis of power spectral density graphs, appeared to be infeasible.
4. The feature (dimple) size has a considerable influence on the accuracy of the noise detection procedure. As it was assumed, when the depth of valley was greater than 10  $\mu\text{m}$  (usually when  $Sq$  of the non-dimple area was smaller than 10% of the  $Sq$  value calculated for all measured detail – containing dimples), the detection of high-frequency measurement errors (with power spectral density analysis certainly) became unobtainable. In addition, when the width of the feature was also

smaller than 10% of the profile width, then the precision of noise disclosure increased.

5. It was suggested to determine the high-frequency measurement noise with noise profile analysis, as a results of S-filtering methods. The properly obtained noise profile should contain only the 'unwanted' (with high-frequency spectrum) components. It was suggested that the removed frequency should be this dominant in the received results after the S-filtering of noise profile.
6. When the received noise profile (surface) were considered with autocorrelation function, it was found that properly defined noise profile (surface) should not contain any periodic waveforms; the noise profile (surface) should be non-periodic (isotropic), in general.
7. For reduction of high-frequency noise from the results of measurement of surface topography, the fast Fourier transform (FFT) was proposed. When compared with commonly-used procedures (e.g. various Gaussian or spline filters) with various cut-off values (e.g. 0.015 mm, 0.025 mm, 0.035 mm), the amplitude of dominant (required high-) frequency was maximised.
8. It is recommended to select the noise reduction procedure with multivariate analysis, combining the profile (areal) eye view (of contour map/profile plots) analysis with assessments of the graphs of power spectral densities and autocorrelation functions.

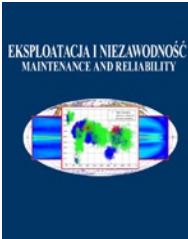
#### Acknowledgements

The author kindly acknowledges the Polish National Science Centre for financial support, project number: 2013/09/N/ST8/04333.

#### References

1. Creath K, Wayant JC. Absolute measurement of surface roughness. *Applied Optics* 1990; 29(26): 3823-3827, <https://doi.org/10.1364/AO.29.003823>
2. Dagnall H. *Exploring Surface Texture*. Rank Taylor Hobson Limited, Leicester, UK, 1986.
3. De Groot P, DiSciaccia J. Surface-height measurement noise in interference microscopy. In: *Proceedings of SPIE 10749, Interferometry XIX*, 2018, 107490Q-7, <https://doi.org/10.1117/12.2323900>
4. De Groot P. The meaning and measure of vertical resolution in optical surface topography measurement. *Applied Sciences* 2017; 7(1): 54, <https://doi.org/10.3390/app7010054>
5. Dong W, Ding H. Full frequency de-noising method based on wavelet decomposition and noise-type detection. *Neurocomputing* 2016; 214: 902-909, <https://doi.org/10.1016/j.neucom.2016.06.072>
6. Dzierwa A, Galda L, Tupaj M, Dudek K. Investigation of wear resistance of selected materials after slide burnishing process. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2020; 22(3): 432-439, <https://doi.org/10.17531/ein.2020.3.5>
7. Gomez C, Su R, Thompson A, DiSciaccia J, Lawes S, Leach R K. Optimisation of surface measurement for metal additive manufacturing using coherence scanning interferometry. *Optical Engineering* 2017; 56(11): 111714, <https://doi.org/10.1117/1.OE.56.11.111714>
8. Haitjema H, Morel M A A. Method for approximate noise elimination in form and roughness measurements. In: *Proceedings of SPIE 2003*; 5190: 203-210.
9. Haitjema H. Uncertainty in measurement of surface topography. *Surface Topography: Metrology and Properties* 2015; 3(3): 035004, <https://doi.org/10.1088/2051-672X/3/3/035004>
10. Hart M, Vass D G, Begbie M L. Fast surface profiling by spectral analysis of white-light interferograms with Fourier transform spectroscopy. *Applied Optics* 1998; 37(10): 1764-1769, <https://doi.org/10.1364/AO.37.001764>
11. He Y F, Tang J Y, Zhou W, Liao D R. Research on the obtainment of topography parameters by rough surface simulation with fast Fourier transform. *Journal of Tribology-Transactions of the ASME* 2015; 137(3): 031401, <https://doi.org/10.1115/1.4029843>
12. ISO 2016 25178-600 Geometrical product specification (GPS) - Surface texture: Areal Part 600: Metrological characteristics for areal-topography measuring methods.
13. ISO 25178-3:2012 Geometrical product specifications (GPS) – Surface texture: Areal – Part 3: Specification operators.
14. ISO 25178-605 Geometrical product specification (GPS) – Surface Texture: Areal – Part 605: Nominal characteristics of non-contact (point autofocus probe) instruments, First Edition 2004.
15. Itoh T, Yamauchi N. Surface morphology characterisation of pentacene thin film and its substrate with under-layers by power spectral density using fast Fourier transform algorithms. *Applied Surface Science* 2007; 253(14): 6196-6202, <https://doi.org/10.1016/j.apsusc.2007.01.056>
16. Kiselev I, Kiselev E I, Drexel M, Hauptmann M. Noise robustness of interferometric surface topography evaluation methods. Correlogram correlation. *Surface Topography: Metrology and Properties* 2017; 5(4): 045008, <https://doi.org/10.1088/2051-672X/aa9459>
17. Kubiak K J, Wilson M C T, Mathia T G, Carval Ph. Wettability versus roughness of engineering surfaces. *Wear* 2011; 271(3-4): 523-528, <https://doi.org/10.1016/j.wear.2010.03.029>

18. Lin C S, Yang S W, Lin H L, Li J W. Measurement of surface profile and surface roughness of fibre-optic interconnect by fast Fourier transform. *Metrology and Measurement Systems* 2017; 24(2): 381-390, <https://doi.org/10.1515/mms-2017-0028>
19. Liu S B, Wang Q. Studying contact stress fields caused by surface tractions with a discrete convolution and fast Fourier Transform algorithm. *Journal of Tribology-Transactions of the ASME* 2002; 24(1): 36-45, <https://doi.org/10.1115/1.1401017>
20. Magdziak M, Wdowik R. Coordinate Measurements of Geometrically Complex Ceramic Parts. *Applied Mechanics and Materials* 2014; 627: 172–176, <https://doi.org/10.4028/www.scientific.net/AMM.627.172>
21. Marinescu I D, Rowe W B, Dimitrov B, Ohmori H. *Tribology of Abrasive Machining Processes*. 2nd Edition. William Andrew Publishing, Norwich, NY, 2013, <https://doi.org/10.1016/C2010-0-67070-2>
22. Pawlus P, Reizer R, Wieczorowski M. Comparison of results of surface texture measurement obtained with stylus methods and optical methods. *Metrology and Measurement Systems* 2018; 25(3): 589–602, <https://doi.org/10.24425/123894>
23. Pawlus P. Digitisation of surface topography measurement results. *Measurement* 2007; 40(6): 672-686, <https://doi.org/10.1016/j.measurement.2006.07.009>
24. Pawlus P. The errors of surface topography measurement using stylus instruments. *Metrology and Measurement Systems* 2002; 9(3): 273-289.
25. Płonka S, Zaborski A. Operational wear of the neck of spindle coating in cooperation with yarn. *Eksploatacja i Niezawodność - Maintenance and Reliability* 2015; 17(4): 496–503, <https://doi.org/10.17531/ein.2015.4.3>
26. Podulka P. Bisquare robust polynomial fitting method for dimple distortion minimisation in surface quality analysis. *Surface and Interface Analysis* 2020; 52(12) 875–881, <https://doi.org/10.1002/sia.6793>
27. Podulka P. Comparisons of envelope morphological filtering methods and various regular algorithms for surface texture analysis. *Metrology and Measurement Systems* 2020; 27(2): 243-263, <https://doi.org/10.24425/mms.2020.132772>
28. Podulka P. Edge-area form removal of two-process surfaces with valley excluding method approach. *MATEC web of conferences* 2019; 252: 05020, <https://doi.org/10.1051/mateconf/201925205020>
29. Podulka P. Proposal of Frequency-Based Decomposition Approach for minimisation of errors in surface texture parameter calculation. *Surface and Interface Analysis* 2020; 52(12) 882–889, <https://doi.org/10.1002/sia.6840>
30. Podulka P. The effect of valley depth on areal form removal in surface topography measurements. *Bulletin of the Polish Academy of Sciences. Technical Sciences* 2019; 67(2): 391-400, <https://doi.org/10.24425/bpas.2019.128597>
31. Podulka P. The effect of valley location in two-process surface topography analysis. *Advances in Science and Technology Research Journal* 2018; 12(4): 97-102, <https://doi.org/10.12913/22998624/100343>
32. Raja J, Muralikrishnan B, Fu S. Recent advances in separation of roughness, waviness and form. *Precision Engineering* 2002; 26(2): 222-235, [https://doi.org/10.1016/S0141-6359\(02\)00103-4](https://doi.org/10.1016/S0141-6359(02)00103-4)
33. Raja J, Radhakrishnan V. Filtering of surface profiles using fast Fourier transform. *International Journal of Machine Tools & Manufacture* 1979; 19: 133–141, [https://doi.org/10.1016/0020-7357\(79\)90003-9](https://doi.org/10.1016/0020-7357(79)90003-9)
34. Rhee H G, Vorburger T V, Fu J, Renegar T B, Song J F. Comparison of roughness measurements obtained with optical and stylus techniques. *Proceedings of the 10th International Conference on Metrology and Properties of Engineering Surfaces*, Saint-Etienne, France, 2005, pp. 39–47.
35. Sainsot P, Lubrecht A A. Efficient solution of the dry contact of rough surfaces: a comparison of fast Fourier transform and multigrad methods. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology* 2011; 225(6): 441-448, <https://doi.org/10.1177/1350650111401535>
36. Schmehl R, Nebeker B M, Hirlman E D. Discrete-dipole approximation for scattering by features on surfaces by means of a two-dimensional fast Fourier transform technique. *Journal Of The Optical Society Of America A-Optics Image Science And Vision* 1997; 14(11): 3026-3036, <https://doi.org/10.1364/JOSAA.14.003026>
37. Thomas T R. *Rough Surfaces*. Second Edition, Imperial College Press, London, 1999.
38. Tien C L, Yang H M, Liu M C. The measurement of surface roughness of optical thin films based on fast Fourier transform. *Thin Solid Films* 2009; 517(17): 5110–5115, <https://doi.org/10.1016/j.tsf.2009.03.193>
39. Vo Q, Fang F, Zhang X, Gao H. Surface recovery algorithm in white light interferometry based on combined white light phase shifting and fast Fourier transform algorithms. *Applied Optics* 2017; 56(29): 8174-8185, <https://doi.org/10.1364/AO.56.008174>
40. Vorburger T V, Rhee H G, Renegar T B, Song J F, Zheng A. Comparison of optical and stylus methods for measurement of surface texture. *The International Journal of Advanced Manufacturing Technology* 2007; 33: 110–118, <https://doi.org/10.1007/s00170-007-0953-8>
41. Wang C, D-Amato R, Gomez E. Confidence Distance Matrix for outlier identification A new method to improve the characterisations of surfaces measured by confocal microscopy. *Measurement* 2019; 137: 484-500, <https://doi.org/10.1016/j.measurement.2019.01.043>
42. Wang Y, Liu Y, Zhang G, Wang Y. A simulation method for non-Gaussian rough surfaces using Fast Fourier Transform and translation process theory. *Journal of Tribology-Transactions of the ASME* 2018; 140(2) 021403, <https://doi.org/10.1115/1.4037793>
43. Wang Y, Yuan P, Ma J, Qian L. Scattering noise and measurement artifacts in a single-shot cross-correlator and their suppression. *Applied Physics B* 2013; 111: 501–508, <https://doi.org/10.1007/s00340-013-5364-y>
44. Wdowik R, Magdziak M, Porzycki J. Measurements of surface roughness in ultrasonic assisted grinding of ceramic materials. *Applied Mechanics and Materials* 2014; 627: 191–196, <https://doi.org/10.4028/www.scientific.net/AMM.627.191>
45. Zhang H, Yuan Y, Piao W. A universal spline filter for surface metrology. *Measurement* 2010; 43(10): 1575-1582, <https://doi.org/10.1016/j.measurement.2010.09.008>



Article citation info:

Kulinowski P, Kasza P, Zarzycki J. Identification of the operating parameters of the friction drum drive in industrial conditions. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 94–102, <http://dx.doi.org/10.17531/ein.2021.1.10>.

## Identification of the operating parameters of the friction drum drive in industrial conditions

Indexed by:



Piotr Kulinowski<sup>a</sup>, Piotr Kasza<sup>a</sup>, Jacek Zarzycki<sup>a</sup>

<sup>a</sup>AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakow, Poland

### Highlights

- Measurement of the friction coefficient in the laboratory cannot fully reflect the actual conditions.
- Characteristics of the operation of the drive-tensioning system allows for determining the slip limit.
- The mobile measuring system enables determining the conveyor's characteristics.
- A technique for measuring the operational friction coefficient of a drum drive has been developed.

### Abstract

The publication presents the results of measurements of the operating parameters of the drum drive of a belt conveyor operating in a copper ore mine. The laboratory and industrial tests of belt conveyor components so far have largely focused on idler sets, the belt and the conveyor route. The authors of the publication notice that the subject literature lacks information on research into the coefficient of friction between the belt and the drum in an industrial facility under real conditions, which may imply that the phenomenon of driving force transmission from the drum surface to the belt has not been thoroughly explored. The investigations described in the publication were aimed at determining the kinetic coefficient of friction between the conveyor belt and the lining of the drive drum under operating conditions. In the first part of the study, preliminary laboratory tests were carried out, whereas in the second part of the study, a mobile, non-invasive measurement system was applied, which allowed for recording the kinematic and dynamic parameters of the conveyor's operation. During the several dozen minutes of the conveyor's operation, there were several start-ups and brakings as well as periods of steady operation under variable load on the material handled. The non-typical planned sequence of switching on the drive motors caused a temporary slip of the belt on the drive drum surface during one of the start-ups. The recorded parameters of the conveyor operation enabled determining the operational limit value of the friction coefficient between the belt and the lining of the drive drum.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

belt conveyor, friction coefficient, frictional contact, exploitation characteristics.

## 1. Introduction

Belt conveyors belong to the most effective means of transporting loose materials. Global trends in the development of the raw materials industry, in which conveyor transport systems are of crucial importance, force the design of ever longer and effective haulage routes. Modern belt conveyors should be characterized by high reliability and durability as well as low investment and operating costs.

Currently conducted research and development works in the field of conveyor belt transport include diagnostics and monitoring as well as laboratory and industrial tests. The knowledge and experience gathered during these investigations help to improve the calculation methods and optimize the conveyor design.

The use of advanced diagnostic and monitoring methods significantly contributes to the improvement of the reliability and safety of belt transport systems operation [17, 18, 26]. In [9], the authors indicate that the development of conveyor transport systems requires modern systems for monitoring the condition of components and effective diagnostic systems that implement the assumptions of Industry 4.0. One of the current trends in the development of conveyor trans-

port is also the reduction of energy consumption through the use of speed control systems as a function of the degree of the conveyor's loading [6].

In the field of laboratory tests, particular attention is paid to the testing of conveyor components, such as idlers and belts [1]. The works carried out in this area confirm that the operational features of the conveyor's subassemblies may be determined in laboratory tests, provided that the actual operating conditions have been reproduced, for example the condition of the idler load with radial and axial forces when measuring the rotational resistance [11, 16].

Industrial testing of belt conveyors is a necessary element in the process of calculation methods verification. The publication [2] emphasizes the need to identify the load of the conveyor with material transported in real conditions, using dedicated specialized equipment. Measurements of conveyor operation parameters in industrial conditions are necessary to verify some calculation factors, such as the main resistance coefficient, which is discussed in the paper [4]. The results of properly planned and systematically conducted industrial tests provide a basis for introducing advanced, nonlinear computa-

E-mail addresses: P. Kulinowski - [piotr.kulinowski@agh.edu.pl](mailto:piotr.kulinowski@agh.edu.pl), P. Kasza - [piotr.kasza@agh.edu.pl](mailto:piotr.kasza@agh.edu.pl), J. Zarzycki - [jacek.zarzycki@agh.edu.pl](mailto:jacek.zarzycki@agh.edu.pl)



tional models into design practice [7]. Tests conducted on real objects are used to optimize the key components of the conveyor, for example, the publication [12] presents the measurement of the condition of idler sets load, which was used to determine the optimal geometric parameters of the conveyor's route.

Currently, there are many IT tools available to support the process of designing belt conveyors. They use various calculation models, such as those specified in the CEMA and DIN 22101 standards, as well as proprietary methods developed in research centres, e.g. at the University of Hannover, Wrocław University of Technology or the AGH University of Science and Technology. All these methods require defining unambiguous input parameters, the values of which should be verified in laboratory and industrial tests on real objects. It is necessary as the comparison of the standard calculation methods used in the publication [20] shows significant discrepancies in the results.

Useful information on the functioning of belt conveyor transfer points is provided by the results of simulation tests employing the discrete element method [21]. The authors use the DEM environment to define guidelines for designing reliable chutes which ensure reduction of belt motion resistance, central feeding of mined rock and limitation of the process of mined rock defragmentation. Verified computational procedures and simulation models are applied to improve the efficiency of the conveyor's work by determining the most advantageous construction parameters of the belt [23, 24], correcting the spacing of idler supports [3] or through a proper selection of belt tensioning systems [5, 22].

In the world literature, there are no mentions of the results of research into the coefficient of friction between the belt and the drum carried out in an industrial facility, which may indicate that the phenomenon of the driving force transmission from the drum surface to the belt has not been thoroughly understood. These tests may be particularly important due to the new materials used for drum linings. One of the current design trends is to apply a new type of rubber-ceramic lining on drive drums [19].

Currently, the development of measuring equipment allows for determining the value of the actual coefficient of friction in variable industrial conditions, which in the future will enable designing a safer and more reliable drum friction drive.

In belt conveyors with a drum drive, the driving force is transferred to the belt thanks to the phenomenon of frictional contact. Striving to build long conveyors, the designers were forced to reduce the value of safety factors, use the belt's strength parameters more fully and reduce and constantly control the tension force at the minimum acceptable level. It was therefore necessary to better understand the phenomena taking place when the belt is going through the drive drum, in order to determine the conditions under which a slip between the belt and the drum occurred. The phenomenon of frictional contact is described by the Euler-Eytelwein dependence ( $S_2 \leq S_2 \cdot e^{\mu\alpha}$ ), which has been known for more than 200 years, where variable parameters are belt-tensioning force  $S_2$ , drive drum's wrap angle  $\alpha$  and friction coefficient  $\mu$  [15]. The belt tensioning force is a parameter controlled by the tensioning system, the wrap angle depends on the geometric configuration of the drive drums, while the value of the friction coefficient is most often determined on the basis of tests carried out in laboratory conditions [25]. In fact, the value of the friction coefficient is a stochastic parameter, depending on, among

other things, the values of unit pressure and slip velocity, the condition of the surface of the friction pair, the type of intermediate layer or moisture. The measurement method used may also have a significant impact on the determined value of the friction coefficient.

This publication presents the results of laboratory investigations into the kinetic friction coefficient and an original method of its determination applied directly in the actual conditions of the belt conveyor's operation.

## 2. Preliminary laboratory tests

During the initial phase of the tests, the friction coefficient was determined in laboratory conditions for a clean and dirty belt, as well as a wet and dry belt, at low values of slip velocity, simulating the slip of the belt unwound on the drum surface, and at pressures corresponding to real conditions. The contamination was copper ore dust. Similar tests, in which the friction pair were spring elements and the results were the values of static and kinetic friction coefficients, have been presented in publication [8]. These tests took into account the different surface condition, different values of slip velocity and variable values of the normal load.

Laboratory tests of the friction coefficient between the belt cover and a fragment of the drum lining were carried out on a special stand (Fig. 1), which consists of a movable frame (1) with an attached and properly tensioned counter-sample (2). A counter-sample is a longitudinally cut tape having dimensions of 800x50 mm. In the stationary clamping system (3), there are mounted samples cut from the material of the drive drums lining in the form of two cubes (4), each of which has a contact surface of 20x20 mm. The drive system consists of an inverter-controlled gear motor (5) and a belt transmission. The frame is advanced by means of a screw and nut mechanism.

The friction coefficient tests were carried out for a fixed pressure force  $N=200N$ , tension  $P_1=P_2=2000N$  and slip velocity  $v=0,6mm/s$ . Before the actual tests, the running-in of the tape-lining friction pair

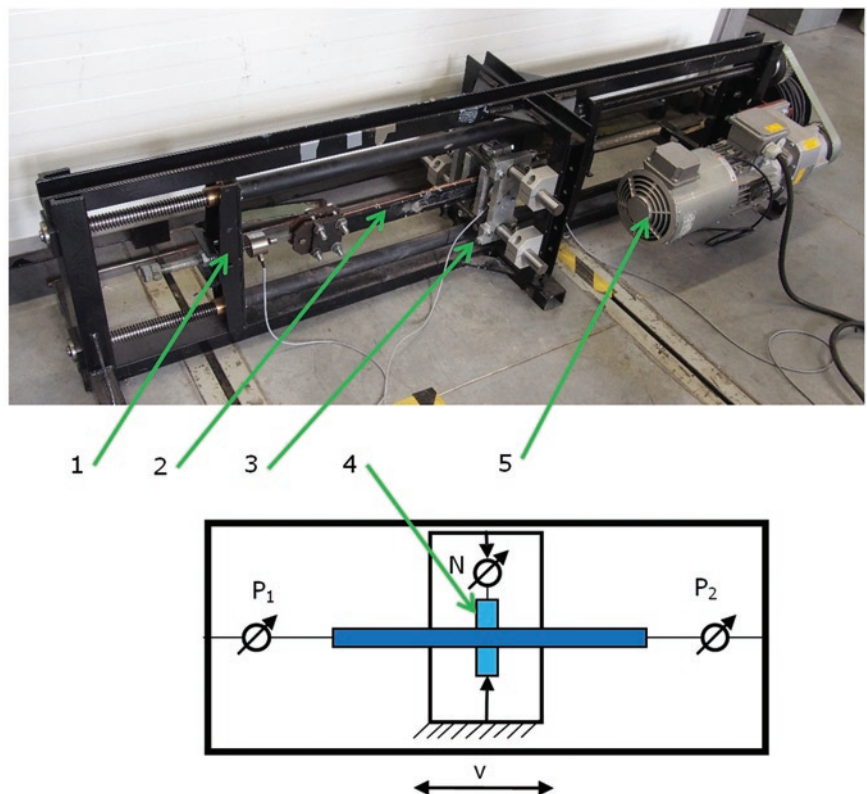


Fig. 1. View and schematic diagram of the stand for testing the friction coefficient [8]: 1 – movable frame, 2 – belt counter-sample, 3 – stationary sample clamping system, 4 – drum lining samples, 5 – drive

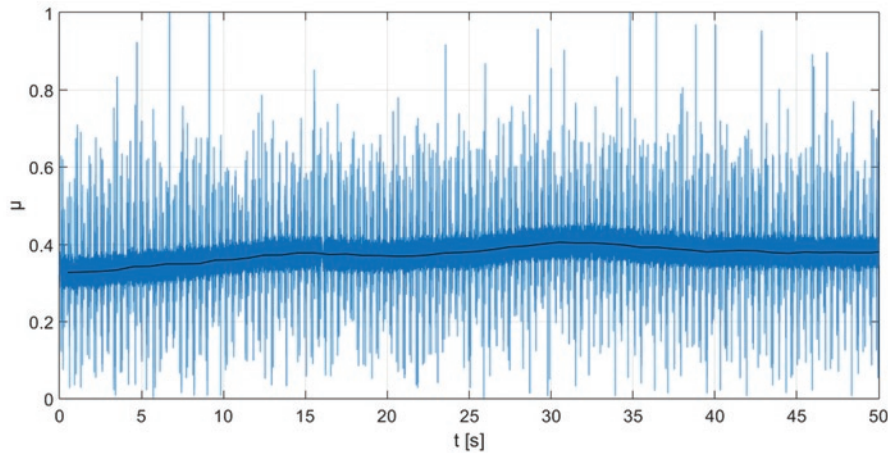


Fig. 2. Course of the friction coefficient for a wet and dust-contaminated friction pair [8]

was performed over a length of approx. 50 mm to ensure stable friction conditions and repeatable results [10].

The tests were carried out many times for clean and contaminated samples, in both directions, while paying attention to the anisotropic nature of the friction phenomenon. An example of the changes in the friction coefficient during one of the tests is shown in Fig. 2

The value of the friction coefficient of the sample contaminated with dust from a finely ground copper ore determined during laboratory tests ranged from 0,37 to 0,53, depending on the degree of moisture and the direction of mutual movement of the samples and the counter-sample. The test results confirmed the random nature of changes in the friction coefficient value as well as the influence of the friction pair cooperation conditions and the method of conducting the measurement test on its value.

### 3. Characteristics of the operation of the drive-tensioning system

The aim of the research team was to determine the friction coefficient between the belt and the drum in industrial conditions. The necessary condition was to make the unrolled belt slip on the drum surface during the start-up of the belt, while simultaneously measuring the kinematic and dynamic parameters of the drive system. Determination of the belt slip conditions is facilitated by a diagram showing the characteristics of operation of the belt conveyor drive-tensioning system. This characteristic takes into account the features of the drive system and the tensioning device by presenting the required values of the belt tensioning force as a function of torque or driving force [13, 15].

#### 3.1. Minimum values of the belt tensioning force

The value of the belt tensioning force for a given conveyor must be selected so that two conditions are met:

- 1) the belt sags between the idler sets of the upper and lower branches must be limited in order to maintain the proper geometric shape of the belt,
- 2) non-slip cooperation between the belt and the drum being driven or braked must be ensured.

The first condition is the minimum force in the belt assuming an acceptable belt sag. The most frequently assumed value is  $f_u/l_g = 0,015$ , and, during the braking of the belt, the quotient  $f_u/l_g = 0,04$  is allowed [15]. This force does not directly depend on the driving torque, but only on the spacing of the idler sets and the transported masses

(Fig. 3). Its value is determined on the basis of the following equation:

$$S_{\min} = \frac{g \cdot (m_t + m_u) \cdot l_d^2}{8 \cdot f_u} \quad [\text{N}] \quad (1)$$

where:  $m_t, m_u$  - unit masses of belt and mined rock, [kg/m];  $l_g$  - distance between upper idler sets, [m];  $f_u$  - belt deflection between idler sets, [m];  $g = 9,81 \text{ m/s}^2$ .

The second condition, ensuring the correct frictional cooperation between the belt and the drum during start-up, is determined by the following dependence:

$$k_p \cdot P_N \leq S_2 (e^{\mu\alpha} - 1) \quad (2)$$

where:  $P_N$  - peripheral driving force, [N];  $k_p$  - slip protection factor, [-];  $S_2$  - force in the belt running off the drive drum, [N],  $\alpha$  - drive drum's wrap angle, [rad];  $\mu$  - friction coefficient [-].

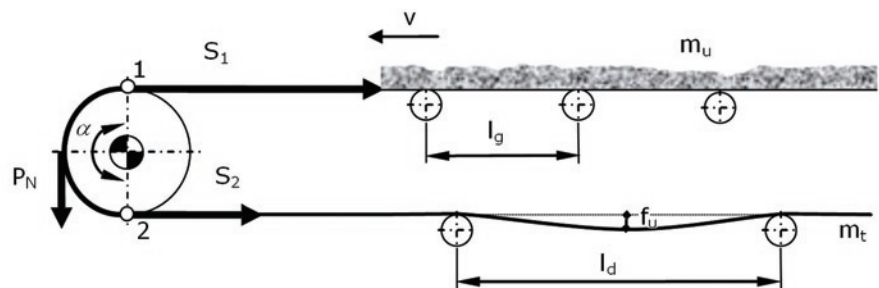


Fig. 3. Forces in the belt during drum driving

For a determined value of the peripheral driving force  $P_N$ , the tensioning device situated in the vicinity of the drive drum should ensure force  $S_2$ , defined by the following dependences:

$$S_2 \geq P_N \frac{k_p}{(e^{\mu\alpha} - 1)} ; S_2 \geq \frac{g \cdot m_t \cdot l_d^2}{8 \cdot f_u} \quad [\text{N}] \quad (3)$$

where:  $l_d$  - distance between lower idler sets, [m].

The value of wrap angle  $\alpha$  for a given conveyor is constant, whereas the friction coefficient  $\mu$  is a random variable. Assuming that the value of  $\mu$  in the above relationships is constant and safe for particular conditions, we will obtain linear dependencies on the basis of which it can be concluded that the minimum forces  $S_2$  which the tensioning device should induce in order to ensure the correct operation of the belt conveyor within the range  $P \in [0; P_{N\max}]$  are defined by lines  $a$  and  $b$  (Fig. 4).

Fig. 4 shows a set of values of force  $S_2$  which are unacceptable during normal operation. A decrease in the value of the force in the (tensioning) belt running off the drive drum below line  $b$  may lead to the belt slip on the drum and result in the drum lining being worn and causing a fire hazard.

#### 3.2. Actual characteristics of the conveyor belt tensioning device

Based on the calculated minimum values of the belt tension force according to the formulas: (1), (2) and (3), the required value of the

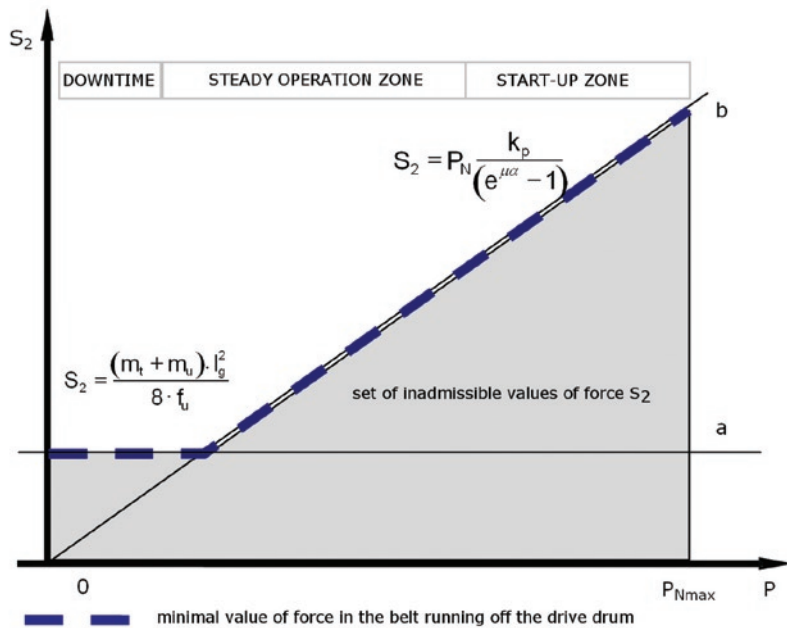


Fig. 4. Minimum values of the belt pre-tension force  $S_2$  as a function of the driving force  $P$  on the driven drum

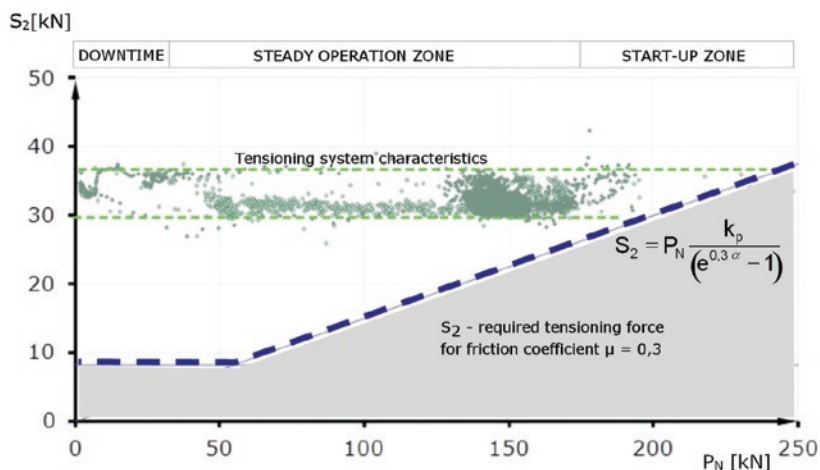


Fig. 5. Operational characteristics of the drive-tensioning system with a gravity tensioning device, determined on the basis of measurements

force in the belt running off the drive drum as a function of the driving force is determined. Next, the characteristics of tensioning devices, obtained on the basis of calculations, simulation and laboratory or industrial tests [13], are placed on a simplified diagram showing these dependencies (Fig. 2).

Fig. 5 below shows an example of the actual characteristics of the tensioning device obtained on the basis of the results of a series of measurements of forces in the conveyor belt during start-ups, steady operation and braking of the conveyor belt. The belt conveyor was equipped with a gravity belt tensioning system, located just behind the main conveyor drive. The belt tensioning force was in the range of  $29 \div 37$  kN. Fluctuations in this force were caused by the inertia of the weights and the efficiency of the tensioner system.

The diagram of the permissible, minimum forces in the belt presented in Fig. 5 indicates that the drive will work non-slip with the actual characteristics of the gravity tensioning system. The necessary condition is to use drum linings and effective cleaning devices ensuring a minimum value of the friction coefficient  $\mu=0,3$ .

The actual characteristics of the tensioning system shown in Fig. 5 were recorded in industrial conditions, using a mobile measuring system described in Chapter 4.

#### 4. Industrial tests

A typical mining belt conveyor that transports copper ore was selected for industrial measurements of the friction coefficient between the drum lining and the conveyor belt. The measuring system described below, intended for the testing of industrial belt conveyors, meets the requirement of measuring stations' mobility and non-invasiveness, which was set at the design stage. The mobility of the stations enables installing them on the conveyor on the day of measurements as well as preparing for data recording in no more than 2 hours. (installation of measuring equipment, sensors, recorders, checking the correct operation of measurement chains). On the other hand, non-invasiveness of the measuring system ensures that none of the conveyor components will be modified or replaced, and that the tested belt conveyor can at any time resume its standard transport task. Moreover, it has been assumed that all measured physical quantities will be registered simultaneously and each of the data acquisition points of the measuring system will be powered autonomously.

A non-invasive mobile measuring system in its basic configuration includes three stations that record measurement signals during start-up, steady operation under variable load and during the conveyor's braking. One station, recording the speed of the upper and lower belt as well as the feed efficiency, is located on the conveyor route, the second one is in the electrical switchboard, and the third one is placed in the area of the conveyor's drive and tensioning system. Measurement data from this stand was used to verify the actual value of the friction coefficient between the belt and the surface of the drive drum.

The measuring station for the drive and tensioning system, shown in Fig. 6, is located in the area of the drive and tensioning station. The station is equipped with a box with a battery power supply system, a measuring computer, sensors for measuring the force in the belt and line, a set of belt speed sensors and optical rotational speed sensors.

During measurements, it is possible to observe and record changes in the following parameters:

- $x_A, x_B$  - belt displacement at stations A and B [m],
- $v_A, v_B$  - belt speed at stations A and B [m],
- $x_l$  - rope displacement in the belt tensioning system [m],
- $S_L$  - force in the stationary line of the belt tensioning system [kN],
- $R_A, R_B$  - reaction forces on the deflecting drums of the measuring system A, B [kN]
- $n$  - rotational speeds of drive motors shafts, high-speed shafts of gears or drum [rpm].

Based on the measured parameters, the following are determined:

- $ST_N, ST_Z$  - forces in the belt at stations A and B, i.e. in the area where the belts run onto or off the drive drums [kN],
- $s$  - the actual safety factor [-],
- $P$  - peripheral driving force on the drive drums, balancing the resistance to the conveyor's movement  $W$  [kN],
- $M$  - driving torque in drive units equipped with hydrodynamic couplings having known characteristics [Nm],
- $N$  - drive power [kW],
- $x_w$  - stroke of the belt tensioning trolley [m],

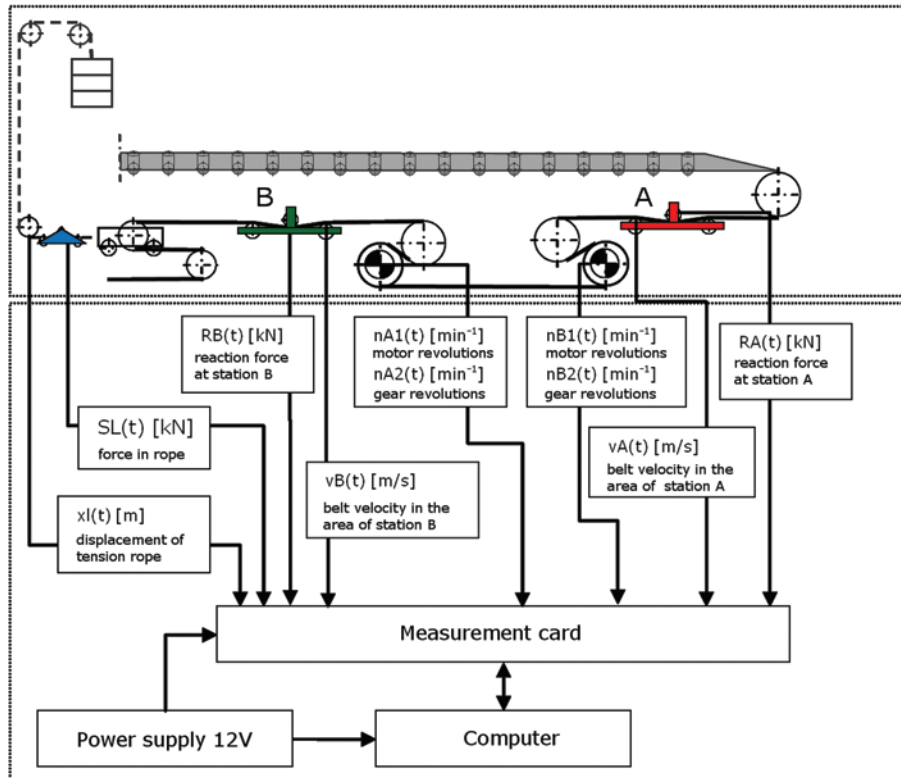


Fig. 6. Diagram of the measuring chains of the complete measuring system of the drive-tensioning system

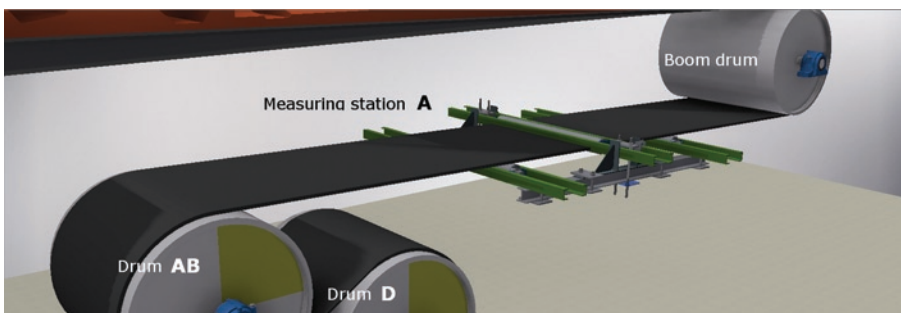


Fig. 7. Stand A for measuring the force in the belt running on the drive drums

- $v_w$  - speed of the belt-tensioning trolley [m / s],
- $\eta$  - efficiency of the tensioning system [-],
- $\mu$  - friction coefficient between the belt and the drive drum surface [-].

Fig. 7. shows the visualization of the installation of measuring station A in the area where the belt runs onto the drum drive system.

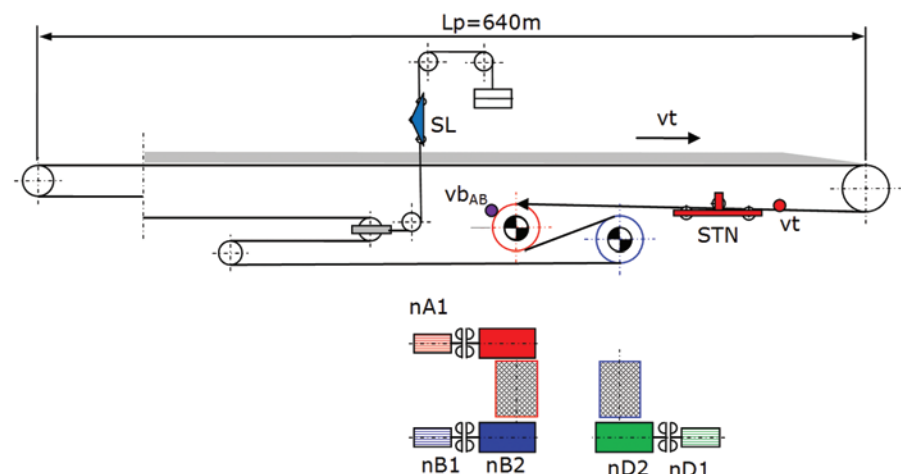
The industrial tests were carried out on a belt conveyor with a gravity belt tensioning system and a two-drum drive system, equipped with three motors and hydrodynamic starter couplings installed in accordance with the diagram in Fig. 8.

Technical and operational parameters of the tested conveyor:

- Conveyor's length :  $L = 640$  m,
- Belt speed:  $v = 2,73$  m/s,
- Average inclination angle of the conveyor route:  $\delta = 2,87^\circ$ ,
- Type of tensioning device: gravity.

During the measurements, a series of start-ups of the conveyor was carried out with variable load of the handled material and for different sequences of switching on motors A, B, D. Among others, the following were recorded: belt speed in the drive area, forces in the belt running onto the drive system (STN) and in the tensioning system line (SL), rotational speed of the motor shafts ( $nA1$ ,  $nB1$ ,  $nD1$ ) and gears ( $nB2$ ,  $nD2$ ) and peripheral speed of drum AB ( $v_{bAB}$ ) - Fig. 8.

The measurement results were used, among others, to determine the friction coefficient between the belt and the lining of the drive drum



- STN - force measured in the belt running onto the drive drum AB [kN]
- STM - calculated force in the belt between drums [kN],  $STM = STZ + P_D$
- STZ - force measured in the belt running off the drive drum D [kN]

Fig. 8. Diagram of the tested conveyor and location of measuring points



Fig. 9. Diagram of a two-drum drive system

coefficient between the belt and the lining of the drive drum AB. The delay in turning on motor D resulted in a macro-slip on drive drum AB.

The conveyor was started up according to the following sequence: start-up of motor A, after 0,7 s start-up of motor B and, after seven seconds, start-up of motor D. The belt speed and the peripheral speed of the drum were recorded using a mobile measuring system.

The difference between the values of the belt speed  $v_t$  and the peripheral speed of the drum  $v_{b_{AB}}$  represents the belt slip on the drum (Fig. 10). The slip began after approx. 1,5 s from the

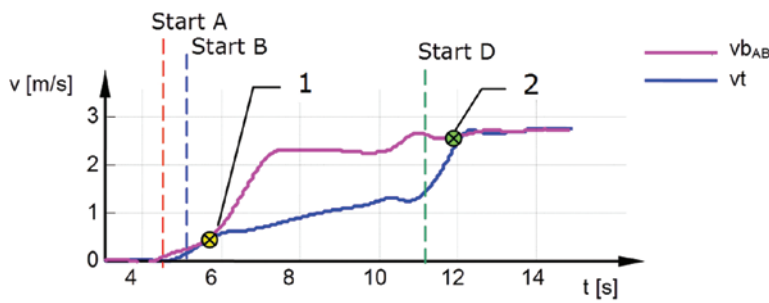


Fig. 10. Method of measuring the belt speed  $v_t$  and the peripheral speed of the drum  $v_{b_{AB}}$

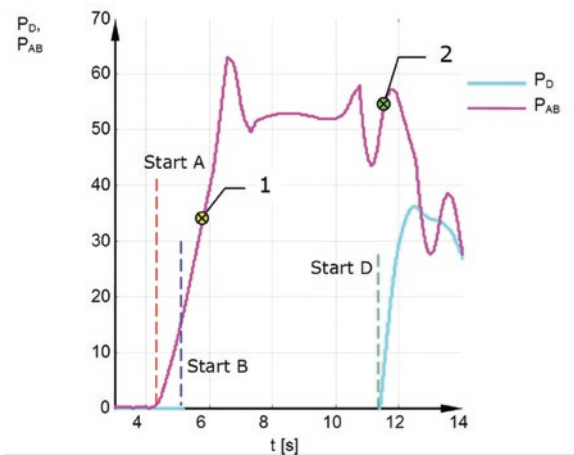


Fig. 13. Course of changes in the peripheral force on drums AB and D.

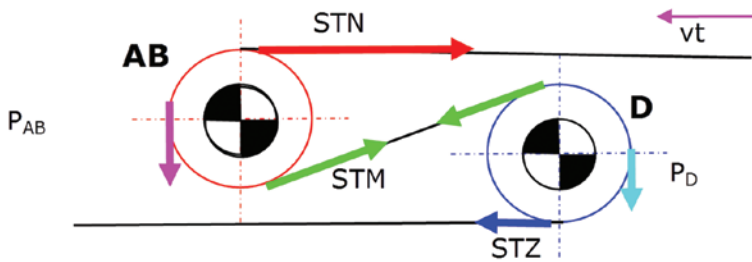


Fig. 11. Course of changes in the belt speed  $v_t$  and the peripheral speed of the drum  $v_{b_{AB}}$

start-up of motor A (Fig. 11 - point 1), lasted about 6 seconds and ended after increasing the belt tension, when motor D was started up (Fig. 11 - point 2).

The course of changes in the peripheral driving force for drums AB and D was determined on the basis of the recorded courses of changes in the slip of the coupling (Fig. 12) and the knowledge of hydrodynamic couplings characteristics, which enable calculating the torque of the coupling  $M_s$ . The characteristics of hydrodynamic couplings were determined in the laboratory of their manufacturer. The knowledge of the driving force  $P_D = M_s \cdot i / R_{AB}$  and the force in the STZ belt running off the drive drum enabled determining the force in the belt between the drive

The case of a slip between the belt and the drum during start-up was analysed.

The possibility of changing the sequence of switching on motors A, B, D allowed for an experiment that involved determining the friction

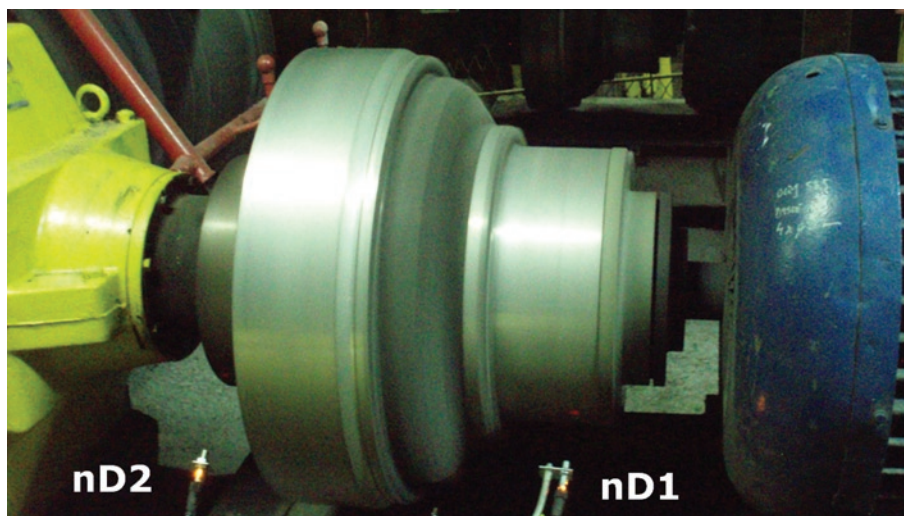


Fig. 12. Measurement of the rotational speed of the pump and turbine part of the hydrodynamic coupling

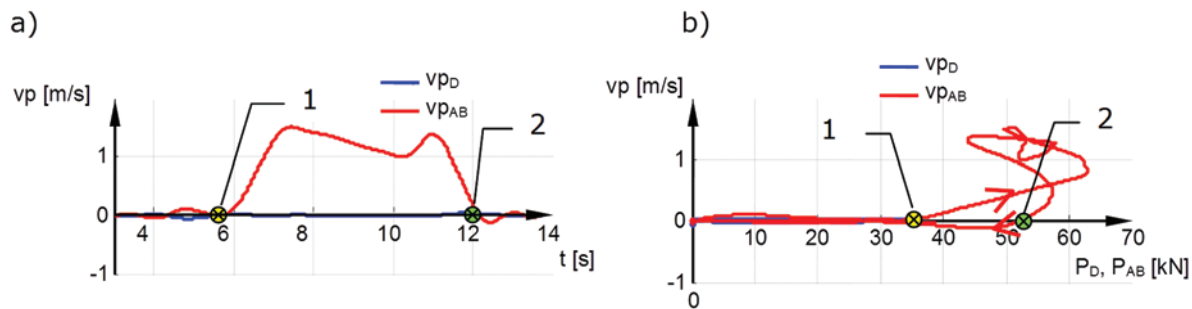


Fig. 14. Course of changes in the speed of belt slip drum AB -  $vp_{AB}$  and D -  $vp_D$  versus time (a) and as a function of peripheral force on drive drums (b)

drums  $STM = P_D + STZ$  (Fig.9). The course of changes of this force is shown in Fig. 16a.

The course of the peripheral driving force for drums AB and D is shown in Fig. 13. Points 1 and 2 mark the beginning and end of the slip on drum AB.

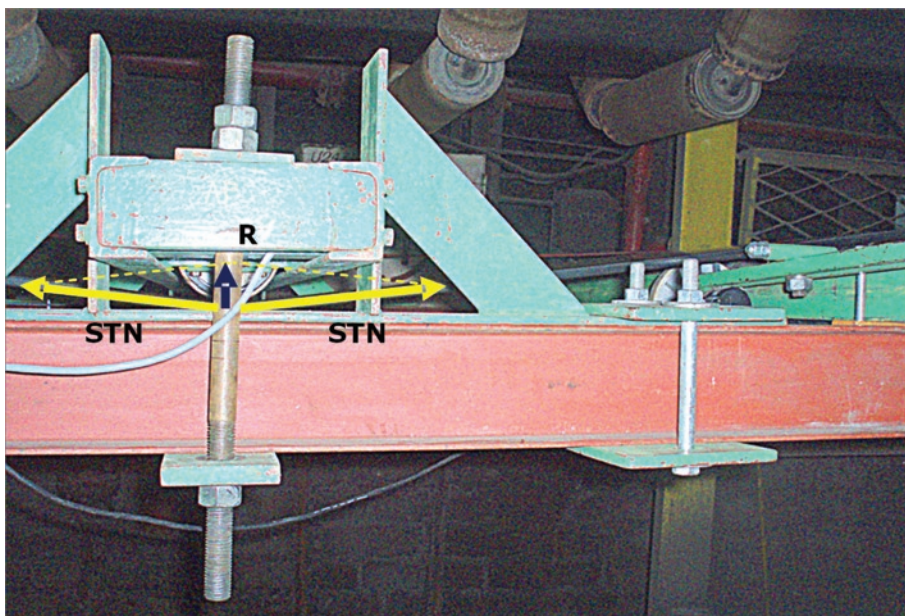


Fig. 15. Method of measuring the force in the STN belt at station A

Determination of the beginning and end of the slip phase enabled presenting the changes in the belt slip speed on drums AB and D as a function of the determined peripheral driving force (Fig. 14b). Slip speed is the difference between the peripheral speed of the drum and the speed of the belt ( $vp_{AB}=vb_{AB}-vt$ ). It can be read from the diagram that the belt slip on drum AB started with a driving force of approx. 34 kN and ended at approx. 53 kN. There was no slip on drum D.

The force in the belt running onto drive drums - STN was determined on the basis of measurements of the value of the reaction force acting on the roller deflecting the belt run (Fig. 15). The force in the belt running off the drive drum was calculated on the basis of measurements of the force in the line of the SL belt tensioning system (Fig. 8).

The knowledge of the force in the belt between the drive drums - STM and the peripheral driving force -  $P_{AB}$  allowed for presenting the characteristics of the drive-tensioning system for drum AB. Figure 14b shows the actual characteristics of the drive-tensioning system and the values of the tensioning force for the friction coefficient in the range of  $0,1 \div 0,5$ , which

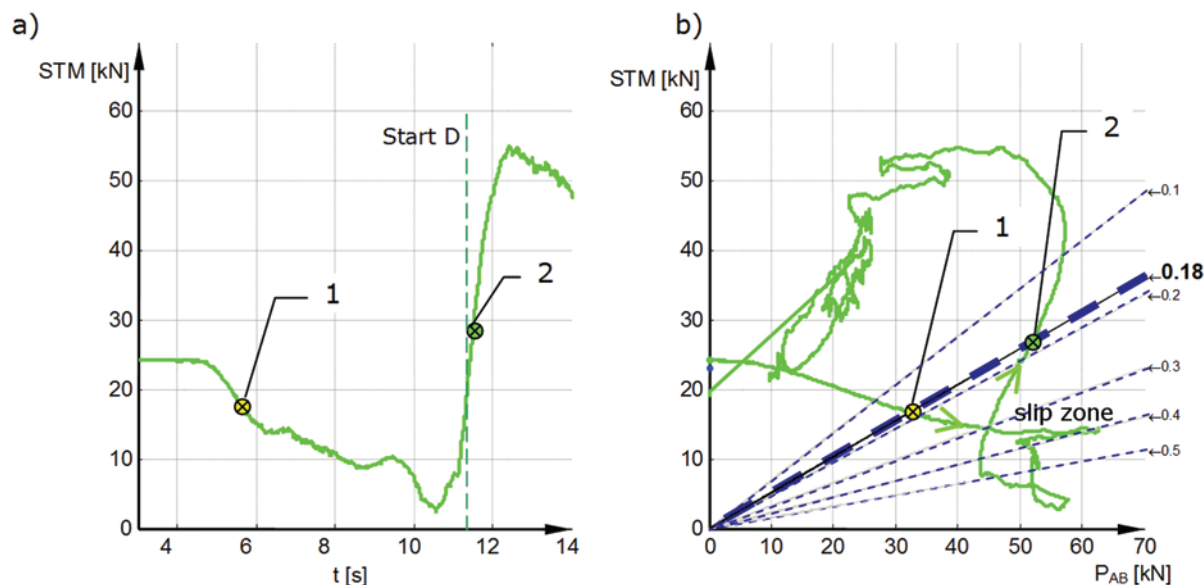


Fig. 16. Course of changes of the force in the belt between the drums during STM start-up versus time (a) and as characteristics of operation of the drive-tensioning system for drive drum AB (b)



Fig. 17. Condition of the drive drum lining during the tests

are required due to the anti-slip protection. The required values of belt tension are marked with dotted lines (Fig. 14b).

The characteristics of the drive-tensioning system (Fig. 16b) have been marked with points representing the beginning (1) and the end of the belt slip (2), taking into account the course of changes in the tensioning force versus time (Fig. 16a) as well as the values of the driving force  $P_{AB}$  that mark the beginning and end of the slip (Fig. 12b item 1 and item 2). The straight line passing through these points corresponds to the required tensioning force determined for the friction coefficient  $\mu=0,18$ .

The value of the coefficient of friction between the contaminated belt lining and the lining of drive drum AB, determined by the indirect measurement method in industrial conditions, was  $\mu=0,18$  at the time of slip. It is a value below the level required for designing a conveyor ( $\mu>0,25$ ), but seems probable considering the operating conditions of the conveyor and the condition of the drum lining (Fig. 17.)

## 5. Summary

The mobile, non-invasive measuring system of belt conveyors and the measuring equipment as well as the appropriate procedures applied proved to be fully suitable for measurements performed in underground mines. The results of measurements can be used for diagnostic purposes. They can demonstrate the correct operation of conveyors, but also indicate errors in design assumptions, calculations and selection of conveyor components, resulting in reduced durability of components and lower reliability of the conveyor, which translates into increased operating costs. An important task for the mobile meas-

urement system is verification of the applied calculation models and selected calculation factors. The conducted industrial tests allowed for extending the scope of calculations of programs supporting the design process and enabled validation and improvement of models used to analyse dynamic phenomena occurring during the operation of the conveyor [13, 14, 15].

As part of the research presented in this publication, the actual performance characteristics of the mining belt conveyor drive and tensioning system were determined. These characteristics illustrated the case of the conveyor start-up in a situation when a slip occurred between the belt and the drive drum. Based on the characteristics, the actual limit value of the friction coefficient between the belt and the drum lining was estimated, which in this case reached  $\mu=0,18$ . The low value of the coefficient is most likely caused by the high degree of wear and contamination of the drive drum lining.

The value of the kinetic friction coefficient determined in laboratory conditions proved to be more than twice higher than the value obtained during industrial measurements. This indicates that it is difficult to perform repeatable and reliable tests of the kinetic friction coefficient for friction pairs used in drive systems. The rubber lining of the drum and the belt work in conditions of high pressure, significant contamination and moisture, which causes their accelerated abrasion and wear. The problem of belt slip on the drive drum of the test conveyor was solved by accelerating the start-up of motor D, which resulted in an increase in the force in the belt running off drum AB and a correct start-up of the conveyor.

## References

1. Andrejiova M, Grincova A, Marasova D. Analysis of tensile properties of worn fabric conveyor belts with renovated cover and with the different carcass type. *Eksplotacja i Niezawodność - Maintenance and Reliability* 2020; 22(3): 472-481, <https://doi.org/10.17531/ein.2020.3.10>
2. Gładysiewicz L, Król R, Kisielewski W. Measurements of loads on belt conveyor idlers operated in real conditions. *Measurement: Journal of the International Measurement Confederation* 2019; 134: 336-344, <https://doi.org/10.1016/j.measurement.2018.10.068>.
3. Gładysiewicz L, Kawalec W, Król R. Selection of carry idlers spacing of belt conveyor taking into account random stream of transported bulk material. *Eksplotacja i Niezawodność - Maintenance and Reliability* 2016; 18(1): 32-37, <https://doi.org/10.17531/ein.2016.1.5>.
4. Gładysiewicz L, Król R, Kisielewski W, Kaszuba D. Experimental determination of belt conveyors artificial friction coefficient. *Acta Montanistica Slovaca* 2017;22(2): 206-214.
5. Goncharov KA, Grishin AV. Theoretical study of influence of belt tension of intermediate belt conveyor drive on value of zone of relative slip of traction and carrying belts. *IOP Conference Series: Earth and Environmental Science* 2017; 87(2): 022008, <https://doi.org/10.1088/1755->

- 1315/87/2/022008.
6. Halepoto IA, Shaikh MZ, Chowdhry BS, Uqaili Muhammad A. Design and Implementation of Intelligent Energy Efficient Conveyor System Model Based on Variable Speed Drive Control and Physical Modeling. *International Journal of Control and Automation* 2016; 9(6): 379-388, <https://doi.org/10.14257/ijca.2016.9.6.36>.
  7. Harrison A. A comparison of friction models for conveyor design. 9th International Conference on Bulk Materials Storage, Handling and Transportation, ICBMH 2007.
  8. Hryciów Z, Krasoń W, Wysocki J. The experimental tests on the friction coefficient between the leaves of the multi-leaf spring considering a condition of the friction surfaces. *Eksploracja i Niezawodność - Maintenance and Reliability* 2018; 20(4): 682-688, <https://doi.org/10.17531/ein.2018.4.19>.
  9. Jurdziak L, Blazej R, Bajda M. Conveyor belt 4.0. *Advances in Intelligent Systems and Computing* 2019; 835: 645-654, [https://doi.org/10.1007/978-3-319-97490-3\\_61](https://doi.org/10.1007/978-3-319-97490-3_61).
  10. Kasza P, Kulinowski P, Zarzycki J. The Influence of an Operating Conditions on the Friction Coefficient in Transportation Machines Drives. *New Trends in Production Engineering* 2020; 3(1): 294-302, <https://doi.org/10.2478/ntpe-2020-0024>.
  11. Król R. Studies of the Durability of Belt Conveyor Idlers with Working Loads Taken into Account. *IOP Conference Series: Earth and Environmental Science* 2017; 95(4): 042054, <https://doi.org/10.1088/1755-1315/95/4/042054>.
  12. Król R, Kisielewski W. Research of loading carrying idlers used in belt conveyor-practical applications. *Diagnostyka* 2014; 15 (1): 67-73.
  13. Kulinowski P. Simulation method of designing and selecting tensioning systems for mining belt conveyors. *Archives of Mining Sciences* 2014; 59(1): 123-138, <https://doi.org/10.2478/amsc-2014-0009>.
  14. Kulinowski P. Simulation studies as the part of an integrated design process dealing with belt conveyor operation. *Eksploracja i Niezawodność - Maintenance and Reliability* 2013; 15 (1): 83-88.
  15. Kulinowski P. Analytical method of designing and selecting take-up systems for mining belt conveyors. *Archives of Mining Sciences* 2013; 58(4): 1301-1315, <https://doi.org/10.2478/amsc-2013-0090>.
  16. Kulinowski P, Kasza P, Zarzycki J. The methodology of testing rotational resistances of the rollers under the operation load. *New Trends in Production Engineering* 2019; 2(1): 337-343, <https://doi.org/10.2478/ntpe-2019-0036>.
  17. Mazurkiewicz D. Computer-aided maintenance and reliability management systems for conveyor belts. *Eksploracja i Niezawodność - Maintenance and Reliability* 2014; 16(3): 377-382.
  18. Mazurkiewicz D. Analysis of the ageing impact on the strength of the adhesive sealed joints of conveyor belts. *Journal of Materials Processing Technology* 2008; 208(1-3): 477-485, <https://doi.org/10.1016/j.jmatprotec.2008.01.012>.
  19. Patton PW, Smith CS. Toward an effective friction factor for ceramic-in-rubber pulley lagging: An analysis of field data. *Bulk Material Handling by Conveyor Belt V.* 2004; 107-109.
  20. Satria I, Rusli M. A Comparison of Effective Tension Calculation for Design Belt Conveyor between CEMA and DIN Standard. *MATEC Web of Conferences* 2018; 166: 01007, <https://doi.org/10.1051/mateconf/201816601007>.
  21. Walker P, Doroszuk B, Król R. Analysis of ore flow through longitudinal belt conveyor transfer point. *Eksploracja i Niezawodność - Maintenance and Reliability* 2020; 22 (3): 536-543, <https://doi.org/10.17531/ein.2020.3.17>.
  22. Wenrong W, Jianhua Q, Wangheng. Research on the influence of tensioning device in the dynamic process of belt conveyor. *ACM International Conference Proceeding Series* 2018; 186-190, <https://doi.org/10.1145/3305275.3305312>.
  23. Wheeler C, Munzenberger P, Ausling D, Beh B. How to design energy efficient belt conveyors. *Bulk Solids Handling* 2015; 6: 40-50.
  24. Woźniak D. Laboratory tests of indentation rolling resistance of conveyor belts. *Measurement: Journal of the International Measurement Confederation* 2020; 150: 107065, <https://doi.org/10.1016/j.measurement.2019.107065>.
  25. Zakharov A, Gerike B, Shiryamov D. The Ultimate Rotating Resistance of the Belt Conveyors Rollers 2018; 176: 227-230, <https://doi.org/10.2991/coal-18.2018.42>.
  26. Zimroz R, Hardygóra M, Blazej R. Maintenance of Belt Conveyor Systems in Poland – An Overview 2015: 21-30, [https://doi.org/10.1007/978-3-319-12301-1\\_3](https://doi.org/10.1007/978-3-319-12301-1_3).





Article citation info:

Ružinskas A, Giessler M, Gauterin F, Wiese K, Bogdevičius M. Experimental investigation of tire performance on slush. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 103–109, <http://dx.doi.org/10.17531/ein.2021.1.11>.

## Experimental investigation of tire performance on slush

Andrius Ružinskas<sup>a\*</sup>, Martin Giessler<sup>b</sup>, Frank Gauterin<sup>b</sup>, Klaus Wiese<sup>c</sup>, Marijonas Bogdevičius<sup>a</sup>

Indexed by:



Department of Mobile Machinery and Railway Transport, Vilnius Gediminas Technical University, Plytinės str., 27 10105 Vilnius, Lithuania  
Institute of Vehicle System Technology, Karlsruhe Institute of Technology, Kaiserstrasse., 2 76131 Karlsruhe, Germany  
Continental Reifen Deutschland GmbH, Jädekamp., 30 30419 Hannover, Germany

### Highlights

- This paper presents the experimental investigation of tire performance on slush.
- Slush is defined as a mixture of snow and water, and a mixture of crushed ice and water.
- Tire friction decreases increasing the mixture amount of snow and water.
- Tire friction increases increasing the mixture amount of crushed ice and water.
- The results of performed tire tests showed a good reproducibility and reliability.

### Abstract

An investigation of tires behaviour on winter roads was always a high importance in the context of road safety. This paper presents the experimental investigation of tire performance on slush that is identified as two mixtures: a mixture of snow and water, and a mixture of crushed ice and water. The measurements of longitudinal and lateral performance including tire traction, braking and cornering were performed. Tire traction tests were performed for both mixtures with different amount of material. A decreasing in the tire friction was observed when amount of the mixture of snow and water was increased twice. For the mixture of crushed ice and water, an opposite trend was observed. The standard deviation values for the peak force coefficient showed a good reproducibility and reliability of performed tire tests.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>) slush, tire, friction, longitudinal force, lateral force, force coefficient.

## 1. Introduction

The frictional behaviour of tires on any surface is important for vehicle safety and control [14, 24]. As tires are in direct contact with the pavement, the vehicle is controlled directly due to its tire friction with the pavement, and parameters of the tires have a great influence on controlling stability and ride safety of the vehicle [21]. The tire is the only component transmitting forces between the vehicle and the ground, thus it has to maintain high forces during all driving maneuvers [15]. Also, a knowledge of the current tire-road friction is essential for future autonomous vehicles [2]. In general, the tire performance is evaluated within longitudinal and lateral slip curves. The shape of the curve is influenced by many parameters, for example the friction characteristics of the tread rubber, which itself is influenced by the texture of the road surface, the medium between the road and the tire (e.g. water, snow, ice etc.) and the tire tread compounds [19].

In research papers a number of experimental studies of tire performance on snow and ice can be found. The effects of operational parameters, namely load on the tire, inflation pressure, toe angle, tread depth, camber angle, ice temperature, ambient temperature and type of ice surface using Standard Reference Test Tire were studied with indoor test method [3] and outdoor test method [4]. Cutini et al. [8] presented a method for testing winter tires in outdoor test facilities on compacted snow and on iced surface. Results showed that the method

was able to measure the small differences between the traction force of the different sets of tires. Hjort et al. [12] presented a comprehensive study of the performance of winter tires on snow, ice, and asphalt using a mobile tire-testing device for snow and asphalt, and using stationary tire-testing facility for ice. Both devices recorded the tire forces and motions, and results suggested that the recorded data represent real vehicle performance. Salimi et al. [20] investigated the friction behaviour of tire with ice and snow using RT3-Curve, a lateral friction measuring device. The study showed that ice had 55 % less friction compared to bare dry conditions. Light, moderate, and heavy snow reduced the dry surface friction significantly by 69, 75, and 81 %, respectively. Yokoyama et al. [23] studied the friction performance of tires on snow covered roads to predict vehicle performance within the grip range. The findings indicated that tire frictional force, as an aspect of vehicle dynamic performance on snow covered roads, can predict subjective evaluation results from actual vehicles by means of tire pattern, compound, and contact patch pressure distribution characteristics.

Unlike the tire performance on snow and ice, only a couple approaches can be found of investigation on tire performance on slush. According to Klein-Paste [17] slush mainly consists of water and ice particles and behaves like an incompressible fluid. The tire braking action on slush is mainly reduced because of hydroplaning. When a

(\*) Corresponding author.

E-mail addresses: A. Ružinskas - [andrius.ruzinskas@vilniustech.lt](mailto:andrius.ruzinskas@vilniustech.lt), M. Giessler - [martin.giessler@kit.edu](mailto:martin.giessler@kit.edu), F. Gauterin - [frank.gauterin@kit.edu](mailto:frank.gauterin@kit.edu), K. Wiese - [klaus.wiese@conti.de](mailto:klaus.wiese@conti.de), M. Bogdevičius - [marijonas.bogdevicius@vilniustech.lt](mailto:marijonas.bogdevicius@vilniustech.lt)

tire rolls on a pavement covered with an incompressible fluid, it is squeezing the fluid from under the footprint. This squeezing process generates pressures on the surface of the tire footprint, delaying or preventing it to get into contact with the underlying pavement. When full hydroplaning occurs it provides almost no braking action as the tire slides over the fluid film without reaching the pavement texture at all. Hydroplaning is largely speed dependent since the speed determines how much time is available for the squeeze out. However, it is also largely dependent on the fluid thickness.

Lee and Huang [13] used a physical tire-snow interaction model and presented results of calibration and validation of the interaction model for wet snow wherein the volume fraction of water content ranges from 3 % to 8 %. The main differences between dry and wet snow were noticed in a higher coefficient of friction for the rear wheel on dry snow than wet, and the traction force for wet snow was higher than for dry indicating lower motion resistance for wet snow. In Bhoopalam et al [5] study a slush with soil definition was used evaluating the tire traction performance on different icy roads. The study showed a 300 % increase in the peak traction during operation on ice covered with soil and slush (shaved ice and water mix), compared to wet ice (ice covered with a thin layer of liquid water). Bogdevičius et al [6] presented a theoretical approach of mechanism of force transmission between a tire and slush-covered pavement. The mathematical approach analysed the system “sub-block-slush layer-drum”, wherein the slush was described as a multi-layer bulk. The model evaluated mass change velocity of slush layer and physical-mechanical properties of sub-block. The obtained velocities of slush layers and friction forces enabled to determine generated heat per time unit at each layer. Results showed that the top layer of slush has the highest velocity and heat flow values.

This paper presents a comprehensive experimental study on tire performance on slush that is identified with two different mixtures. The experimental setup including test equipment, track preparation, test procedure and the tire friction values for longitudinal and lateral tire performance including tire traction, braking and cornering are presented and discussed.

## 2. Experimental setup

In the following chapter experimental setup including test equipment, track preparation and test procedure are presented. The Internal Drum Test Bench in Karlsruhe Institute of Technology was used to perform measurements of tire force transmission on the different amount and consistency of slush layer.

### 2.1. Test equipment

The Internal Drum Test Bench of Karlsruhe Institute of Technology, showed in Fig. 1 allows tire measurements in controllable winter conditions such as ice or snow. The test rig consists of the internal drum (4) with a diameter of 3.8 m, wherein the tire (2), mounted on a rigid wheel suspension, rolls on the installed track (5). Wheel and drum can be driven independently for braking and traction tests. Slip and camber angles, and vertical force are adjusted by hydraulic system. The main technical specifications of the test bench are presented in [11, 22]. All tire response data is measured inside the wheel hub by a six-component measurement hub (3) and the main outputs are: longitudinal force ( $F_x$ ), lateral force ( $F_y$ ) and vertical (wheel reaction) force ( $F_z$ ).

The test rig is surrounded by a climate chamber with an air conditioning system which allows cool down the testing room to  $-20^{\circ}\text{C}$ . For measurements on ice and snow tracks the test rig can be additionally equipped with a snow production machine, a roller for snow compaction, a blade (7) for cutting ice or snow layer and infrared camera for temperature measurements.

A special winter tire 205/55 R16 was used for the measurements. The tire footprint at about 4000 N vertical load and 220 kPa inflation pressure is showed in Fig. 2.

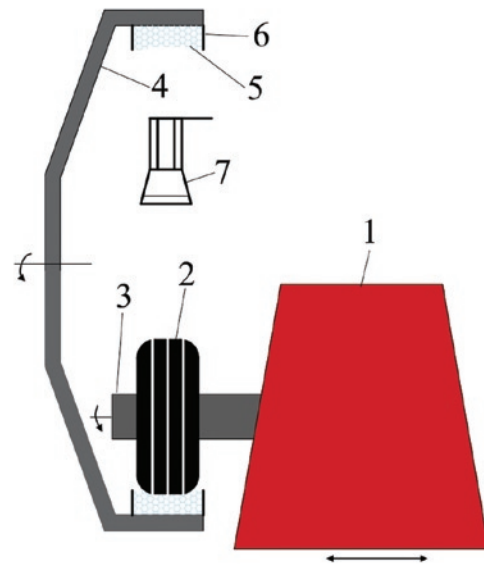


Fig. 1. Schematic view of the test bench. 1 – test stand; 2 – tested tire; 3 – measuring hub; 4 – internal drum; 5 – track; 6 – borders; 7 – hydraulic blade.

As it is seen from Fig. 2, the tire has a non-symmetrical tread pattern. Its depth was 8 mm and shore hardness was 56A. The tire was always kept in the testing room before the each measurement.



Fig. 2. The footprint of the tested tire

### 2.2. Slush track preparation

According to [10] slush is a mixture of water and snow, at the temperature of zero or a little above the freezing point. Slush behaves as a liquid in the sense that only insignificant shear forces may be transmitted. The density of slush depends on its water content and is always less than the density of water. The viscosity of slush is always larger than that of water, and always increases with decreasing water content. In The International Classification for Seasonal Snow on the Ground [9] and [18] slush is presented as a soaked snow which has a small air content and liquid water content is greater than 15 %. The temperature of snow is equal or higher than  $0^{\circ}\text{C}$ .

The main challenge preparing the slushy track was to keep the mixture on the drum as much as possible to get a good track reproducibility when tire is rolling. For this reason, the special borders (6), showed in Fig. 1, were installed not allowing to splash away the mixture from the drum when tire was rolling. The next challenge was to equalize

the slush mixture on the drum and don't let it to fall down when drum is rotating. After some attempts, it was decided to throw mixed material directly on the rotating drum expecting that centrifugal forces  $F_c$  showed in Fig. 3 will make it stick on the drum surface and prevents from sliding. For this reason, a required drum speed was calculated for different angle situations.

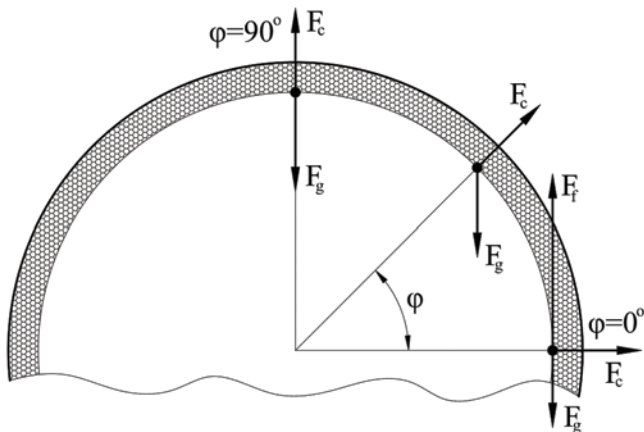


Fig. 3. Forces acting on the slush at different angle when drum is rotating

When slush mixture is at  $90^\circ$  angle, the centrifugal  $F_c$  and gravity  $F_g$  forces are equalized:

$$mg = \frac{mV_D^2}{R} \quad (1)$$

where  $V_D$  refers to drum speed and  $R$  refers to drum radius. Then drum speed can be expressed:

$$V_D = \sqrt{gR} \quad (2)$$

A force equilibrium can be expressed accordingly to the angle  $0^\circ < \varphi < 90^\circ$  and drum speed is calculated:

$$mg = \frac{mV_D^2}{R} \sin \varphi \quad (3)$$

$$V_D = \sqrt{\frac{gR}{\sin \varphi}} \quad (4)$$

At  $0^\circ$  a cohesion force occurs, thus equalizing total friction force  $F_f$  with gravity force  $F_g$ , we obtain:

$$\frac{mV_D^2}{R} \mu + F_c = mg \quad (5)$$

Then from equation (5) friction coefficient  $\mu$  and drum speed  $V_D$  are expressed:

$$\mu = \frac{R}{mV_D^2} (mg - F_c) \quad (6)$$

$$V_D = \sqrt{\frac{R}{\mu m} (mg - F_c)} \quad (7)$$

Assuming that cohesion force  $F_c$  which should increase the friction between the mixture and the drum surface is neglected, the minimum required speed of the drum is about 55 km/h (or about 277 rpm). The calculated friction value was 0.07 wherein  $F_c$  was also neglected.

It was calculated that for 2 cm of height 62 l of mixture are required. According to the calculation, the ratios of the mixture were selected.

In this study slush was identified as a mixture of snow and water, and a mixture of crushed ice and water. According to the slush definition presented in [9], the mixture should have more than 15 % of water. Therefore, for the mixture of snow and water the ratios of 5.5:1 (55 l of snow and 10 l of water) and 11:2.2 were selected. For the mixture of crushed ice and water the ratios of 6:0.75 (60 l of crushed ice and 7.5 l of water) and 11:1.5 were selected.

For the mixture of snow and water, a snow was produced by cutting ice with the blade (7) showed in Fig. 1. The ice was formed on the drum by pouring the water at the temperature below  $0^\circ\text{C}$ . Usually, for the tire-snow measurements snow is produced by snow production machine that is using a liquid nitrogen. In our case, the presented snow production method appeared to be more cost effective and allowed to get a loose snow of about  $400 \text{ kg/m}^3$  density. Further snow is mixed up with the water and density of the mixture is determined. An average density value for the snow and water mixture was  $815 \text{ kg/m}^3$  with the standard deviation of  $56 \text{ kg/m}^3$ . Then prepared mixture (Fig. 4a) is thrown on the rotating drum at already calculated speed (55 km/h). After stopping the drum (Fig. 4c), an initial height of the mixture is measured that is about 2 cm or 4 cm according to the ratio of the mixture.

The mixture of crushed ice and water (Fig. 4b) is prepared accordingly to the mixture of snow and water. Simple ice cubes that are crushed into small pieces (ice particles) and mixed up with the water were used. An average density of the mixture was about  $674 \text{ kg/m}^3$  with the standard deviation of  $39 \text{ kg/m}^3$ . The prepared mixture is thrown on the rotating drum and the same initial height of 2 cm and 4 cm was obtained according to the ratio of the mixture.

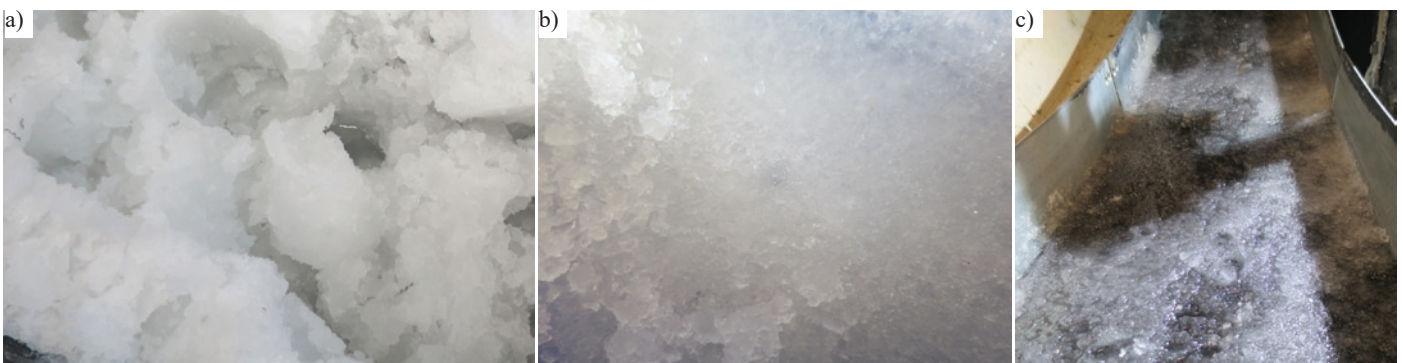


Fig. 4. Slush production. a) – mixture of snow and water. b) – mixture of crushed ice and water. c) – mixture of snow and water on the drum

### 2.3. Test procedure

Once the track is prepared, the tire is installed on the test bench and measurements could be performed recording the force and moment data. The longitudinal tire slip  $S_x$  is calculated by:

$$S_x = \frac{\Omega R_{dyn} - V_D}{V_D} \quad (8)$$

where  $\Omega$  refers to the angular velocity of the tire,  $R_{dyn}$  refers to dynamic rolling radius of the tire that is determined while the wheel is freely rotating before the measurement starts. Positive values represent traction and negative values braking slip.

The values of test parameters are presented in Table 1. For every test the ambient temperature in the cooling chamber was kept about 0-1°C. The tire was rolling on the drum loaded with 4000 N, inflated with the pressure of 220 kPa and with 0° camber angle. The driving speed (the drum speed) was selected about 50 km/h constant speed. Then the tire is accelerated or decelerated with in the test bench. The cornering tests were performed as pure cornering, i.e. with a very low slip. Up to 80 % slip is reached when tire is accelerated and about 70 % slip when it is braked. The tire turning (slip) angle has ranged from -4° to 4° and was limited because of installed borders on the drum.

Table 1. The values of test parameters

Test parameter	Value
Ambient temperature, °C	0-1
Vertical load, N	4000
Tire inflation pressure, kPa	220
Tire camber angle, °	0
Driving speed, km/h	50
Slip ratio, %	-70 - 80
Slip angle,	-4 - 4

Three types of measurements were performed: traction, braking and cornering. Traction tests were performed for both mixtures with different ratios, and for braking and cornering only with the mixture of snow and water with the ratio of 5.5:1. For every test condition at least three measurements were performed wherein after two measurements the track was renewed with the hydraulic blade showed in Fig. 1 equalizing the layer of the mixture and then one more measurement was performed. For the next test the track was renewed with the completely new mixture. The one traction or braking measurement lasted up to 15 s and cornering measurement lasted up to 35 s since a special test protocol was used specially designed for the snow track.

The results are presented as slip curves wherein a longitudinal force coefficient (the ratio of longitudinal  $F_x$  and vertical  $F_z$  forces) is the function of slip, and a lateral force coefficient (the ratio of lateral  $F_y$  and vertical  $F_z$  forces) is the function of slip angle. For longitudinal slip curves, a Magic Formula (MF) [16] was used to fit the data using a least square minimization [1]. The standard deviation was calculated for the peak values of force coefficient. For the longitudinal tire performance, the peak values were obtained from slip curves fitted with MF.

### 3. Results

The average slip curves for tire traction tests for different mixtures and ratios are presented in Fig. 5 and Fig. 6. Also the views of the track after the measurements are presented to get a better insight of tire performance.

A peak value of force coefficient for the mixture of snow and water with the ratio 5.5:1 is 0.3 and is reached at about 20 % slip. Increasing the amount of mixture twice decreases the force coefficient and the peak value is 0.27 that is reached at the same slip. The standard deviation for both peak values was 0.031 and 0.026 respectively, and each value was calculated from three measurements.

A peak value of force coefficient for the mixture of crushed ice and water with the ratio 6:0.75 is 0.25 and is reached at about 15 % slip. Increasing the amount of mixture twice an opposite trend was observed. The force coefficient is increasing and the peak value is 0.31 reached at about 10 % slip. The standard deviation for both peak values was 0.016 and 0.031 respectively, and was calculated from three

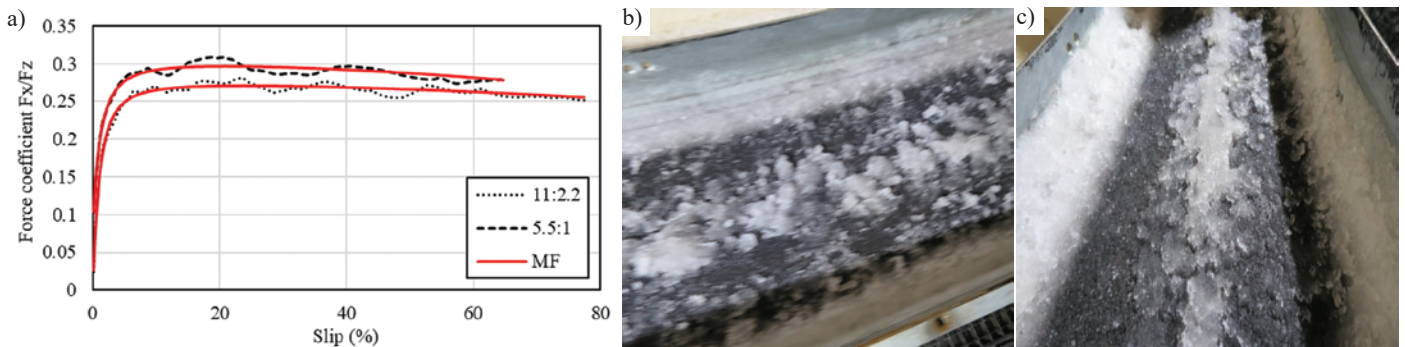


Fig. 5. Results of traction test for the mixture of snow and water a) Averaged slip curves fitted with Magic Formula for different mixture ratio b) Track after measurements for the mixture ratio 5.5:1. c) Track after measurements for the mixture ratio 11:2.2.

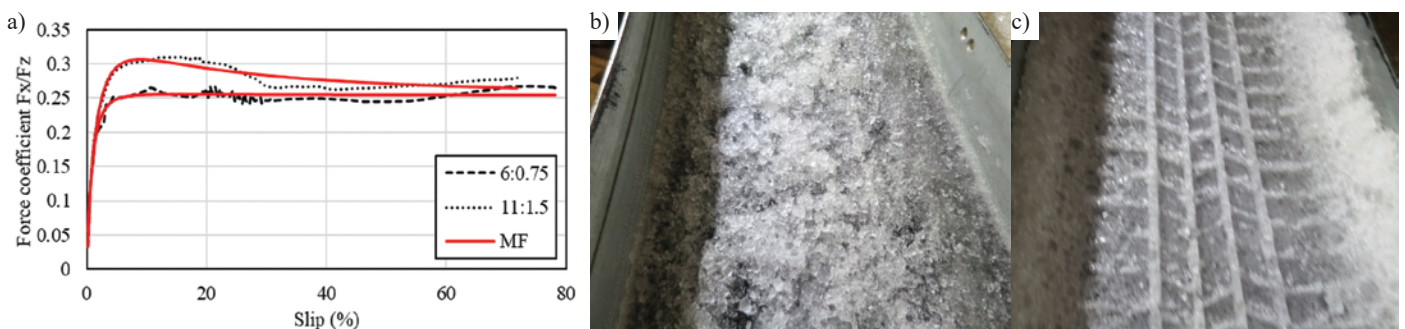


Fig. 6. Results of traction tests for the mixture of crushed ice and water. a) Averaged slip curves fitted with Magic Formula for different mixture ratio. b) Track after measurements for the mixture ratio 6:0.75. c) Track after measurements for the mixture ratio 11:1.5.

measurements for the mixture ratio 6:0.75, and from four measurements for the mixture ratio 11:1.5.

An average slip curve with Magic Formula for the tire braking test is presented in Fig. 7.

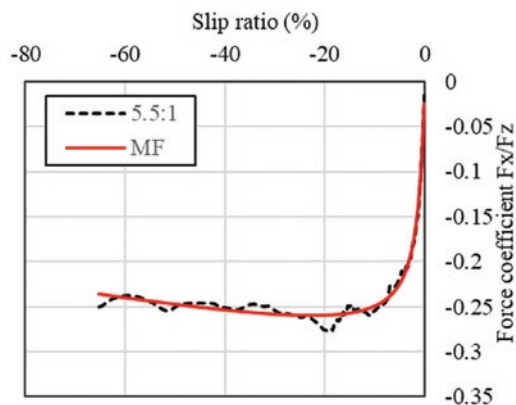


Fig. 7. Averaged braking slip curve fitted with Magic Formula for the mixture of snow and water with the ratio 5.5:1.

A peak value of the braking force coefficient for the mixture of snow and water is equal to 0.26 and is reached at about 20% slip, the same value as for traction test. The peak value of braking is lower than the peak value of traction (0.30) with the same mixture ratio. The standard deviation of 0.02 for the peak value was calculated from three measurements.

An average slip curve for the tire cornering test is presented in Fig. 8.

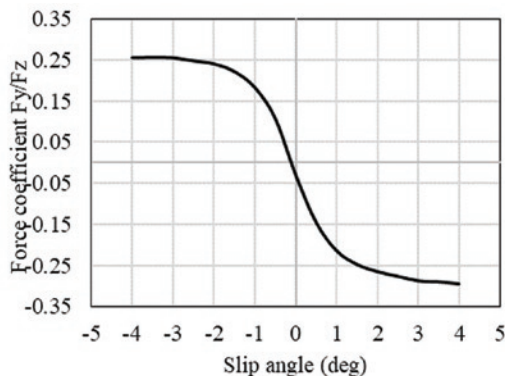


Fig. 8. Averaged cornering slip curve for the mixture of snow and water with the ratio 5.5:1.

For the cornering, the peak values of force coefficient at negative and positive angles aren't equal. At negative slip angle the peak value is 0.26 and for the positive angle the value is higher, and is equal to 0.3. Both values are reached at about 4° of slip angle. The standard deviation for both peak values is less than 0.01, that was calculated from three measurements.

#### 4. Discussion

In this chapter the obtained results are explained and for this reason the schematic view of the tire interaction with slush mixture is presented in Fig. 9.

In experimental studies, when a tire interacts with the mixture of snow and water, three zones appear in the contact area wherein different forces act on the tire surface (Fig 9a).

In the first zone, the tire surface is subjected to a high hydrodynamic pressure. The magnitude of the pressure depends on the thickness and rate of change of the slush layer, slush density and tire speed in the horizontal and vertical directions.

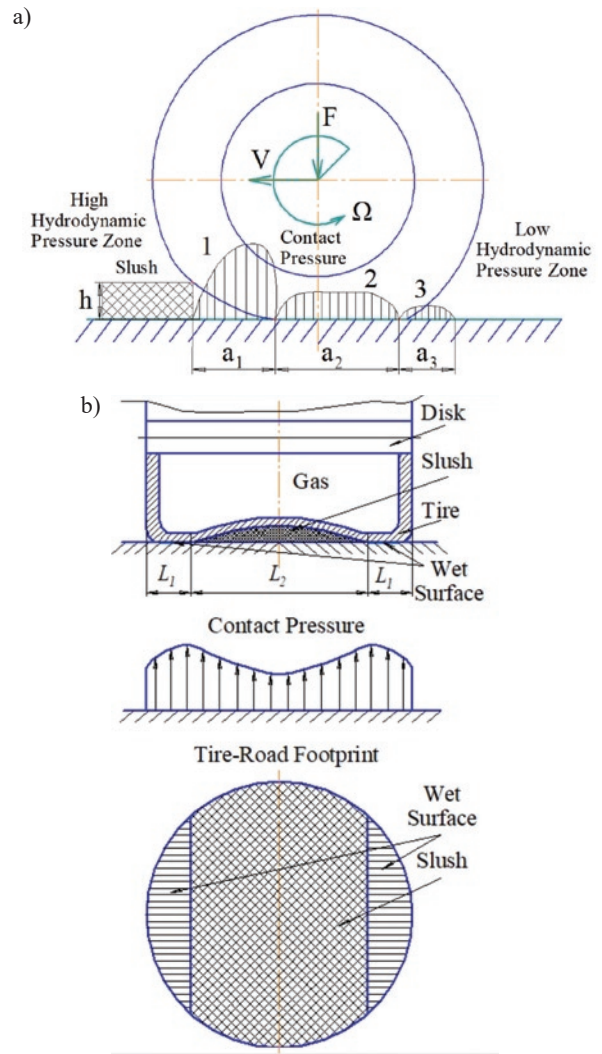


Fig. 9. Schematic view of the tire interacting with slush mixture. (a) Contact zones. (b) Pressure distribution between tire and surface

The viscosity of slush (fluidized snow) is about 50 times higher than that of water ( $94.10 / 1.7914 = 52.53 \text{ Pa}\cdot\text{s}$ ). The viscosity of slush depends on the amount of snow and water. The higher the amount of water, the closer the viscosity of the slush gets to the viscosity of the water. Therefore, when the tire interacts with slush, the hydrodynamic pressure is higher than the tire interacting with water.

In the second zone, the tire has contact with the road surface (the drum) and slush layer. Due to the lower radial stiffness of the tire, the tire deforms more at the centre of contact, resulting in lower slush pressure at the centre of contact than it is at the tire edges. When the tire presses on the slush layer, the speed of slush outflow will also change. Because the tire edges are less deformed, in this area the tire squeezes the slush out faster and mechanical contact between the tire and the drum occurs. The slush becomes closed at the tire contact centre, i.e. the slush is unable to escape from the contact area and this zone concentrates the increased amount of slush (Fig. 5b and c).

As a result, the tangential stresses (frictional forces) at the contact between the tire and the slush will be lower when the amount of water in the slush is higher. This explains the lower tire friction with higher amount of mixture in Fig. 5a.

When the tire presses on the mixture of crushed ice and water, the space between the ice particles is reduced (air is squeezed) and the mixture becomes significantly stiffer. With each revolution of the wheel, the mixture in the rotating drum is densified. The vertical force and centrifugal forces acting on the wheel cannot disturb the structure of such a mixture, and after several revolutions of the drum, the physical and mechanical properties of the layer become

close to those of the ice cover. In such a mixture, a small amount of water (7.5 l) fills the gaps between ice and snow particles, and the mixture becomes more compressible, but the applied forces cannot destroy the structure of the mixture. The wheel is rolling on the rigid surface of the mixture. Increasing the amount of water in the mixture of ratio 11:1.5 increases the space between the hard pieces of ice, making the mixture less stiff.

The forces acting on the mixture (wheel contact force and centrifugal forces) disrupt the structure of the mixture and the tire tread digs into the mixture layer. As a result, the contact area of the tire increases with the frictional forces. The higher water content in the mixture increases the tire friction. This explains the higher values of force coefficient obtained with the higher amount of mixture in Fig. 6a.

The lower friction for braking (Fig. 7) compared with the traction is explained as follows: when tire is accelerated, the pressure in the first zone (Fig. 9a.) increases and when it is braked, the tire angular velocity  $\Omega$  decreases and the pressure in the first zone is lower as similar to aquaplaning [7]. The wheel reaction  $F_z$  and longitudinal  $F_x$  forces also decrease, but due to increased area of slush (Fig. 9b.) the longitudinal force  $F_x$  is decreasing at higher rate than reaction force.

The difference in lateral tire friction at positive and negative slip angles (Fig. 8) could be led by the non-symmetrical tire tread pattern (Fig. 2).

A good repeatability and reliability were observed with the peak values of force coefficient for all performed tests. The values of standard deviation ranged from 0.01 to 0.031 indicating that the peak values of force coefficient lied in a narrow band.

## 5. Conclusion

In this study, an experimental investigation of tire performance on slush is proposed. The experimental tests were performed with Internal Drum Test Bench wherein the tire is able to run on the pre-in-

stalled road track. The slush was identified as two mixtures: a mixture of snow and water, and a mixture of crushed ice and water. Based on slush definition presented in [9] the mixture ratios were selected. For the mixture of snow and water, it was 5.5:1, i.e. 55 l of snow and 10 l of water, and 11:2.2 accordingly. For the mixture of crushed ice and water, the ratios were as follows: 6:0.75 and 11:1.5. The required drum speed, at which should be thrown a mixture on the drum to get equally distributed layer, was calculated from the forces acting on slush in rotating drum. Tire traction tests were performed for both mixtures. A decreasing in the tire friction was observed when amount of the mixture of snow and water was increased twice. The peak value of force coefficient decreased from 0.3 to 0.27. For the mixture of crushed ice and water, an opposite trend was observed. Increasing the amount of mixture increased tire friction. The peak value of force coefficient increased from 0.25 to 0.31. The tire braking and cornering tests were performed with the mixture of snow and water with the ratio 5.5:1. It was observed that tire has lower friction for braking compared with the traction for the same ratio of mixture. For the tire cornering, a lower tire friction was observed for the negative slip angle compared with positive angle. The low standard deviation values for the peak force coefficient of all performed tests showed a good reproducibility and reliability of presented test method.

## Acknowledgement

*The research work has been supported by the Research Council of Lithuania Fund within the project "Investigation of tribological micro-macro processes in multi-phase layer with ice-snow particles and their contribution to tire-layer-surface system", project code P-MIP-17-233.*

## References

1. Alagappan V A, Rao N V K, Kumar K R. A comparison of various algorithms to extract Magic Formula tyre model coefficients for vehicle dynamics simulations. *Vehicle System Dynamics* 2015; 53(2): 154-178, <http://dx.doi.org/10.1080/00423114.2014.984727>
2. Albinsson A, Bruzelius F, Jacobson B, Fredriksson J. Design of tyre force excitation for tyre-road friction estimation. *Vehicle System Dynamics* 2017; 55(2): 208-230, <http://dx.doi.org/10.1080/00423114.2016.1251598>
3. Bhoopalam K A, Sandu C, Taheri S. Experimental investigation of pneumatic tire performance on ice: Part I – Indoor study. *Journal of Terramechanics* 2015; 60: 43-54, <http://dx.doi.org/10.1016/j.jterra.2015.02.006>
4. Bhoopalam K A, Sandu C, Taheri S. Experimental investigation of pneumatic tire performance on ice: Part II – Outdoor study. *Journal of Terramechanics* 2015; 60: 45-62, <http://dx.doi.org/10.1016/j.jterra.2015.03.001>
5. Bhoopalam K A, Sandu C, Taheri S. Tire traction of commercial vehicles on icy roads. *SAE International Journal of Commercial Vehicles* 2014; 7(2): 357-365, <http://dx.doi.org/10.4271/2014-01-2292>
6. Bogdevičius M, Ružinskas A, Vadlūga V, Bogdevičius P, Kačianauskas R, Maknickas A, Gauterin F. Investigation of tire force transmission on interaction with slush. *Problemy Transportu – Transport Problems* 2019; 14(1): 13-21, <http://dx.doi.org/10.21307/tp.2019.14.1.2>
7. Cho R J, Lee W H, Sohn S J, Kim J G, Woo S J. Numerical investigation of hydroplaning characteristics of three-dimensional patterned tire. *European Journal of Mechanics A/Solids*, 2006; 25: 914-926, <https://doi.org/10.1016/j.euromechsol.2006.02.007>
8. Cutini M, Brambilla M, Toscano P, Bisaglia C, Abbati G, Meloro G. Evaluation of drwabar performance of winter tyres for special purpose vehicles. *Journal of Terramechanics* 2020; 87: 29-36, <https://doi.org/10.1016/j.jterra.2019.10.002>
9. Fierz C, Armstrong R L, Durand Y, Etchevers P, Green E, McClung D M, Nishimura K, Satyawali P K, Sokratov S A. The International Classification for Seasonal Snow on the Ground. IHP-VII Technical Documents in Hydrology Nr. 83, 2009.
10. Fristedt K, Norrbom B. Studies of contaminated runways. The Aeronautical Research Institute of Sweden, 1980.
11. Giessler M, Gauterin F, Wiese K, Wies B. Influence of friction heat on tire traction on ice and snow. *Tire Science and Technology* 2010; 38(1): 4-23, <http://dx.doi.org/10.2346/1.3298679>
12. Hjort M, Eriksson O, Bruzelius F. Comprehensive study of the performance of winter tires on ice, snow, and asphalt roads: The influence of tire type and wear. *Tire Science and Technology* 2017; 45(3): 175-199, <https://doi.org/10.2346/tire.17.450304>
13. Lee H J, Huang D. Vehicle-wet snow interaction: testing, modeling and validation. *Journal of Terramechanics* 2016; 67: 37-51, <http://dx.doi.org/10.1016/j.jterra.2016.08.001>
14. Levulytė L, Žuraulis V, Sokolovskij E. The research of dynamic characteristics of a vehicle driving over road roughness. *Eksplloatacija i Niezawodnosc – Maintenance and Reliability* 2014; 16(4): 518-525
15. Linke T, Wiese K, Wangenheim M, Wies B, Wallashek J. Investigation of snow milling mechanics to optimize winter tire traction. *Tire Science and Technology* 2017; 45(3): 162-174, <https://doi.org/10.2346/tire.17.450302>

16. Pacejka, B H. *Tire and vehicle dynamics*. Amsterdam: Elsevier, 2012.
17. Paste A K. Airplane braking friction on dry snow, wet snow or slush contaminated runways. *Cold Regions Science and Technology*, 2018; 150: 70-74, <http://dx.doi.org/10.1016/j.coldregions.2017.02.004>
18. Piskova M, Jarausch S, Kopf Ch, Novak P, Trappmann D. An introduction to the basic properties and mechanics of snow. *Snow and Avalanches*, Winter 2008; 09: 1-22.
19. Riehm P, Unrau H J, Gauterin, Torbrügge S, Wies B. 3D brush model to predict longitudinal tyre characteristics. *Vehicle System Dynamics* 2019; 57(1): 17-43, <https://doi.org/10.1080/00423114.2018.1447135>
20. Salimi S, Nassiri S, Bayat A, Halliday D. Lateral coefficient of friction for characterizing winter road conditions. *Canadian Journal of Civil Engineering* 2016; 43(1): 73-83, <http://dx.doi.org/10.1139/cjce-2015-0222>
21. Sapragonas J, Keršys A, Makaras R, Lukoševičius V, Juodvalkis D. Research of the influence of tire hydroplaning on directional stability of vehicle. *Transport* 2014; 28(4): 374-380, <http://dx.doi.org/10.3846/16484142.2013.865673>
22. Wassertheurer B, Gauterin F. Investigation of winter tire characteristics on different track surfaces using a statistical approach, *Tire Science and Technology* 2015; 43(3): 195-215, <https://doi.org/10.2346/tire.15.430304>
23. Yokoyama T, Hiratsuka K, Notoni Sh. A study of tire characteristics and vehicle performance on snow-covered roads. *SAE Technical Paper* 2015-01-1522, 2015; 1-9, <https://doi.org/10.4271/2015-01-1522>
24. Žuraulis V, Garbinčius G, Skačkauskas P, Prentkovskis, O. Experimental study of winter tyre usage according to tread depth and temperature in vehicle braking performance. *Iranian Journal of Science and Technology, Transactions of Mechanical Engineering* 2020; 44: 83-91, <https://doi.org/10.1007/s40997-018-0243-0>

## Alarms management by supervisory control and data acquisition system for wind turbines

Indexed by:



Isaac Segovia Ramirez<sup>a</sup>, Behnam Mohammadi-Ivatloo<sup>b,c</sup>, Fausto Pedro García Márquez<sup>a\*</sup>

<sup>a</sup>Ingenium Research Group, Universidad Castilla-La Mancha, 13071 Ciudad Real, Spain

<sup>b</sup>Faculty of Electrical and Computer Engineering, University of Tabriz, 5166616471 Tabriz, Iran

<sup>c</sup>Department of Energy Technology, Aalborg University, 9220 Aalborg East, Denmark

### Highlights

- We propose a new approach for signal processing, fault detection and diagnosis.
- A New approach is based on principal component analysis and artificial neural networks
- We analyse the signals and detect the alarm activation pattern.
- The dataset has been reduced by 93%
- The performance of the neural network is incremented by 1000%.

### Abstract

Wind energy is one of the most relevant renewable energy. A proper wind turbine maintenance management is required to ensure continuous operation and optimized maintenance costs. Larger wind turbines are being installed and they require new monitoring systems to ensure optimization, reliability and availability. Advanced analytics are employed to analyze the data and reduce false alarms, avoiding unplanned downtimes and increasing costs. Supervisory control and data acquisition system determines the condition of the wind turbine providing large dataset with different signals and alarms. This paper presents a new approach combining statistical analysis and advanced algorithm for signal processing, fault detection and diagnosis. Principal component analysis and artificial neural networks are employed to evaluate the signals and detect the alarm activation pattern. The dataset has been reduced by 93% and the performance of the neural network is incremented by 1000% in comparison with the performance of original dataset without filtering process.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

alarm management, maintenance management, principal component analysis, SCADA, wind turbines.

## 1. Introduction and Objectives

The wind energy capacity in the world energy production is increasing every year, being one of the fastest growing renewable energy. This growth in recent years is due to the increasing size, increment of complexity of wind turbines (WTs) and favourable policies adopted by governments. The wind energy new installations exceeded the 60,4 GW at the end of 2019, with the United States and China as the most relevant markets [19]. The impact of COVID-19 pandemic on this growth has not been quantified yet and, therefore, it is required technical and economic advances to ensure the reliability of this technology.

WTs are complex electromechanical systems formed by a rotor that transforms the wind energy into mechanical energy and it is converted into electrical power. The blades transmit the mechanical energy through the low speed shaft to the high-speed shaft connected to the generator. The yaw system rotates the nacelle to align the blades with the direction of the wind and the gearbox regulates the speed. Different subsystems, e.g., meteorological units, refrigeration, brakes, security levels, etc., increase the complexity of the WTs [30]. The working conditions of WTs produce a variety of potential component failures with high failure rates and economic losses. Different researches reported that the gearbox, yaw and hydraulic system, electrical control

and blades, concentrate the 60% of the total failures [11], being necessary the application of novel techniques or methodologies [10].

The objective of the maintenance management is to ensure an accurate behaviour of WTs, minimizing the use of human and material resources, reducing reduced costs and avoiding energy production losses due to downtimes [35]. The operation and maintenance (O&M) costs are estimated between 15-25% in onshore [27], and they are higher in offshore WTs [32]. The increment of faults reduces the availability of energy generation due to downtimes and unplanned maintenance activities [31]. New improvements in maintenance management are needed through novel conditions monitoring systems (CMS) and data analysis to ensure proper levels of reliability, availability, maintainability and safety (RAMS) [9,20].

Several data acquisition systems and CMS are installed to determine the condition of WTs. The measurement techniques are based on traditional monitoring, e.g., vibration analysis with low-frequency ranges for fault detection [36], thermal analysis for failure of electric elements, ultrasonic waves or acoustic emissions generated with energy pulses to detect blade failures [14, 16], among others. Catelani et al. [5] proposes a set of techniques combined in a single system for advance detection and the identification of the anomalies and failures in WTs, based on data processing. The results lead to inform with enough time to make the appropriate decision to maintain their WTs,

(\*) Corresponding author.

E-mail addresses: I. Segovia Ramirez - [isaac.segovia@uclm.es](mailto:isaac.segovia@uclm.es), B. Mohammadi-Ivatloo - [bmohammadi@tabrizu.ac.ir](mailto:bmohammadi@tabrizu.ac.ir), F.P. García Márquez - [faustopedro.garcia@uclm.es](mailto:faustopedro.garcia@uclm.es)



similar to [3]. Despite these techniques, new inspection techniques are being developed for early fault detection, e.g., noise waves analysis. Supervisory control and data acquisition (SCADA) system is widely implanted in the wind energy industry and this system collects main parameters of any WT from sensors and measurement systems.

The SCADA system provides alarm records and signal dataset, usually every ten minutes. The signals are numerical values about certain parameters, such as temperature, energy production or vibration. The alarm is defined as an operational indicator to identify different anomalies or issues that has already done in the WT. Faults in the sensors, inaccurate design of the diagnostic model or measurement uncertainty may lead to false alarms [6]. The false alarms identification is a critical issue in the wind energy market, since it is produced elevated alarm flood, unnecessary stops and maintenance interventions [28]. Several researches are focused on false alarm determination to reduce or reset the maximum alarms as possible, employing statistical analysis, deep learning, machine learning or neural networks, etc. The SCADA data volume and complexity increment the difficulty in the data analysis and the computational cost is elevated. Reliable and robust algorithms are required for false alarm detection and signal analysis. Fault tree analysis together with binary decision diagrams are employed for quantitative and qualitative analysis of these problems [13,23]. Machine learning algorithms are largely applied to evaluate the state of the WT and process a large amount of data. Artificial neural networks (ANNs) are based on the biological nerve system formed by neurons. This computational structure is formed by several layers with different weights and transfer functions for the connection between neurons. They can work with non-linear problems and the training phase defines the learning of the network [21]. ANNs have been used in several applications, highlighting pattern recognition and image processing with high accuracy [33]. ANNs are used in wind energy maintenance for forecasting and prediction, design, control optimization and fault detection and diagnosis (FDD) [18]. Gomes and Castro [15] employed the autoregressive moving average and ANNs to develop a wind energy forecasting with better results than the reference model. Li et al [22] probed that the type of ANN has an relevant influence in the analysis and different ANNs may present better capabilities in each case. Other authors have developed different researches about the false alarm identification. Adouni et al. [2] used a new ANN to increase the FDD procedure and the robustness to avoid false alarms. Bangalore et al. [4] developed filtering methods to ensure a strict training to improve the fault detection with the ANN. The advantages of the methodology are tested in different case studies. Peco et al. [7] developed a novel approach to detect false alarms in gearbox bearing with a partitioning methodology. None of these researches consider a preselection of the signals and data using statistical methods to reduce the computational cost and the accuracy results of the ANN.

This paper presents a novel approach in alarm and signal filtering to increase the accuracy in fault detection. The alarms are analyzed by Pareto chart to select the most critical, stabilizing interest periods of study. The signal dataset is reduced using principal component analysis (PCA), decreasing the amount of data an incrementing the accuracy of ANN. This approach is validated with a real case study based on different alarms and signals acquired from a WT.

The paper is divided in the following sections: Section II defines the overview and the methodology based on different phases and the information about the algorithms of the approach, a case study is proposed in Section III to test the approach; Section IV validates the results obtained with neural network and finally, Section V summarized the main findings from this research work.

## 2. Method

This work proposes a data filtering process in alarms and signals to increase the reliability of the ANN in fault detection due to the volume of the data and the number of alarms. Figure 1 shows flowchart of the proposed approach.

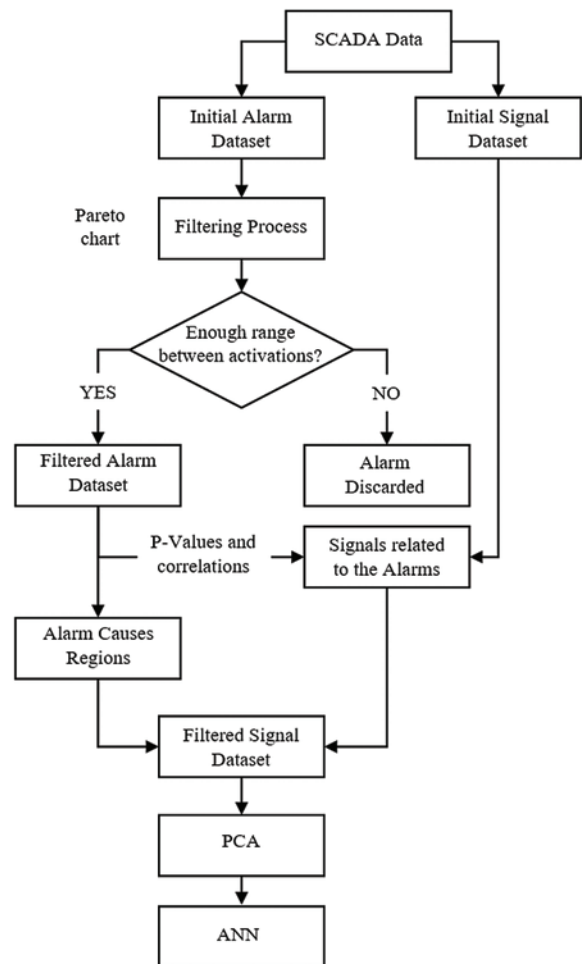


Fig. 1. Approach diagram

The SCADA data is divided into alarms and signals. The alarms are filtered analysing its main parameters. The critical alarms are identified with Pareto chart, considering its main characteristics, e.g., number of activations. Pareto chart shows the frequencies of the selected variables and, in this paper, is employed for determining the critical alarm. The user may introduce different weights regarding on the needs of the system. The main characteristics are the number of activations, time of activation and difference between alarms. For this work, the alarm must present several activations with elevated time periods with no failures in order to increase the reliability of the method. Schlechtingen and Ferreira [34] considered the alarm analysis for prediction one day before the failure. Reduced periods with less than one day between alarm activation are discarded, ensuring suitable range periods for the analysis. It is necessary to obtain the dataset of signals related to the critical alarm determined in previous phase. The signals with no correlation with the alarm increase the complexity of the operation and do not provide valuable information. Once the alarms and signals are properly determined, the regions of interest for the study are determined analysing the region with alarm deactivation. The data closer to the alarm activation is considered more relevant to the alarm since it is more probable to find data patterns that generate the activation of the alarm in that data range. A diagram of alarm and signal ranges related to the temporal scale is shown in Figure 2. The region of interest is defined in the graph as alarm causes range. The other region, registered as safe period, presents data related to the alarm activation but, in this period, the probability of alarm activations is low, being considered as training zone for the next phases.

Despite the filtering process, the signal dataset may present large volume of data. PCA has been employed in this paper to reduce this data volume. PCA is a multivariate analysis that decreases the dimen-



Fig. 2. Alarm and signal periods

sions of the initial dataset to a reduced number of uncorrelated variables or principal components [12,26]. The patterns in multivariate datasets may be extracted by PCA and the transformed space is defined by the eigenvalues and eigenvectors of the covariance matrix.

### 2.1. P-values

The P-value contributes with statistical measurements about the relation between signals, being possible the identification of more relevant signals [24]. The P-values are defined as the maximum probability under the null hypothesis, providing information about the combination of the data. In first place, it is necessary to calculate the distribution of the test statistic  $T$  by equation (1):

$$T = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad (1)$$

being  $\hat{p}$  the sample proportion;  $p_0$  the population proportion in the null hypothesis, and;  $n$  the sample size. The P-value is defined by equation (2):

$$P - value = \text{Probability}(T \geq ts | \text{Hypothesis is true}) = cdf(ts) \quad (2)$$

where  $ts$  the observed value of  $T$ , and;  $cdf$  the cumulative distribution function of the test static.

For this work, P-values of 5% are employed to confirm that the model is suitable for this dataset, because of the common criteria in several researches is that p-value lower than 0,05 is needed [8].

### 2.2. Principal Component Analysis

Large datasets may have correlated variables with redundant information that increase the complexity of the data analysis. PCA is a dimensionality-reduction tool to reduce and simplify the dimensionality of large datasets, developing an orthogonal linear transformation with the original data, to obtain a maximum variance in the new dataset [25]. The initial matrix data  $\mathbf{X}$  is defined with  $n \times p$  dimensions, being  $n$  samples and  $p$  variables, and the objective is the obtention of a reduced dataset [17]. The principal components  $Y_j$  are new uncorrelated variables, defined by (3):

$$Y_j = a_{j1}X_1 + a_{j2}X_2 + a_{j3}X_3 + \dots + a_{jp}X_p \quad (3)$$

being  $a_{j1}, a_{j2}, \dots, a_{jp}$ , constants to develop the linear correlation. The sum of the square of these constants is usually 1. The first principal component is defined to take the greatest variance in the data set. The transformed space is given by the eigenvectors of the covariance matrix  $\mathbf{S}$ , defined by equation (4):

$$\mathbf{S} = \frac{\mathbf{X}^T \mathbf{X}}{n-1} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (4)$$

being  $\mathbf{\Lambda}$  the diagonal matrix with real eigenvalues and  $\mathbf{U}$  the matrix with eigenvectors in columns. The correlation between variable and principal component  $Y_j$  s given by equation 5:

$$r_{ij} = \sqrt{\frac{a_{ij}^2 \text{Var}(Y_j)}{s_{ii}}} \quad (5)$$

being  $s_{ii}$ , the variance-covariance matrix of the original data. For this work, the PCA definition depends on the dataset filtered in previous phases of the approach.

### 2.3. Artificial Neural Network

The ANN is formed by large number of interconnected neurons with the capacity of process information [1]. The ANN may be classified regarding on the presentation of the information, the topology of the network, the relation between input and out and the training method (unsupervised and supervised). In Figure 3, it is observed the two types of ANN employed in this work. Figure 3 shows a multilayer perceptron (MLP) structure of ANN architecture, with the neurons organized in different layers.

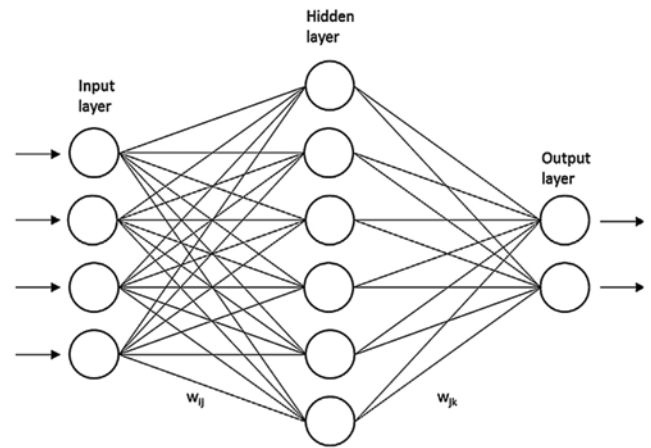


Fig. 3. ANN structure [29]

ANN requires large datasets and training process to design the interconnection weights. The connections between neurons are determined with number reference or weight. Equation (6) determines the output  $h_i$  of neuron  $i$ :

$$h_i = \sigma \left( \sum_{j=1}^N W_{ij} x_j + T_i^{hidden} \right) \quad (6)$$

being  $\sigma()$  the transfer function;  $N$  the number of input neurons;  $W_{ij}$  the weights;  $x_j$  inputs to the input neurons, and;  $T_i^{hidden}$  the threshold of the hidden neurons.

This paper considers a MLP ANN since the objective of this work is the definition of the approach. The simulations show that the best configuration is an ANN network architecture with 100 hidden layers. The inputs are variables, regarding on the filtering phase developed in previous phase. The validation and test phase of the ANN is developed dividing the dataset into 65% for the training, 20% for validation and 15% for testing. The adjustment of the network is obtained with the training phase.

### 3. Case study and results

A real database from the OPTIMUS European Project has been employed [60]. The database is composed of a set of parameter measurements and an alarm report. The SCADA generates an alarm when a set of variables reach a certain threshold. The SCADA system shows a sampling rate of 10 minutes.

The SCADA system also provides a detailed alarm report. This information will be employed for validating the model proposed. The collected data are:

- Date of the alarms.
- Code of the alarm (confidential).
- Cause or description of the alarm.
- The state of the alarm (activation or deactivation).
- The severity.
- Required tasks.

A real case study with SCADA data from one WT for 400 days and a data frequency of one minute, is employed to validate the approach. The WT presents more than 100 alarms with different activation periods. The initial signal dataset is composed by 96 signals analysing different parameters of the WT, e.g., temperature, electric performance, among others. Each signal is acquired per minute presenting more than 500000 data, and it must be related to the critical alarm with the aim of predict failures with accuracy. Due to non-disclosure agreements with the operators, the location of the wind farm and the alarm nomenclature are omitted. The alarm selection criterion has been established by the plant operator according to the needs.

Figure 4 shows the Pareto chart, considering the number of activations, the period of the activated alarm, the maximum period of the alarm and the difference between alarms without filtering process, being the red line the accumulated data in each case.

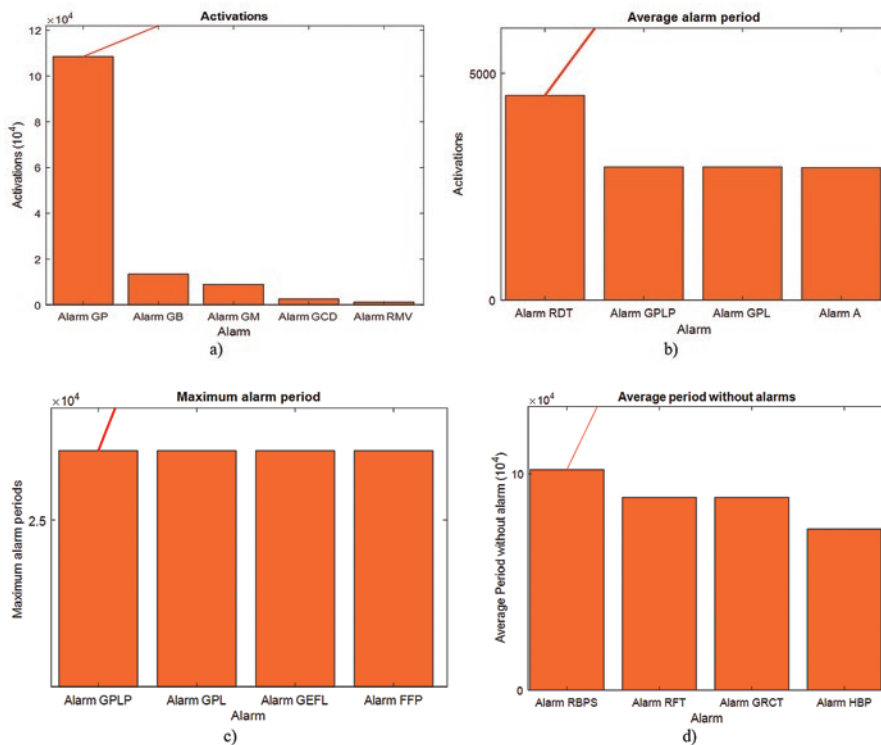


Fig. 4. Alarm activations with no filtering process. a) Alarms activation. b) Maximum period of each alarm activation. c) Average alarm period. d) Average period without alarms

Table 1 shows the alarm codification following the alarm log defined in the wind farm. Several alarms present elevated activations, but this information does not provide value to the analysis. It is required a further filtering phase to ensure proper ranges and avoid alarms with several correlated activations with inadequate data for the analysis.

The range of interest is defined in previous section in one day before the alarm activation. The alarms that do not comply with these requirements are discarded. Operators may provide different guidelines to determine the critical alarm. For this case study, the critical alarm is determined following the internal criterion of the plant operators and it is related with the overspeed of the WT. This alarm is activated in several periods but only 22 ranges between alarms achieve the condition of at least one day between activations (other case is considered

Table 1. Alarm codification.

Alarm	Alarm definition
Alarm A	Brake alarm
Alarm GB	Regulation of the pump motor
Alarm GEFL	Out-of-bounds winding
Alarm GM	CW motor alarm
Alarm GP	WT yaw mechanism alarm
Alarm GCD	Degradation alarm
Alarm GPL	Limitation of the cosine of ( $\varphi$ )
Alarm GPLP	Active power limitation
Alarm GRCT	Voltage drop of the grid
Alarm RFT	Voltage failure of the grid
Alarm HBP	Low pressure hydraulic group
Alarm RDT	Failure in the thermal ventilators of the gearbox
Alarm RMV	Slow operation of the ventilator
Alarm RBPS	Low pitch value at the WT stop

as false alarm). This process ensures the reduction of the data volume, the computational cost and increasing the accuracy of the ANN. Figure 5 shows the separation between alarms and the activation between periods.

The signals related to the critical alarm are set by the P-values and correlations, obtaining thermal signals in the 71% of the cases and the rest are vibration signals. PCA is applied on this new dataset employing the data analysis showed in Figure 2 to reduce the data volume and select fundamental periods of time. Figure 6 shows that 99% of the original dataset can be defined with two principal components, being the red line the accumulated data.

PCA modifies the initial dataset to obtain a new dataset with the same patterns, and the reconstruction error of PCA must be determined to analyze the importance of the selected variables. Figure 7 shows the unexplained variance depending on the number of components defined in the PCA process. In this case, it is confirmed that two components are suitable to reduce the uncertainty in the variance and define a new suitable dataset.

The original dataset has been reduced by 93%, from 96 signals to a new PCA dataset. The ANN will be trained and designed using this new dataset, with the same capabilities and patterns.

#### 4. Validation

The validation of the filtering process is essential for methodology acceptance. It is proposed a comparison between the scenario with no statistical analysis and the filtering method proposed in this work to validate this method. It is designed a MLP neural network with ten hidden layers and its performance is analysed with both datasets.

Figure 8 compares the performance of both ANNs, with filtering and data treatment, and with the original dataset. The cross-entropy quantifies the error between the defined outputs and the desired outputs in the training data. Minimizing cross-entropy leads to better networks. The ANN performance with the original dataset is showed in Figure 8.a) and in this case, the network employs more epochs in the stabilization, producing elevated errors and reducing the possibilities of finding reliable patterns. Figure 8.b) shows a better performance

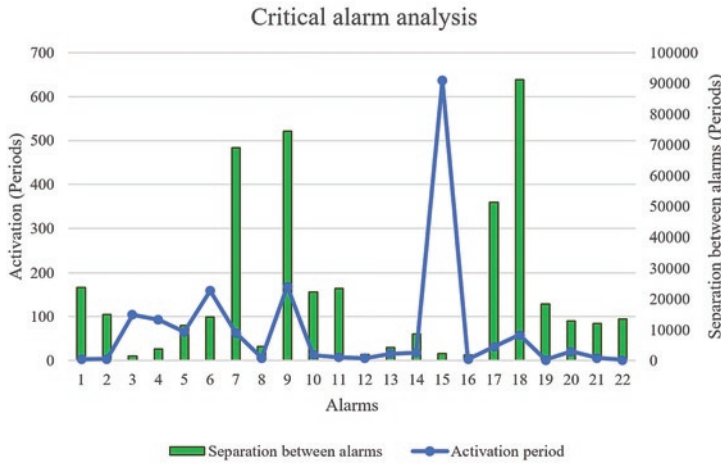


Fig. 5. Critical alarm analysis with filtered ranges

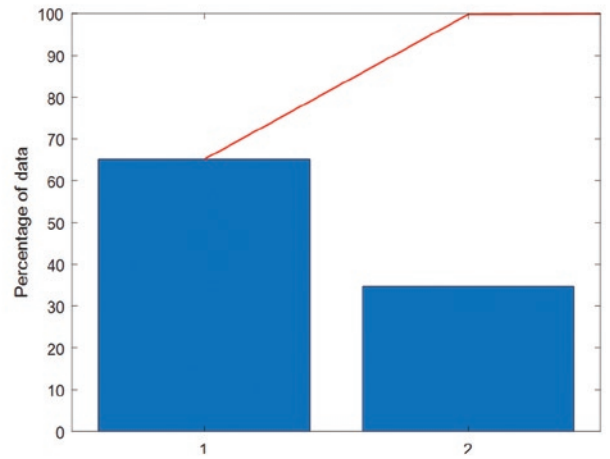


Fig. 6. Percentage of the data explained with different principal components

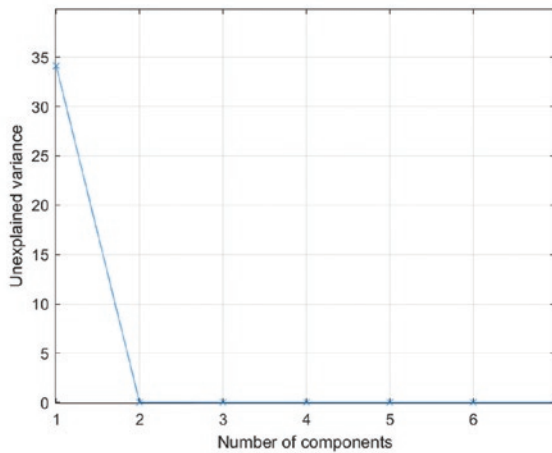


Fig. 7. Unexplained variance regarding on the number of components

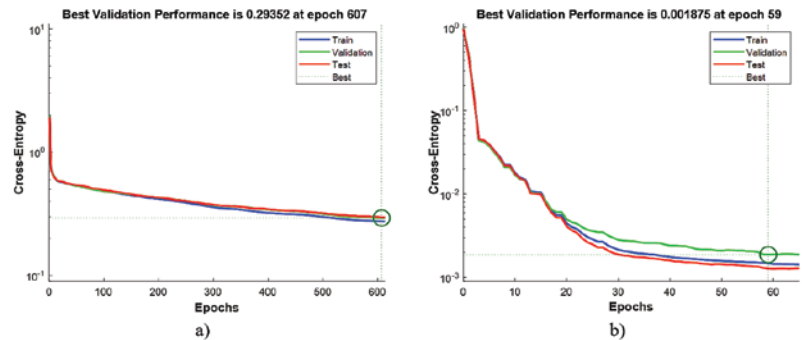


Fig. 8. a) Performance of the neural network with original dataset. b) Performance of the MLP neural network with PCA.

of the ANN, with reduced number of epochs and cross-entropy. The stabilization of this network is developed with less epochs due to an optimized and reduced dataset. The methodology proposed in this work increments the ANN performance by 1000% due to an efficient data filtering and selection.

The confusion matrix for the MLP with the filtering and data treatment process is showed in Figure 9. The number of positive cases properly identified by the algorithm is high. The overall accuracy of the ANN with the filtered dataset is about 94%, proving a high capacity to classify the periods before the activation of an alarm. It is concluded that the 6% of the alarms studied were activated without evidence of real damage in the WT, according to the signals acquired with the SCADA system.

## 5. Conclusions

The volume and type of data acquired by supervisory control and data acquisition system implies the need of robust and accuracy methods and algorithms to analyse the data. This paper presents a novel approach that filter the alarms and signals with the aim of increasing the accuracy of neural networks employed for signal processing. The initial phase is focused on the identification of the critical alarm by means of statistical analysis and Pareto chart. The signal dataset is filtered, analysed regarding with the critical alarm and reduced with principal component analysis to employ reliable and filtered data in the neural network. It is presented a real case study formed by a new dataset with thermal and vibration dataset, being electric signals filtered. The 93.5% of the cases are accurately classified, validating the methodology proposed for this paper.

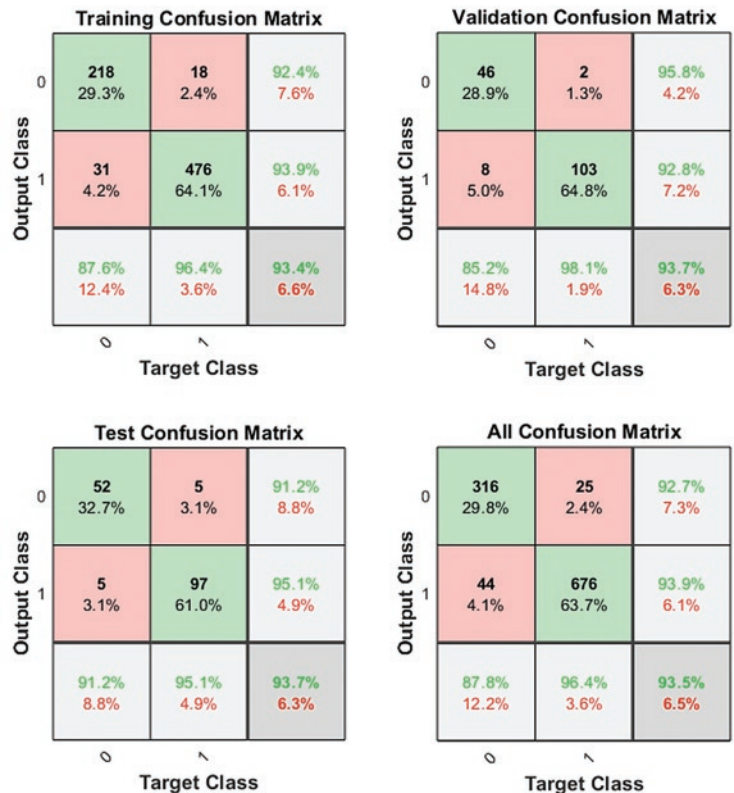


Fig. 9. Confusion matrix for the MLP ANN developed with the approach

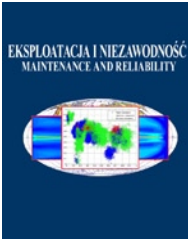
### Acknowledgments

The work reported herewith has been financially by the Dirección General de Universidades, Investigación e Innovación of Castilla-La Mancha, under Research Grant ProSeaWind project (Ref.: SBPLY/19/180501/000102) and the Spanish Ministerio de Economía y Competitividad, under Research Grant (DPI2015-67264-P).

### References

1. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* 2018; 4: e00938, <https://doi.org/10.1016/j.heliyon.2018.e00938>.
2. Adouni A, Chariag D, Diallo D, Ben Hamed M, Sbata L. FDI based on artificial neural network for low-voltage-ride-through in dfig-based wind turbine. *ISA Transactions* 2016; 64: 353-364, <https://doi.org/10.1016/j.isatra.2016.05.009>.
3. Alcaide-Moreno BA, Fuerte-Esquivel CR, Glavic M, Van Cutsem T. Electric power network state tracking from multirate measurements. *IEEE Transactions on Instrumentation and Measurement* 2017; 67: 33-44, <https://doi.org/10.1109/TIM.2017.2754838>.
4. Bangalore P, Letzgs S, Karlsson D, Patriksson M. An artificial neural network-based condition monitoring method for wind turbines with application to the monitoring of the gearbox. *Wind Energy* 2017; 20: 1421-1438, <https://doi.org/10.1002/we.2102>.
5. Catelani M, Ciani L, Galar D, Patrizi G. Optimizing maintenance policies for a yaw system using reliability centered maintenance and data-driven condition monitoring. *IEEE Transactions on Instrumentation and Measurement* 2020, <https://doi.org/10.1109/TIM.2020.2968160>.
6. Chacón AMP, Márquez FPG. False alarms management by data science. *Data Science and Digital Business* 2019; 301-316, [https://doi.org/10.1007/978-3-319-95651-0\\_15](https://doi.org/10.1007/978-3-319-95651-0_15).
7. Chacón AMP, Ramírez IS, Márquez FPG. False alarms analysis of wind turbine bearing system. *Sustainability* 2020; 12: 7867, <https://doi.org/10.3390/su12197867>.
8. Feise R.J. Do multiple outcome measures require p-value adjustment? *BMC Medical Research Methodology* 2002; 2: 8, <https://doi.org/10.1186/1471-2288-2-8>.
9. Galar D, Gustafson A, Tormos Martínez BV, Berges L. Maintenance decision making based on different types of data fusion. *Eksplatacja i Niezawodność - Maintenance and Reliability* 2012; 14: 135-144.
10. Gao J-X, An Z-W, Ma Q, Bai X-Z. Residual strength assessment of wind turbine rotor blade composites under combined effects of natural aging and fatigue loads. *Eksplatacja i Niezawodność - Maintenance and Reliability* 2020; 22: 601-609, <https://doi.org/10.17531/ein.2020.4.3>.
11. García Márquez F, Pliego Marugán A, Pinar Pérez J, Hillmansen S, Papaelias M. Optimal dynamic analysis of electrical/electronic components in wind turbines. *Energies* 2017; 10: 1111, <https://doi.org/10.3390/en10081111>.
12. García Márquez FP, García-Pardo IP. Principal component analysis applied to filtered signals for maintenance management. *Quality and Reliability Engineering International* 2010; 26: 523-527, <https://doi.org/10.1002/qre.1067>.
13. García Márquez FP, Segovia Ramírez I, Mohammadi-Ivatloo B, Marugán AP. Reliability dynamic analysis by fault trees and binary decision diagrams. *Information MDPI* 2020; 11: 324, <https://doi.org/10.3390/info11060324>.
14. Glowacz A. Diagnostics of direct current machine based on analysis of acoustic signals with the use of symlet wavelet transform and modified classifier based on words. *Eksplatacja i Niezawodność - Maintenance and Reliability* 2014; 16: 554.
15. Gomes P, Castro R. Wind speed and wind power forecasting using statistical models: Autoregressive moving average (ARMA) and artificial neural networks (ANN). *International Journal of Sustainable Energy Development* 2012; 1, <https://doi.org/10.20533/ijesd.2046.3707.2012.0007>.
16. Gómez Muñoz CQ, García Márquez FP. A new fault location approach for acoustic emission techniques in wind turbines. *Energies* 2016; 9: 40, <https://doi.org/10.3390/en9010040>.
17. Holland S.M. Principal components analysis (pca). Department of Geology, University of Georgia, Athens, GA 2008: 30602-32501.
18. Iqbal R, Maniak T, Doctor F, Karyotis C. Fault detection and isolation in industrial processes using deep learning approaches. *IEEE Transactions on Industrial Informatics* 2019; 15: 3077-3084, <https://doi.org/10.1109/TII.2019.2902274>.
19. Joyce Lee F.Z. Global wind report 2019 (global wind energy council); <https://gwec.net/global-wind-report-2019/>, 2020.
20. Kabiri M, Amjady N. A new hybrid state estimation considering different accuracy levels of pmu and scada measurements. *IEEE Transactions on Instrumentation and Measurement* 2018; 68: 3078-3089, <https://doi.org/10.1109/TIM.2018.2872446>.
21. Kalogirou SA. Artificial neural networks in renewable energy systems applications: A review. *Renewable and Sustainable Energy Reviews* 2001; 5: 373-401, [https://doi.org/10.1016/S1364-0321\(01\)00006-5](https://doi.org/10.1016/S1364-0321(01)00006-5).
22. Li G, Shi J, Zhou J. Bayesian adaptive combination of short-term wind speed forecasts from neural network models. *Renewable Energy* 2011; 36: 352-359, <https://doi.org/10.1016/j.renene.2010.06.049>.
23. Li Y-F, Huang H-Z, Liu Y, Li H. A new fault tree analysis method: Fuzzy dynamic fault tree analysis. *Eksplatacja i Niezawodność - Maintenance and Reliability* 2012; 14.
24. Ludbrook J. Statistical techniques for comparing measurers and methods of measurement: A critical review. *Clinical and Experimental Pharmacology and Physiology* 2002; 29: 527-536, <https://doi.org/10.1046/j.1440-1681.2002.03686.x>.
25. Marquez FG. An approach to remote condition monitoring systems management. *IET International Conference on Railway Condition Monitoring* 2006: 156 - 160, <https://doi.org/10.1049/ic:20060061>.
26. Márquez FPG. A new method for maintenance management employing principal component analysis. *Structural Durability & Health Monitoring* 2010; 6: 89-99.
27. Márquez FPG, Karyotakis A, Papaelias M. *Renewable energies: Business Outlook 2050*. Springer: 2018.
28. Marugán AP, Márquez FPG, Papaelias M. In *Multivariable analysis for advanced analytics of wind turbine management*, Proceedings of the Tenth International Conference on Management Science and Engineering Management, 2017; Springer: 319-328, [https://doi.org/10.1007/978-981-10-1837-4\\_28](https://doi.org/10.1007/978-981-10-1837-4_28).
29. Marugán AP, Márquez FPG, Perez JMP, Ruiz-Hernández D. A survey of artificial neural network in wind energy systems. *Applied Energy* 2018; 228: 1822-1836, <https://doi.org/10.1016/j.apenergy.2018.07.084>.
30. Pérez JMP, Márquez FPG, Tobias A, Papaelias M. Wind turbine reliability analysis. *Renewable and Sustainable Energy Reviews* 2013; 23: 463-472, <https://doi.org/10.1016/j.rser.2013.03.018>.

31. Pliego Marugán A, Garcia Marquez FP, Lev B. Optimal decision-making via binary decision diagrams for investments under a risky environment. *International Journal of Production Research* 2017; 55: 5271-5286, <https://doi.org/10.1080/00207543.2017.1308570>.
32. Rademakers L, Braam H, Zaaier M, Van Bussel G. In Assessment and optimisation of operation and maintenance of offshore wind turbines, *Proc. EWEC*, 2003.
33. Saufi SR, Ahmad ZAB, Leong MS, Lim MH. Gearbox fault diagnosis using a deep learning model with limited data sample. *IEEE Transactions on Industrial Informatics* 2020; 16: 6263-6271, <https://doi.org/10.1109/TII.2020.2967822>.
34. Schlechtingen M, Ferreira Santos I. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical Systems and Signal Processing* 2011; 25: 1849-1875, <https://doi.org/10.1016/j.ymsp.2010.12.007>.
35. Walford CA. *Wind turbine reliability: Understanding and minimizing wind turbine operation and maintenance costs*; Sandia National Laboratories 2006, <https://doi.org/10.2172/882048>.
36. Zuber N, Bajrić R. Gearbox faults feature selection and severity classification using machine learning. *Eksploracja i Niezawodność - Maintenance and Reliability* 2020; 22: 748-756, <https://doi.org/10.17531/ein.2020.4.19>.



Article citation info:

Bocewicz G, Szwarc E, Wikarek J, Nielsen P, Banaszak Z. A competency-driven staff assignment approach to improving employee scheduling robustness. *Eksploatacja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 117–131, <http://dx.doi.org/10.17531/ein.2021.1.13>.

## A competency-driven staff assignment approach to improving employee scheduling robustness

Indexed by:



Grzegorz Bocewicz<sup>a</sup>, Eryk Szwarc<sup>a</sup>, Jarosław Wikarek<sup>c</sup>, Peter Nielsen<sup>b</sup>, Zbigniew Banaszak<sup>a</sup>

<sup>a</sup>Faculty of Electronics and Computer Science, Koszalin University of Technology, ul. Śniadeckich 2, 75-453 Koszalin, Poland

<sup>b</sup>Department of Materials and Production, Aalborg University, Fibigerstræde 16, 9220 Aalborg, Denmark

<sup>c</sup>Department of Information Systems, Kielce University of Technology, Al. 1000-lecia Państwa Polskiego, 25-314 Kielce, Poland

### Highlights

- Projects are subject to disruptions (employee absences) that influence their execution.
- Redundancy of employee competences affects the efficiency of projects driven enterprises.
- The proposed definition of robustness allows to find redundant competency frameworks.
- The considered problem is implemented in a constraint programming environment.
- The proposed approach is verified with an example of a real-life project portfolio.

### Abstract

Presented paper concerns the competency-driven staff assignment and scheduling approach to the management of project portfolios subject to perturbations caused by employee absences and/or unexpected arrival of high priority jobs. Proactive strategy is considered, which exploits the concept of employee substitutability to improve the robustness of personnel allocation in the case of occurrence of specific types of disruptions.

Solutions obtained using the model of a constraint satisfaction problem developed in this study are validated in series quantitative and qualitative experiments. With a view to future implementation in a Decision Support Systems dedicated to prototyping of proactive personnel allocation, a methodology employing the concept of a competency framework-based robustness measure is proposed. Implemented in a declarative framework, the proposed approach allows one to find a redundant competency framework robust to a given set of disruptions.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

competency assignment, competency framework, personnel scheduling, robustness measure.

## 1. Introduction

The objective of manpower planning, factors that determine employment planning, such as workforce allocation and personnel scheduling, are associated with the arrangement of work schedules and the assignment of personnel to shifts, in order to meet the demand for human resources that varies over time. In this context, a pivotal role is played by so-called project-centered planning [6], which is used in companies that divide their work into projects to which they assign different groups of employees. Typical examples of such firms include job production companies such as ship-building, bridge-building, and construction companies; companies that manufacture one-off products (e.g., yachts); businesses that produce handmade craft items like furniture; or engineer-to-order companies, in which employees must be qualified to perform creative tasks [20, 33].

In the literature of the subject [2, 22, 45], competencies are defined as a set comprising theoretical knowledge, practical skills, behaviors, and qualifications that allow workers to successfully execute their tasks. During the scheduling phase, a personnel roster (or work assignment) is constructed by assigning the available personnel resources (employees with specific personal competencies) to specific duties. In other words, planning decisions regard the allocation of project tasks (which require specific employee competencies) to

resources (employees with given competencies). Projects are often subject to various disruptions that influence the duration of activities. This means that it is necessary to develop effective approaches that allow for the generation of robust project schedules which are less sensitive to disruptions caused by such uncontrollable factors such as employee absences or the unexpected arrival of a priority job [19]. In order to deal with these uncertainties [36], organizations need to adopt proactive and reactive scheduling strategies to protect the personnel roster and to respond to operational variability, respectively. Methods must then be developed to support decision-makers in situations that require responding to dynamic changes to organizational settings, e.g., frequent changes in the scope and structure of objectives, tasks, and resources. It should be noted, however, that while the existing literature describes many methods for the assessment and determination of competency frameworks [44], the problem of constructing robust personnel rosters has received only limited attention. This is the reason why a proactive approach based on the employee substitutability concept, i.e. taking into account employees specific competencies in the event of disturbances [42], is being proposed.

The considered problem of redundant competency framework synthesis that take into account the specificity of human resources and

E-mail addresses: G. Bocewicz - [grzegorz.bocewicz@tu.koszalin.pl](mailto:grzegorz.bocewicz@tu.koszalin.pl), E. Szwarc - [eryk.szwarc@tu.koszalin.pl](mailto:eryk.szwarc@tu.koszalin.pl), J. Wikarek - [j.wikarek@tu.kielce.pl](mailto:j.wikarek@tu.kielce.pl), P. Nielsen - [peter@mp.aau.dk](mailto:peter@mp.aau.dk), Z. Banaszak - [zbigniew.banaszak@tu.koszalin.pl](mailto:zbigniew.banaszak@tu.koszalin.pl)

issues concerning projects planning, fits within the framework of the well-known Redundancy Allocation Problem [44].

The present study is a continuation of our previous work, which explored methods of fast prototyping of solutions to workforce allocation and personnel scheduling problems that are robust to a given type of disruptions occurring in the course of the execution of multiple projects [12]. The main contributions of this paper are as follows:

- 1) Proposed approach to the prototyping of robust competency-driven staff assignments and schedules takes into account both: the projects are subject to disruptions (employee absences) that influence their execution, the redundancy of employee competences affects the efficiency of projects driven enterprises. Thus it allows for the construction of more realistic, i.e., more accurate, models, taking into account proactive strategies that guarantee robust arrangement of work schedules and robust assignment of personnel to a project portfolio.
- 2) The proposed definition of robustness measure allows to find redundant competency frameworks. Consequently, introducing sufficient conditions for the existence of competency structures resistant to a given type of disturbance, provides an attractive analytical method as an alternative to the currently used simulation based methods.
- 3) The considered problem is implemented in a constraint programming environment and verified with an example of a real-life project portfolio. Its constraint satisfaction (CSP) model [5, 28] allows one to search for competency-driven staff assignments and schedules robust to employee absences.

In Section 2, the overview of the literature is provided. An example introducing to the competency-driven staff assignment approach is provided in Section 3. A reference model of a CSP which allows one to find competency frameworks robust to a selected set of anticipated types of disruption is presented in Section 4. Evaluation of computational experiments verifying the proposed method is presented in Section 5. In Section 6 the conclusions and directions for further research are presented.

## 2. Related works

The last two decades have seen a rapidly increasing interest in the problems of workforce allocation and personnel scheduling in reference to the arrangement of work schedules and the assignment of personnel to shifts. There is a fast-growing body of literature on these topics [2, 7, 13, 22, 29, 31, 32, 34], which encompasses nearly all areas associated with production and services management, in particular those regarding the issues of personnel scheduling [43], e.g., crew scheduling, shift scheduling, and personnel assignment [1, 33], e.g., competency-driven staff assignment. This especially refers to settings where a creative task must be performed, for instance in engineering-to-order companies. The interlacing problems of scheduling and manpower assignment involve, allocation of employees with different competences to activities carried out in the given time intervals. Both problems are combinatorially hard [38].

As manufacturers increasingly convert their production systems from make-to-stock production systems to make-to-order or assemble-to-order production systems, in which products or parts are assembled once an order has been received, there is a growing focus on human resource management in these jobbing production environments. Jobbing production, which involves the manufacture of one-off products such as yachts, furniture, and artificial limbs, or software development, tends to be labor intensive, and requires a multi-skilled workforce. In companies that produce custom goods the problem of worker assignment, with special focus on technical and human skills, becomes particularly important [33].

One commonly used approach to improving the robustness of task assignments is to introduce time buffers or capacity buffers [10, 11, 14]. One kind of buffers, refers to the reserve staff (reserve crew, etc.) used in services, (transport, health management, etc.) in which dis-

ruptions include events such as employee sickness [27] or technical failures [17, 40]. Other commonly used approaches to staff allocation and scheduling problems that are worth mentioning include AI methods, especially those based on genetic algorithms [3], stochastic and fuzzy set-based techniques [12, 18, 32, 40], linear programming [15, 16], constraint logic programming [8], and Hungarian methods [37]. Studies [23, 25] have shown that resource redundancy affects the efficiency of an organization. However, the related works have not provided a quantitative assessment of the impact of the competencies of the existing staff on the quality of the processes carried out in an organization and their robustness to disruptions. In general, current-state focuses on methods dedicated to solving a narrowly understood problem Redundancy Allocation Problem [4, 26, 41], e.g. ignoring the specificity of available human resources (and in particular the sets of competences that characterize them), the specifics of the functioning and organization of project teams (in particular, carrying out several activities simultaneously), etc.

Recently conducted research [1, 6] focuses primarily on finding employee allocations that enable timely execution of production orders in situations caused by: disruptions (employee absenteeism), different personality types of employees (affecting the time of performing tasks), robot worker interaction [6]. The methods used are dominated by approaches based on computer or AI simulation techniques, in particular multi-agent models [1, 6].

Because project management, in essence, consists of building an order fulfillment workflow plan that is robust to disruptions (caused by employee absenteeism, unforeseen urgent production order occurrence, machine breakdowns, and so on) and results in the shortest project makespan possible, the generation of robust schedules and staff assignments as well as the measurement of their robustness have to be considered simultaneously. The concept of robustness, especially in relation to project plans, has not yet been well defined. The few studies regarding this problem that have been published are fragmentary and have a conceptual character [14, 16]. The main focus is on robustness measures. The solutions proposed in this area are related to the evaluation of the insensitivity of the schedule/assessment criteria used, and to interference caused by a given kind of disruptions. Examples of measures of this type include employee substitutability [16], quality robustness [43], schedule robustness [21], surrogate (slack-based) robustness measures [14], and others.

The literature review shows a large number of research contributions aiming to optimize resources allocation and related schedules and costs with and without considering uncertainties and abnormalities occurring in the course their usage. In general, most of these studies investigate optimization problems assuming implicitly the existence of feasible solutions (e.g. no replacement for an absent employee).

In this context the research gap that can be identified in studies conducted in the considered area concerns decision problems related to the reachability of the assumed states as well as the development of analytical methods aimed at staff assignment and employee robust scheduling. An example of such situation concerns the problem of determining whether the possible substitutions guarantee the timely execution of an order in a given case of employee absenteeism. In other words, solutions are sought that guarantee approximate but quick resolution of NP-hard decision problems. This means that the approach proposed in this paper, introducing sufficient conditions for the existence of competency framework resistant to a given type of disturbance, provides an attractive analytical method as an alternative to the currently used simulation based methods.

A review of studies that deal with robust personnel allocation and scheduling problems shows that research in this area is still in its initial phase - considered problem is the NP-hard. Results of research of synthesis competency frameworks robust to a selected set of disruptions [38, 39] confirm the attractiveness of approaches based on the declarative modeling paradigm.



### 3. Introductory example

This section introduces to the competency-driven staff assignment approach aimed at increasing robustness of the assumed competency framework. In this context a measure of competency framework robustness enables one to improve job production resistance with regard to employees' absence.

Consider a job production system where two individual jobs are performed:  $Q = \{Q_1, Q_2\}$  (Fig. 1a). The following sets of tasks:  $Z_1 = \{Z_1, \dots, Z_6\}$ ,  $Z_2 = \{Z_7, \dots, Z_{12}\}$ , (where  $l_i$  is a duration of tasks  $Z_i$  - Fig. 1c) are assigned to particular jobs  $Q_i$ . Given is a set of employees  $P = \{P_1, \dots, P_8\}$  each of which has different competences. The competency framework  $G$  adopted in the model is shown in Figure 1b, where cell values show whether a given employee  $P_k$  has the competency (value "1") to execute task  $Z_i$ .

Assuming that the tasks are non-preemptive and employees working time do not exceed 8 u.t. the answer to the following question is sought: Is it possible to assign tasks to currently available employees, guaranteeing their implementation according to the schedule shown in Figure 2?

Figure 3a illustrates, a task assignment for the case of an absence of employee  $P_6$ . Absences of a larger number of employees are presented on Figure 3b-d: cases of absence of two ( $P_6, P_8$ ), three ( $P_1, P_3, P_6$ ), and four ( $P_1, P_2, P_3, P_8$ ) employees. All of these scenarios ensure that the portfolio of projects  $Q$  is completed by the available staff within the given time horizon of 16 u.t. In the general case, however, e.g., when employees ( $P_1, P_4$ ) are absent, there are no suitable replacements able to take over their duties.

In the examples considered above, it is assumed that cases/types of absence are known before the projects in portfolio  $Q$  are executed. In practice, however, employees may be absent from work at any time during the execution of the project portfolio (due to accidents, illness, etc.). This means that, depending at which time point they occur, absences may have a different effect on the timely execution of jobs.

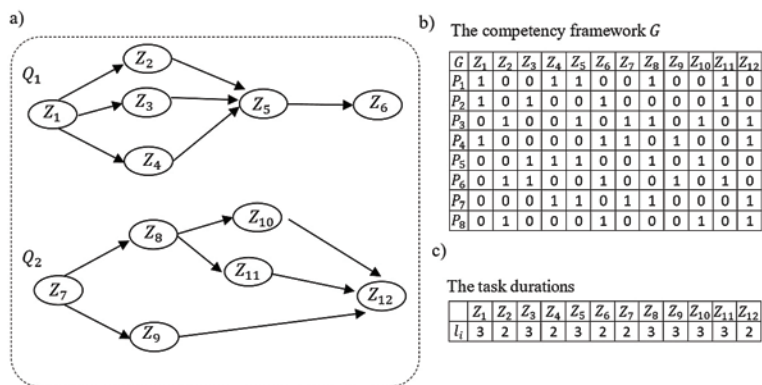


Fig. 1. a) structure of jobs  $Q_1, Q_2$ , b) competency framework  $G$ , c) task durations

An example of a schedule of job is presented in Figure 2. It is assumed that only one employee can be assigned to each task  $Z_i$ . For example, tasks  $Z_3$  and  $Z_{11}$  have been assigned to employee  $P_2$ .

In order to assess the robustness of the staff of employees  $P$  implementing the project portfolio  $Q$  to the simultaneous absenteeism of  $\omega$  employees the following concept of *Robustness of a Competency Framework* is used:

$$R_P^Q(\omega, t) = \frac{|LP_{\omega, t}|}{|U_{\omega}|}, \quad (1)$$

where:

$U_{\omega}$  – family of  $\omega$ -element employee absence scenarios:

$U_{\omega} = \{u_i \mid u_i \subseteq P; |u_i| = \omega\}$ . In the case of two employees absence see Fig. 3b) the set  $U_{\omega} = \{\{P_1, P_2\}, \{P_1, P_3\}, \{P_1, P_4\}, \dots, \{P_7, P_8\}\}$  contains 28 absence scenarios.

$LP_{\omega, t}$  – subset of set  $U_{\omega}$  ( $LP_{\omega, t} \subseteq U_{\omega}$ ) containing scenarios  $u_i$  which guarantee timely completion of the portfolio of projects  $Q$ , in the event of absences of employees, at time point  $t$ .

The values of function  $R_P^Q(\omega, t)$  belong to the range  $[0, 1] \subset \mathbb{R}$ , where:

- $R_P^Q(\omega, t) = 0$  – means no robustness, i.e., there is no scenario  $u_i$  for which the replacement guaranteeing timely completion of the planned project portfolio  $Q$  exists.
- $R_P^Q(\omega, t) = 1$  – means full robustness, i.e., for each scenario  $u_i$ , there exists at least one replacement guaranteeing timely completion of the project portfolio  $Q$ .

Assuming that in the example in Figure 1 a disruption occurs at time point  $t = 0$  with  $\omega = \{1, \dots, 4\}$  employees being absent from work, the values of  $R_P^Q(\omega, 0)$  determined from equation (1) are:

$$R_P^Q(1, 0) = \frac{7}{8} = 0.875; \quad R_P^Q(2, 0) = \frac{20}{28} = 0.71; \quad R_P^Q(3, 0) = \frac{30}{56} = 0.53; \quad R_P^Q(4, 0) = \frac{14}{70} = 0.2,$$

This means that for seven scenarios of one employee absence ( $R_P^Q(1, 0)$ ), there exists the replacement guaranteeing completion of the portfolio of projects  $Q$ . Similarly, in the case of a simultaneous absence of four employees  $R_P^Q(4, 0)$ , project portfolio  $Q$  can be completed on time in 14 out of the 70 possible absence scenarios. Values of  $R_P^Q(\omega, t)$  have been deter-

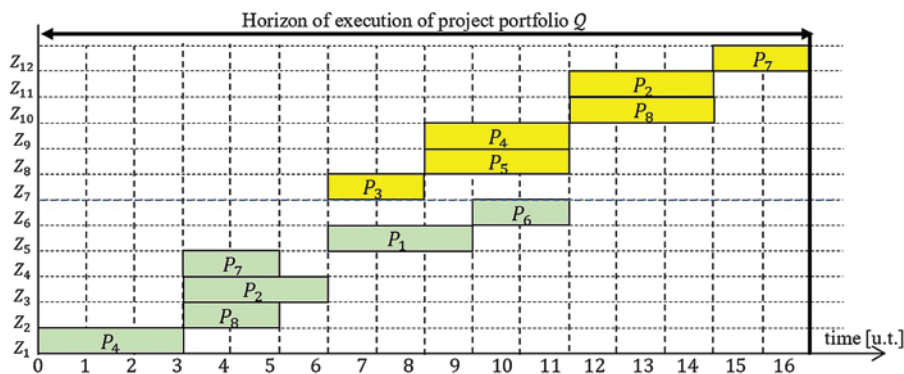


Fig. 2. Schedule of jobs  $Q_1, Q_2$

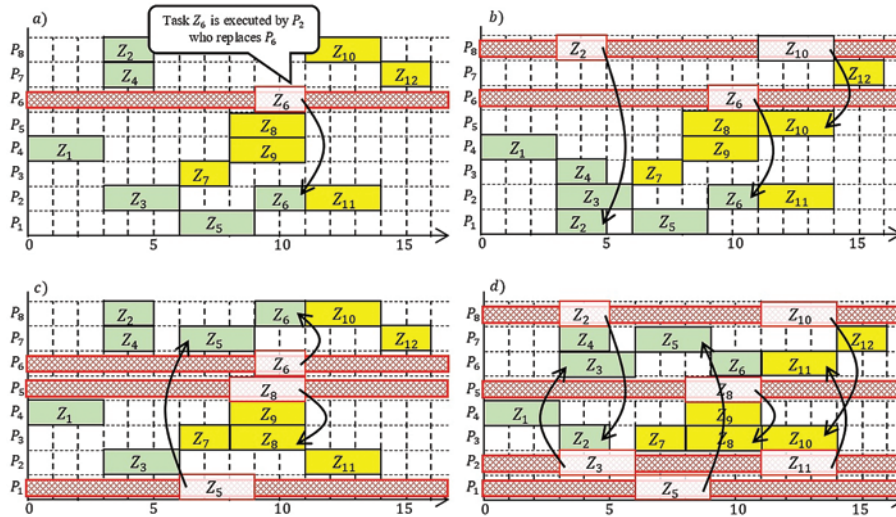


Fig. 3. An illustration showing examples of replacement options in selected cases of absences of a) one employee, b) two employees, c) three employees, d) four employees

mined in a similar way for other time points  $t = 0 \dots 16$  along the time horizon  $H$  :

$$R_{\mathcal{P}}^Q(1,t) = \frac{7}{8} = 0.875; R_{\mathcal{P}}^Q(2,t) = \frac{20}{28} = 0.71; R_{\mathcal{P}}^Q(3,t) = \frac{30}{56} = 0.53; R_{\mathcal{P}}^Q(4,t) = \frac{14}{70} = 0.2$$

for  $t = 0 \dots 8$ ,

$$R_{\mathcal{P}}^Q(1,t) = \frac{7}{8} = 0.875; R_{\mathcal{P}}^Q(2,t) = \frac{21}{28} = 0.75; R_{\mathcal{P}}^Q(3,t) = \frac{37}{56} = 0.66; R_{\mathcal{P}}^Q(4,t) = \frac{22}{70} = 0.31$$

for  $t = 9, 10$ ,

$$R_{\mathcal{P}}^Q(1,t) = \frac{8}{8} = 1.0; R_{\mathcal{P}}^Q(2,t) = \frac{28}{28} = 1.0; R_{\mathcal{P}}^Q(3,t) = \frac{46}{56} = 0.82; R_{\mathcal{P}}^Q(4,t) = \frac{43}{70} = 0.61$$

for  $t = 11, 12, 13$ ,

$$R_{\mathcal{P}}^Q(1,t) = \frac{8}{8} = 1.0; R_{\mathcal{P}}^Q(2,t) = \frac{28}{28} = 1.0; R_{\mathcal{P}}^Q(3,t) = \frac{56}{56} = 1.0; R_{\mathcal{P}}^Q(4,t) = \frac{65}{70} = 0.93$$

for  $t = 14, 15, 16$ .

It is easy to see that the value of robustness  $R_{\mathcal{P}}^Q(\omega, t)$  of the adopted competency framework varies depending on the time point  $t$  at which the disruption occurs. The changes in robustness are shown in graphic form in the radar charts in Figures 4–6. In these figures, robustness  $R_{\mathcal{P}}^Q(\omega, t)$  of the project portfolio  $Q$  is marked in blue, while green and yellow mark robustness values for the individual jobs  $Q_1, Q_2$ . It is worth noting that the robustness of the portfolio  $Q$  corresponding to different cases of absence ( $\omega = 1 \dots 4$ ) of employees increases monotonically with time to completion of the portfolio. In addition, the values of robustness of jobs  $Q_1, Q_2$  are not less than the robustness of the entire portfolio  $Q$ ; this is due to the fact that an absence of employees can disrupt the execution of only one of the jobs without affecting the robustness of other jobs to be completed as part of the portfolio.

It is noteworthy that in the case under study the differences between robustness values  $R_{\mathcal{P}}^Q(\omega, t)$  and at the beginning and end of time horizon  $H$  are relatively large. This observation naturally raises the question of whether it is possible to restructure the competency framework  $G$  in such a way as to ensure that robustness values change in a predetermined manner.

The essence of this question is illustrated in Figures 5 and 6, which show the expected robustness values (so called robustness thresholds  $*R_{\mathcal{P}}^Q(\omega, t)$ , see the red line) that guarantee timely implementation of the considered portfolio of projects along the entire time horizon  $H$ .

The value 0.8 is reachable only for the entire portfolio  $Q$  at time point  $t = 14$  u.t. (from  $t = 11$  u.t. for job  $Q_1$  and from  $t = 14$  u.t. for job  $Q_2$ ). Which employee should acquire which competencies in the new competency framework  $G'$  to guarantee timely completion?

In general, the robustness can be expressed by matrix, and the corresponding matrix of thresholds robustness describing different thresholds of robustness for different number of absent employers ( $\omega$ ) and different time ( $t$ ). In that context the required level of an individual employee's absenteeism may be higher than of two absent employers and so on. The considered problem of robust competency frameworks synthesis boils down to the following question:

*Does there exist, for the given portfolio of projects  $Q$  executed by a staff of employees  $\mathcal{P}$ , a competency framework  $G$ , which, at any time point along the adopted time horizon  $H$  guarantees robustness values  $R_{\mathcal{P}}^Q \geq *R_{\mathcal{P}}^Q$ ?*

In this context, since the selection of redundant competencies that ensure the  $R_{\mathcal{P}}^Q$  value at a given  $*R_{\mathcal{P}}^Q$  level enables the protection of the execution of given production orders against the effects of specific disruptions, hence the fulfillment of condition  $R_{\mathcal{P}}^Q(\omega, t) \geq *R_{\mathcal{P}}^Q(\omega, t)$  guarantees the existence of the sufficient solution ensuring the timely execution of considered order.

#### 4. Modelling and problem description

The formalism of the CSP seems to be best suited for modelling of the robust competency frameworks synthesis problem. Moreover, it can be implemented in a constraints programming environment to generate feasible scenarios of execution of the projects portfolio in terms of appropriate workforce allocation and personnel scheduling.

##### 4.1. A reference model

An organization's production potential and the requirements posed by the production orders placed (hereinafter referred to as the "project portfolio") can be represented as part of the reference model, which consists of a model of the portfolio of projects executed in the system and a model of the framework of the competencies possessed by the organization's personnel.

**Project Portfolio  $Q$ .** The portfolio is assumed to include projects that are executed at a customer's order or are the organization's own undertakings (e.g., modernization or execution of production orders). A formula is adopted in which  $Q = \{Q_1, \dots, Q_j, \dots, Q_{lq}\}$  stands for a project portfolio, where  $Q_j$  is the  $j$ -th job that involves a set of tasks (activities)  $Z_j \subseteq Z = \{Z_1, \dots, Z_i, \dots, Z_n\}$ , and  $Z$  is a set of tasks  $Z_i$  to be executed by the organization. A task  $Z_i$  is defined as follows:

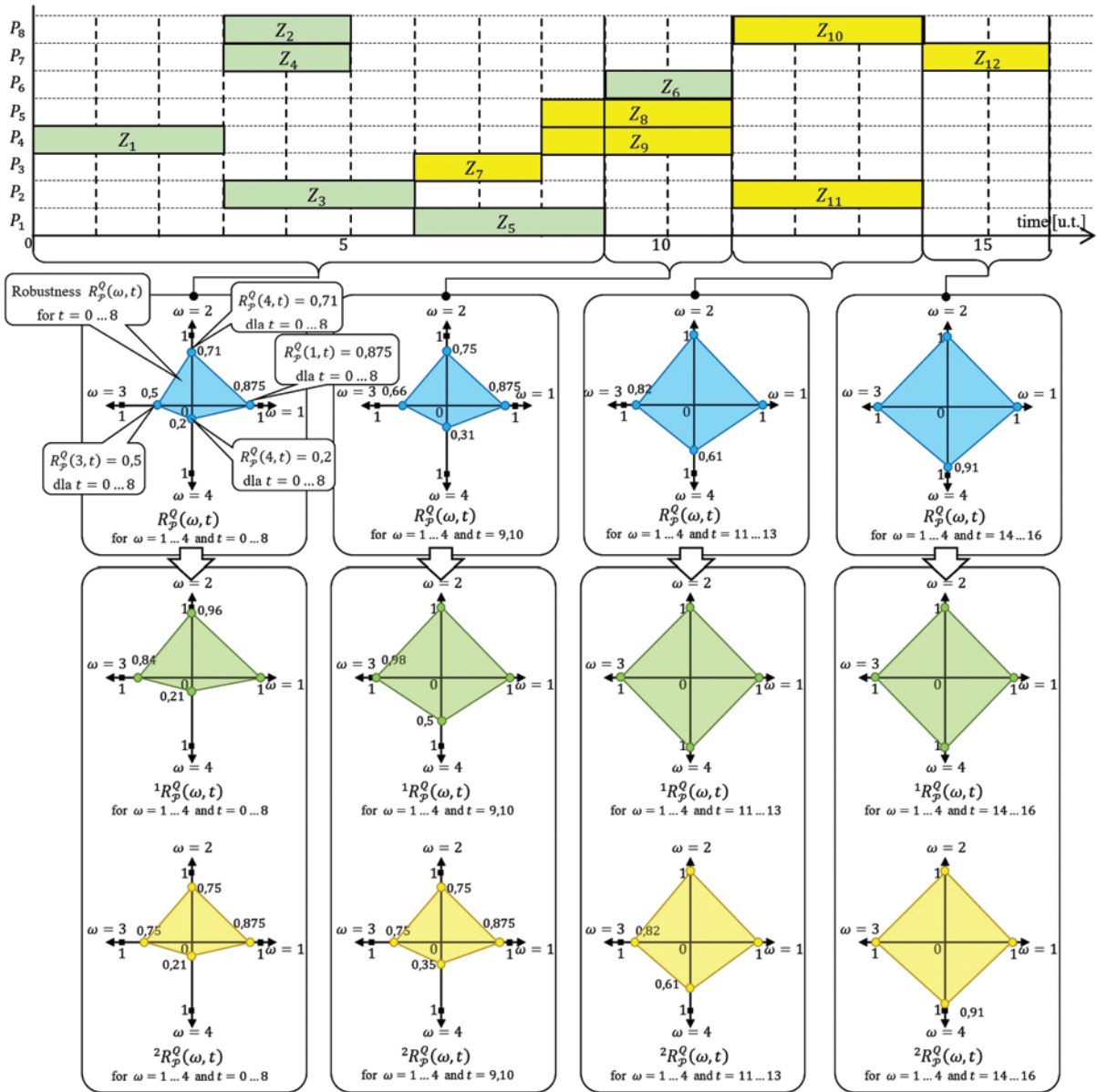


Fig. 4. Changes in robustness  $R_p^Q(\omega, t)$ ,  $= 1 \dots 4$ , of competency framework at selected time points along the time horizon

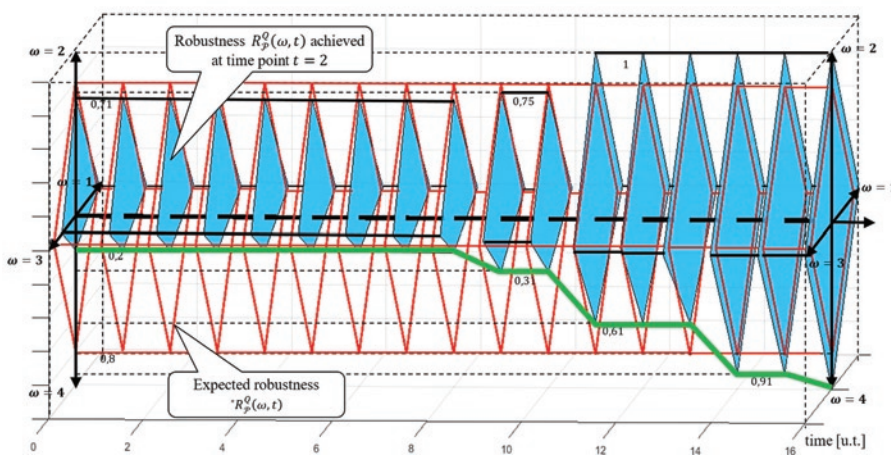


Fig. 5. Graphs of observed and expected changes in robustness  $R_p^Q(\omega, t)$  of competency framework  $G$

$$Z_i = (y_i, l_i, w_i, \Phi_i), \quad (2)$$

where:

- $y_i$  : starting time of task  $Z_i$ ,
- $l_i$  : duration of task  $Z_i$ ,
- $w_i$  : set of tasks that exclude the execution of task  $Z_i$ ,  $w_i \subseteq Z$ ; task  $Z_i$  and task  $Z_a \in w_i$  are said to be mutually exclusive when they cannot be performed by the same employee,
- $\Phi_i$  : number of employees necessary to complete the task  $Z_i$ .

It is assumed that job  $Q_j$  is characterized by a network of tasks that can be represented as a Task-on-Node (TN) network diagram in which tasks  $Z_i$  are assigned to nodes, and precedence relationships are represented by arcs (see Fig. 1). The task network can be represented as the digraph  $DG_j = (\mathbb{Z}_j, E_j)$  where  $\mathbb{Z}_j$  refers to the set of tasks of job  $Q_j$  and  $E_j \subseteq \mathbb{Z}_j \times \mathbb{Z}_j$  are the set of arcs.

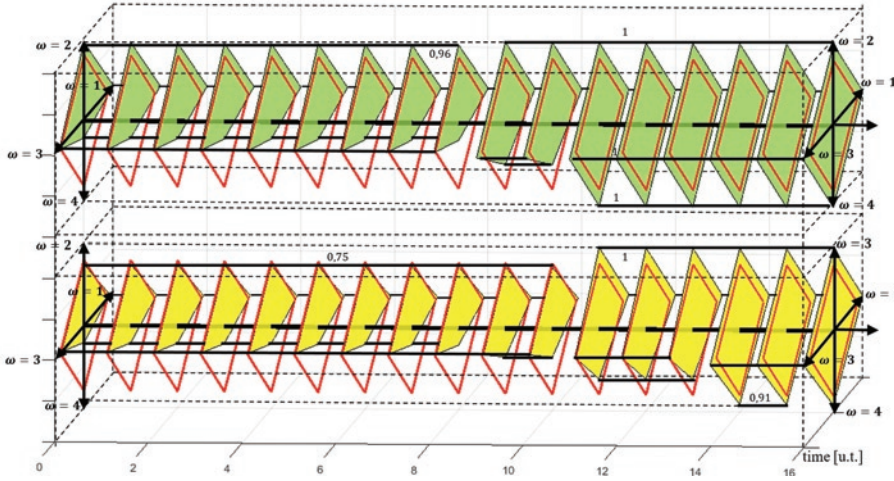


Fig. 6. Graphs of observed and expected changes in robustness  $R_P^O(\omega, t)$  and  $R_P^E(\omega, t)$  of competency framework  $G$  for projects  $Q_1$  and  $Q_2$  respectively

In addition, it is assumed that:

- project portfolio  $\mathcal{Q}$  is completed (i.e., all tasks in portfolio  $\mathcal{Q}$  are to be completed) in a given horizon time  $H$ ,
- tasks are indivisible in time, i.e., once started, a task cannot be interrupted until it has been completed,
- tasks are completed by a staff of  $\varphi_i$  competent employees.

**Staff.** Set  $\mathcal{P} = \{P_1, \dots, P_k, \dots, P_m\}$  represents a employees, where  $P_k$  is a pair:

$$P_k = (s_k, z_k) \quad (3)$$

where  $s_k$  and  $z_k$  determines the minimum/maximum working hours of  $P_k$ .

For the set  $\mathcal{P}$  the competency framework  $G$  is defined:

$$G = [g_{k,i}]_{k=1..m; i=1..n}, \quad (4)$$

where:

$$g_{k,i} = \begin{cases} 1 & \text{when employee } P_k \text{ has the competencies to execute task } Z_i \\ 0 & \text{in remaining cases} \end{cases}$$

Assignment  $X$  defined by the following matrix determines the tasks assigned to employees from the set  $\mathcal{P}$ :

$$X = [x_{k,i}]_{k=1..m; i=1..n}, \quad (5)$$

where:  $x_{k,i} \in \{0,1\}$ ,

$$x_{k,i} = \begin{cases} 1 & \text{when task } Z_i \text{ is executed by employee } P_k \\ 0 & \text{in remaining cases} \end{cases}$$

For example, assignment  $X$  corresponding to the plan from Figure 2, has the following form:

**Disruptions.** Considered type of disruptions is characterized by the set  $U_\omega = \{u_i | u_i \subseteq \mathcal{P}; |u_i| = \omega\}$  imposing ensuing from this the set of  $\omega$ -element scenarios of employee absences. The disruptions occurrence makes the search for assignment  $X$  guaranteeing timely completion of considered projects portfolio  $\mathcal{Q}$ . Consequently, the

concept of robustness  $R_P^O(\omega, t)$  (1) is used to evaluate this kind of assignment.

To put this type of problems into formal terms, the following reference model is introduced:

### Sets:

- $Z$ : tasks executed as part of the project portfolio  $\mathcal{Q}$ :  $Z = \{Z_1, \dots, Z_n\}$ ,
- $H$ : horizon of completion of project portfolio  $\mathcal{Q}$ :  $H = \{0, 1, \dots, h\}$ ,
- $\mathcal{P}$ : set of employees,  $\mathcal{P} = \{P_1, \dots, P_m\}$ ,
- $U_\omega$ : family of  $\omega$ -element employee absence scenarios:

$$U_\omega = \left\{ u_f \mid u_f \subseteq \mathcal{P}; |u_f| = \omega; f = 1 \dots \binom{|\mathcal{P}|}{\omega} \right\},$$

where  $u_f$  is a set of absence employee (employee absence scenario),

$LP_{\omega,t}$ : subset of set  $U_\omega$  ( $LP_{\omega,t} \subseteq U_\omega$ ) containing scenarios for which  $G$  ensure timely completion of the projects in the event employees are absent at time point  $t$ .

pletion of the projects in the event employees are absent at time point  $t$ .

### Parameters:

- $n$ : number of tasks executed as part of the project portfolio  $\mathcal{Q}$ ,
- $m$ : number of employees of staff  $\mathcal{P}$ ,
- $\omega$ : number of employees of staff  $\mathcal{P}$ ,  $\omega < m$ ,
- $l_i$ : duration of task  $Z_i$ ,
- $y_i$ : start time of task  $Z_i$ ,
- $\varphi_i$ : number of employees needed to execute task  $Z_i$ ,
- $s_k$ : minimum limit of working time of  $P_k$ ,
- $z_k$ : maximum limit of working time of  $P_k$ ,
- $w_i$ : set of tasks that exclude the execution of task  $Z_i$ ,  $w_i \subseteq Z$ ,
- $*R_P^O$ : expected robustness of the competency framework, given by the matrix:  $*R_P^O = [ *R_P^O(\omega, t) ]_{\omega \times t}$  where:  $*R_P^O(\omega, t) \in [0, 1]$  - expected robustness to the absence of  $\omega$  employees at time point  $t$ .

### Decision variables:

- $G$ : competency framework given by matrix  $G = [g_{k,i}]_{k=1..m; i=1..n}$  where:
- $g_{k,i} = \begin{cases} 1 & \text{when employee } P_k \text{ has the competencies to execute task } Z_i \\ 0 & \text{in remaining cases} \end{cases}$
- robustness of competency framework  $G$ , given by matrix:  $R_P^O = [R_P^O(\omega, t)]_{\omega \times t}$ , where:  $R_P^O(\omega, t) \in [0, 1]$  is the robustness to disruption by the absence of  $\omega$  employees at time point  $t$ ,
- $G^{u_f}$ : competency framework taking into account absences from the employee absence scenario  $u_f \in U_\omega$ :
- $G^{u_f} = [g_{k,i}^{u_f}]_{k=1..m; i=1..n}$  where:
- $g_{k,i}^{u_f} = \begin{cases} 1 & \text{when } k \notin u_f \text{ and } P_k \text{ has the competencies to execute task } Z_i \\ 0 & \text{in remaining cases} \end{cases}$
- $X$ : assignment of tasks in the portfolio  $\mathcal{Q}$  to employees of staff  $\mathcal{P}$ ,
- $X = [x_{k,i}]_{k=1..m; i=1..n}$  where:
- $x_{k,i} = \begin{cases} 1 & \text{when task } Z_i \text{ is executed by employee } P_k \\ 0 & \text{in remaining cases} \end{cases}$

$X^{uf}$ : assignment taking into account absences from the employee absence scenario  $u_f \in U_\omega$ :  $X^{uf} = \left[ x_{k,i}^{uf} \right]_{k=1..m; i=1..n}$  where:

$$x_{k,i}^\Theta = \begin{cases} 1 & \text{when task } Z_i \text{ is executed by employee } P_k, \\ 0 & \text{in remaining cases} \end{cases}$$

$c^{uf}$ : a variable that specifies whether there exists assignment  $X^{uf}$  ensuring timely completion of project portfolio  $\mathcal{Q}$ .

### Constraints:

The matrix  $G^{uf}$  takes the value 0 for cells  $g_{k,i}^{uf}$  corresponding with absent employees ( $P_k \in u_f$ ):

$$g_{k,i}^{uf} = \begin{cases} g_{k,i} & \text{when } P_k \notin u_f \\ 0 & \text{when } P_k \in u_f \end{cases} \quad (6)$$

Employees who have the appropriate competencies can execute the tasks:

$$x_{k,i}^{uf} \leq g_{k,i}^{uf}, \text{ for } k = 1..m; i = 1..n; u_f \in U_\omega. \quad (7)$$

At a given time point, an employee executes at most one task:

$$\left( (y_\alpha + l_\alpha \leq y_\beta) \vee (y_\beta + l_\beta \leq y_\alpha) \right) \Rightarrow \left( x_{k,\alpha}^{uf} + x_{k,\beta}^{uf} \leq 1 \right). \\ \alpha, \beta = 1..n; k = 1..m; u_f \in U_\omega. \quad (8)$$

Each task is executed by exactly  $\varphi_i$  employees:

$$\left( \sum_{k=1}^m x_{k,i}^{uf} = \varphi_i \right) \Leftrightarrow \left( c_{1,i}^{uf} = 1 \right), \text{ for } i = 1..n; u_f \in U_\omega. \quad (9)$$

Workload of  $P_k$  should not to exceed the minimum/maximum limit  $s_k / z_k$ :

$$\left( \sum_{i=1}^n x_{k,i}^{uf} \cdot l_i \geq s_k \right) \Leftrightarrow \left( c_{2,k}^{uf} = 1 \right), \text{ for } P_k \in \mathcal{P} \setminus u_f; u_f \in U_\omega. \quad (10)$$

$$\left( \sum_{i=1}^n x_{k,i}^{uf} \cdot l_i \leq z_k \right) \Leftrightarrow \left( c_{3,k}^{uf} = 1 \right), \text{ for } P_k \in \mathcal{P} \setminus u_f; u_f \in U_\omega. \quad (11)$$

Execution of mutually exclusive tasks:

$$(Z_b \in w_i) \Rightarrow \left( x_{k,i}^{uf} + x_{k,b}^{uf} \leq 1 \right), \text{ for } i = 1..n, k = 1..m; u_f \in U_\omega. \quad (12)$$

According to (1) the robustness  $R_P^Q(\omega, t)$  is calculated as the following ratio:

$$R_P^Q(\omega, t) = \frac{|LP_{\omega,t}|}{|U_\omega|} \geq {}^*R_P^Q(\omega, t), t \in H, \quad (13)$$

$$|LP_{\omega,t}| = \sum_{u_f \in U_\omega} c^{uf}, \quad (14)$$

$$c^\Theta = \prod_{i=1}^n c_{1,i}^{uf} \prod_{k=1}^m c_{2,k}^{uf} \prod_{k=1}^m c_{3,k}^{uf}. \quad (15)$$

The open structure of the proposed model, allowing it to be easily expanded by various combinations of various criteria and restrictions that occur in practice, implies a choice of constraint programming (CP) formalism implementing the paradigm of declarative modeling, the essence of the CSP problem formulation. In this context, the original element of research is the proposed measure  $R_P^Q(\omega, t)$  of robustness of competency framework  $G$  to the absences of  $\omega$  employees. A feature of the measure that has been recognized as a result of the research is its monotonic course, which increases with the approaching project completion date. This fact finds its practical use in computer-aided interactive resource allocation planning systems.

### 4.2. Problem definition

The structure of the adopted model allows one, in a natural way, to formulate the synthesis problem of robust competency framework  $G$  as a Constraints Satisfaction Problem:

$$CS = ((\mathcal{V}, \mathcal{D}), \mathcal{C}), \quad (16)$$

where:

$\mathcal{V} = \left\{ G, G^{u_f \in U_\omega}, X^{u_f \in U_\omega}, R_P^Q \right\}$  – set of decision variables, including  $G$  and competency subframeworks  $G^{u_f \in U_\omega}$  corresponding to a situation of simultaneous absence of  $\omega$  employees, assignments  $X^{u_f \in U_\omega}$ , and  $R_P^Q$ ,

$\mathcal{D}$  – a finite set of decision variable domains  $\left\{ G, G^{u_f \in U_\omega}, X^{u_f \in U_\omega}, R_P^Q \right\}$ :

$$g_{k,i} \in \{0,1\}, g_{k,i}^{uf} \in \{0,1\}, x_{k,i}^{uf} \in \{0,1\}, R_P^Q(\omega, t) \in [0,1],$$

$\mathcal{C}$  – a set of constraints specifying the relationships among the variables  $G, Z, R_P^Q$  (constraints (6)–(15)).

To solve  $CS$  problem (16), we have to find the values of variables  $G$  (personnel competency framework),  $X^{uf}$  (assignment), and  $R_P^Q$ , for which all the constraints given in set  $\mathcal{C}$  are satisfied. In other words, the solution to  $CS$  is a variant of competency framework  $G$  which guarantees the given value of  $R_P^Q$  for a given type of disruptions.

In general the CSP (16) can be treated as an optimization constraint optimization problem (COP) [44] given by the formula:

$$CO = ((\mathcal{V}, \mathcal{D}), \mathcal{C}, F), \quad (17)$$

where:  $(\mathcal{V}, \mathcal{D}), \mathcal{C}$  are defined as in (16), and  $F$  is the objective function:

$$F(G) = \sum_{k=1..m}^{i=1..n+\lambda} g_{k,i}. \quad (18)$$

To solve  $CO$  (17), one has to determine such values of decision variable  $G_{OPT}$  for which all constraints given in the set  $\mathcal{C}$  are satisfied and for which function  $F$  has a minimum value (a minimum number of changes have to be made to the original competency framework  $G$ ) or, stated differently, returns a minimum competency framework. In general,  $CO$  (17) allows one to synthesize (minimum) robust competency frameworks. In addition to the aforementioned benefits resulting from the adoption of the declarative modeling paradigm (enabling, among others, the implementation of the introduced measure of competency framework robustness  $R_P^Q(\omega, t)$ ) another of its advantages is the possibility to simultaneously evaluate all employee absence scenarios with a given (currently analyzed) variant of

the set of competences. The proposed approach is illustrated in Figure 7 (corresponding to the  $\omega = 1$  instance). The adopted approach assumes that each considered instance of competency framework  $G$  corresponds to the set of competency frameworks  $G^{u,f}$  (representing subsequent cases of employee absenteeism) and the corresponding  $X^{u,f}$  allocation. This means that the mechanisms (implemented in CP environments) used to search for solutions by specifying the values of subsequent elements  $g_{j,i}$  (see the red line in Fig. 7) of competency framework  $G$  determine the degree of compliance with restrictions for each case of absence, which allows the determination of the value of robustness level  $R_P^O(\omega, t)$  (13). In this manner it becomes possible to determine the competency framework  $G_{OPT}$  guaranteeing robustness value  $R_P^O \geq R_P^O^*$  without iterative necessity, determining the  $R_P^O(\omega, t)$  values characteristic of imperative programming methods.

In other words, the space for potential solutions is screened against the criterion of meeting (at a given level determined by  $R_P^O^*$ ) the constraints ((6)–(15)) for all variants of absence of the considered instance of the problem. The benefit of this fact is that once obtained, confirmation of existence of an admissible assignment  $X^{\{P_j\}}$  for the absence scenario  $P_j$  (answer YES in Fig. 7) does not need to be confirmed again in the further synthesis process of the competency framework  $G$ . Consequently, this allows the search process to be limited to those scenarios for which there is no allocation of  $X^{\{P_j\}}$  that meets the restrictions (6)–(15). The limitation of the search space

is implemented by mechanisms of constraints propagation and variables distribution implemented in CP environments.

## 5. Computer experiments

The quantitative and qualitative evaluation of the effectiveness of the proposed method of synthesizing robust competency frameworks was verified in a series of experiments. In the constraint optimization problem from (17) being solved the input data used was an archival data collected from selected project-driven organizations.

### 5.1. Qualitative assessment

Consider the project portfolio shown in Figure 1, which is executed by a staff of  $\mathcal{P} = \{P_1, \dots, P_8\}$  employees. The method of planning tasks for the individual jobs is shown in the schedule in Figure 2. As Figures 4 and 5 clearly demonstrate, the adopted competency framework  $G$  (Figure 1b) does not guarantee the required level of robustness  $R_P^O(\omega, t) \geq 0.8$  to the absences of  $\omega \in \{1, \dots, 4\}$  employees within time horizon  $H = \{0, 1, \dots, h\}$ .

Which employee should acquire which competencies from the new competency framework  $G'$  to guarantee  $R_P^O(\omega, t) \geq 0.8$ ?

To answer this question, the problem CO (17) was solved (implementation in the GUROBI/Intel i7-4770, 8GB RAM). The obtained minimum competency framework  $G_{OPT}$  (time computation = 1s.) is shown in graphic form in the Table 1.

This shows that employees must improve their qualifications by acquiring nine new redundant competencies: employee  $P_1$  should acquire the competencies necessary to execute tasks  $Z_3$  and  $Z_9$ ;  $P_3$  competencies for tasks  $Z_4$  and  $Z_9$ ;  $P_5$  competencies for task  $Z_2$ ;  $P_6$  competencies for tasks  $Z_1$  and  $Z_{10}$ ; and  $P_8$  competencies for tasks  $Z_7$  and  $Z_8$ . Acquisition of these competencies guarantees robustness  $R_P^O(\omega, t) \geq 0.8$  across the time horizon  $H$ . Robustness  $R_P^O(\omega, t)$  for time points  $t = 0 \dots 16$  along time horizon  $H$  takes the following values (see Fig. 8):

$$R_P^O(1,t)=1; R_P^O(2,t)=1; R_P^O(3,t)=0.98; R_P^O(4,t)=0.8 \text{ for } t=0 \dots 8,$$

$$R_P^O(1,t)=1; R_P^O(2,t)=1; R_P^O(3,t)=0.98; R_P^O(4,t)=0.87 \text{ for } t=9,10,$$

$$R_P^O(1,t)=1; R_P^O(2,t)=1; R_P^O(3,t)=0.98; R_P^O(4,t)=0.9 \text{ for } t=11,$$

$$R_P^O(1,t)=1; R_P^O(2,t)=1; R_P^O(3,t)=1; R_P^O(4,t)=0.9 \text{ for } t=12,13,$$

$$R_P^O(1,t)=1; R_P^O(2,t)=1; R_P^O(3,t)=1; R_P^O(4,t)=0.99 \text{ for } t=14,15,$$

$$R_P^O(1,t)=1; R_P^O(2,t)=1; R_P^O(3,t)=1; R_P^O(4,t)=1 \text{ for } t=16.$$

The approach proposed in this paper has been verified in several experiments involving: 10–100 employees and 2–4 jobs (consisting of a different number of tasks (10–100)). Calculations were made to determine the time needed to synthesize a competency framework () robust to the absences of employees across the time horizon (which results from the critical path of the jobs being executed). The obtained results (Table 2) show when the size of the problem does not exceed three jobs and 60 tasks, can be found in less than one hour.

It is worth noting that in real-life settings, project portfolios are executed in parallel with other jobs run simultaneously, often involving the same employees. This means that some of the workers can be engaged in executing a given project portfolio only during certain periods along time horizon  $H$ , which strongly limits the possibility of finding replacements for absent staff members. Figure 9 shows a schedule for project portfolio

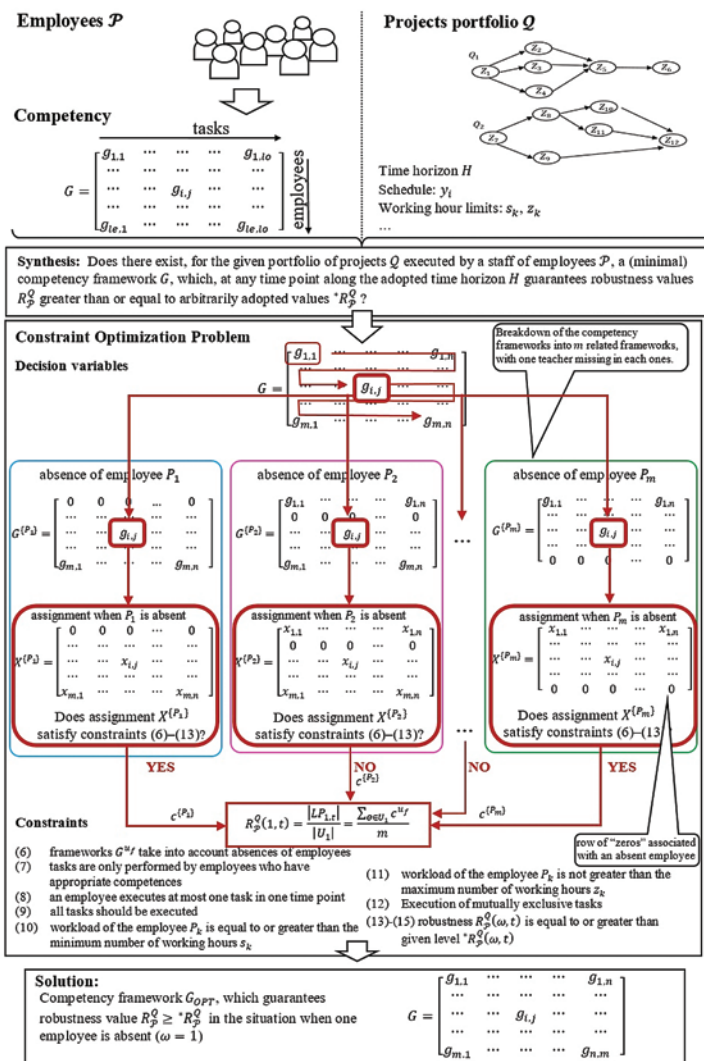


Fig. 7. The idea of COP (17) usage for synthesis of robust competency framework ( $\omega = 1$ )

$Q$ , which incorporates examples of employee unavailability intervals for a staff of employees  $\mathcal{P}$  (the set of unavailability intervals is hereafter referred to as a mask  $M$  – set of additionally tasks with fixed assignment).

As it turns out, when additional tasks  $M$  are taken into consideration, the robustness  $R_P^Q(\omega, t)$  of portfolio  $Q$  is affected. For example, consider a situation in which the tasks scheduled as in Figure 9 are executed by employees  $\mathcal{P}$  who have competencies defined by framework  $G_{OPT}$  of Table 1.

Does competency framework  $G_{OPT}$  guarantee robustness  $R_P^Q(\omega, t) \geq 0.8$  of portfolio  $Q$  when mask  $M$ , which specifies the unavailability of employees over time, is considered (as in Fig. 9)? The following values of  $R_P^Q(\omega, t)$  are obtained:

Table 1. Competency framework which guarantees  $R_P^Q(\omega, t) \geq 0.8$  (the competencies the employees have to acquire are given in bold)

$G_{OPT}$	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12
P1	1	0	<b>1</b>	1	1	0	0	1	<b>1</b>	0	1	0
P2	1	0	1	0	0	1	0	0	0	0	1	0
P3	0	1	0	<b>1</b>	1	0	1	1	<b>1</b>	1	0	1
P4	1	0	0	0	0	1	1	0	1	0	0	1
P5	0	<b>1</b>	1	1	1	0	0	1	0	1	0	0
P6	<b>1</b>	1	1	0	0	1	0	0	1	<b>1</b>	1	0
P7	0	0	0	1	1	0	1	1	0	0	0	1
P8	0	1	0	0	0	1	<b>1</b>	<b>1</b>	0	1	0	1

Table 2. Calculation times for  $G_{OPT}$  guaranteeing robustness  $R_P^Q(\omega, t) = 1$  for  $\omega \in \{1, \dots, 4\}$  and  $t \in H$

Number of jobs	Number of tasks	Number of employees	Calculation time $G_{OPT}$ for			
			$\omega = 1$	$\omega = 2$	$\omega = 3$	$\omega = 4$
			T[s]	T[s]	T[s]	T[s]
2	10	5	1	1	1	2
2	20	10	1	1	34	68
2	30	15	2	8	56	234
3	40	20	4	23	346	1245
3	50	25	7	54	896	3245
3	60	30	10	78	1456	>3600
4	70	35	18	124	2753	>3600
4	80	40	23	198	>3600	>3600
4	90	45	35	234	>3600	>3600
4	100	50	50	345	>3600	>3600

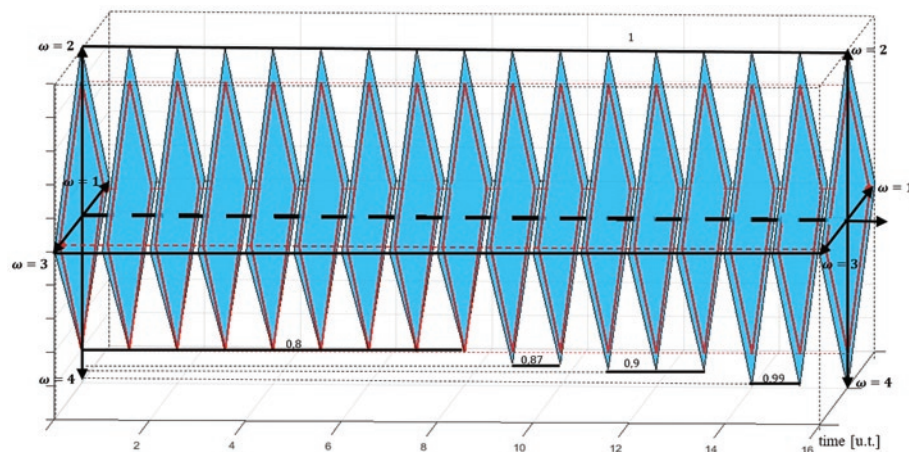


Fig. 8. Graphs of observed and expected changes in robustness  $R_P^Q(\omega, t)$  of competency framework  $G_{OPT}$  from Table 1

$R_P^Q(1, t) = 1; R_P^Q(2, t) = 0.86; R_P^Q(3, t) = 0.39; R_P^Q(4, t) = 0.06$  for  $t = 0 \dots 4$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 0.89; R_P^Q(3, t) = 0.49; R_P^Q(4, t) = 0.12$  for  $t = 5, 6, 7$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 0.93; R_P^Q(3, t) = 0.49; R_P^Q(4, t) = 0.13$  for  $t = 8$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 0.93; R_P^Q(3, t) = 0.53; R_P^Q(4, t) = 0.22$  for  $t = 9, 10$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 0.96; R_P^Q(3, t) = 0.73; R_P^Q(4, t) = 0.46$  for  $t = 11, 12, 13$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 0.96; R_P^Q(3, t) = 0.88; R_P^Q(4, t) = 0.78$  for  $t = 14, 15$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 1; R_P^Q(4, t) = 1$  for  $t = 16$ .

Figure 10 depicts the results in graphic form. It is easy to see that the constraints which take into account the structure of  $M$  significantly change the values of robustness  $R_P^Q(\omega, t)$ . The expected value of 0.8 is obtained as late as the last time unit ( $t = 16$ ) of job execution.

This observation naturally leads to another question: Is it possible to further extend competency framework  $G_{OPT}$  in such a way as to guarantee the expected value of robustness  $R_P^Q(\omega, t)$  across time horizon  $H$ ? When an appropriate synthesis problem CO (17) was solved, a negative result was obtained. The maximum values of  $R_P^Q(\omega, t)$  determined for the so-called full competency framework (each employee has the competency to execute each task:  $g_{k,i} = 1$ ) were as follows:

$R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 0.8; R_P^Q(4, t) = 0.41$  for  $t = 0 \dots 4$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 0.8; R_P^Q(4, t) = 0.49$  for  $t = 5 \dots 10$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 1; R_P^Q(4, t) = 0.93$  for  $t = 11, 12, 13$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 1; R_P^Q(4, t) = 1$  for  $t = 14, 15, 16$ .

This means that a full competency framework does not ensure the expected level of robustness when  $\omega = 4$  employees are absent (see Fig. 11). Robustness  $R_P^Q(4, t)$  at the initial stage of execution of job ( $t = 0 \dots 4$ ) does not exceed 0.41. This means that the available staff of employees  $\mathcal{P}$ , despite having all the necessary competencies, are not able to secure the completion of the project portfolio in the event of an absence of four employees.

Again, the question naturally comes to mind whether it is possible to enlarge the existing staff of employees (by hiring additional employees) in such a way as to build a competency framework that guarantees an expected level of robustness  $R_P^Q(\omega, t)$  across time horizon  $H$ . In order to answer this question, CO (17) was solved. The solution we obtained is shown in Table 3 and Figure 12.

The minimum competency framework  $G_{OPT}'$  (Table 3) shows that staff  $\mathcal{P}$  should be supplemented with two new employees,  $P_9$  and  $P_{10}$ , with eight competencies between them. In addition, the existing employees must improve their qualifications by acquiring five new competencies: employee  $P_1$  should acquire the competency to execute task  $Z_{11}$ ;  $P_4$  competencies for tasks  $Z_3, Z_4$ , and  $Z_{11}$ ; and  $P_5$  the competency for task  $Z_{12}$ . The acquisition of these competencies guarantees robustness  $R_P^Q(\omega, t) \geq 0.8$  across the time horizon  $H$ . More specifically, robustness  $R_P^Q(\omega, t)$  for time points  $t = 0 \dots 16$  along time horizon  $H$  obtains the following values (see Fig. 8):

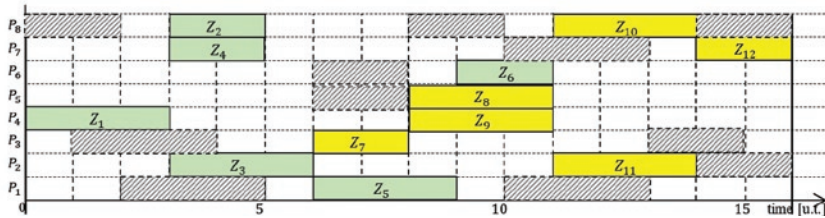


Fig. 9. Schedule for project portfolio  $Q$  from Figure 4, with mask  $M$  defining the unavailability of employees over time horizon  $H$

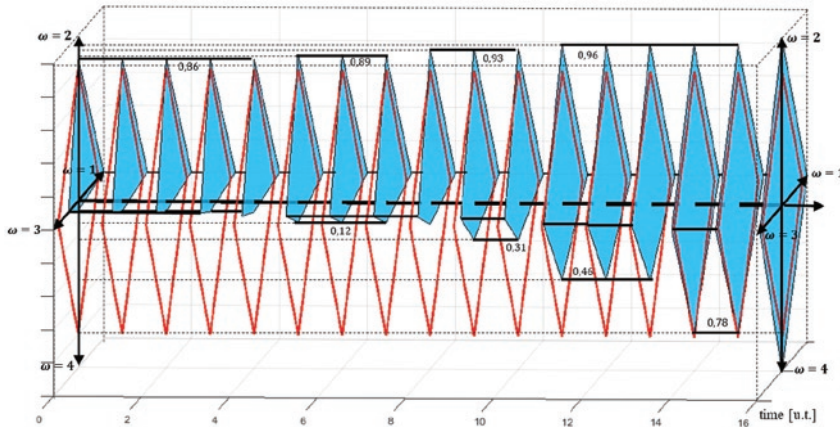


Fig. 10. Graphs of changes in robustness  $R_P^Q(\omega, t)$  of competency framework  $G_{OPT}$  when mask  $M$  of employee unavailability is considered

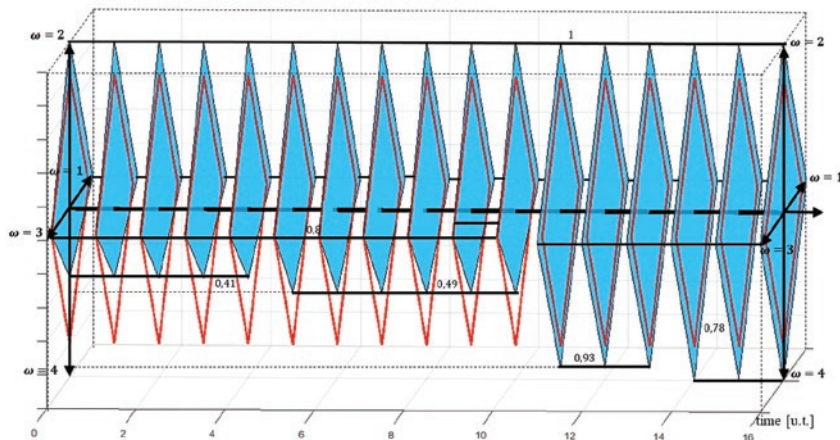


Fig. 11. Curves of change in robustness  $R_P^Q(\omega, t)$  for a full competency framework

Table 3. Competency framework which guarantees  $R_P^Q(\omega, t) \geq 0.8$  (the competencies that the employees need to acquire are given in bold)

$G_{OPT}$	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12
P1	1	0	<b>1</b>	<b>1</b>	1	0	0	1	<b>1</b>	0	1	<b>1</b>
P2	1	0	1	0	0	1	0	0	0	0	1	0
P3	0	1	0	<b>1</b>	1	0	1	1	<b>1</b>	1	0	1
P4	1	0	<b>1</b>	<b>1</b>	0	1	1	0	1	0	<b>1</b>	1
P5	0	<b>1</b>	1	1	1	0	0	1	0	1	0	<b>1</b>
P6	<b>1</b>	1	1	0	0	1	0	0	1	<b>1</b>	1	0
P7	0	0	0	1	1	0	1	1	0	0	0	1
P8	0	1	0	0	0	1	<b>1</b>	<b>1</b>	0	0	1	0
P9	0	<b>1</b>	0	<b>1</b>	<b>1</b>	0	0	<b>1</b>	<b>1</b>	<b>1</b>	0	0
P10	0	0	0	0	0	0	0	<b>1</b>	<b>1</b>	0	0	0

$R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 0.98; R_P^Q(4, t) = 0.8$  for  $t = 0 \dots 4$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 0.98; R_P^Q(4, t) = 0.82$  for  $t = 5 \dots 7$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 0.98; R_P^Q(4, t) = 0.84$  for  $t = 8$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 0.98; R_P^Q(4, t) = 0.86$  for  $t = 9, 10$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 0.98; R_P^Q(4, t) = 0.88$  for  $t = 11, 12, 13$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 1; R_P^Q(4, t) = 0.98$  for  $t = 14, 15$ ,  
 $R_P^Q(1, t) = 1; R_P^Q(2, t) = 1; R_P^Q(3, t) = 1; R_P^Q(4, t) = 1$  for  $t = 16$ .

The results of the experiments demonstrate the competitiveness of the adopted model (which allows one to solve strongly non-linear combinatorial optimization problems) and the computational efficiency of the constraints programming techniques used to analyze it.

The examples provided above illustrate selected options for formulating questions related to different situations in workforce allocation and personnel scheduling processes. The solutions presented, which focus on variants of robust personnel allocation and scheduling, show that the model can be used to design competency frameworks robust to absences of employees for a portfolio of up to four projects. It is worth noting that the concept of mask, introduced in the last example, in addition to solutions robust to employee absenteeism, allows one to search for solutions robust to disruptions caused by the arrival of new jobs during the execution of planned ones.

## 5.2. Quantitative assessment

The proposed approach was evaluated using data of project-driven company carrying out different orders at the same time. The case under consideration relates to a situation in which 49 employees are recruited for six production orders forming the project portfolio  $Q$ . More precisely, the portfolio consist of six jobs:  $Q = \{Q_1, \dots, Q_6\}$  including  $n = 214$  tasks:  $Z = \{Z_1, Z_2, \dots, Z_{214}\}$  with a total of 14,100 hours. The example of parameters  $y_i$  (starting time of task),  $l_i$  (duration of task), and  $w_i$  (set of excluded tasks) are collected in Table 5, and the project portfolio schedule determined by them in Figure 13. Due to the scale of the network of activities describing the order of implementation of individual tasks from among all jobs  $Q_1, \dots, Q_6$ , their graphical representations are omitted in Figure 13. The project portfolio  $Q$  should be completed within a time horizon of 77 days.

Particular tasks are carried out by  $m = 49$  members of the employee team  $\mathcal{P} = \{P_1, P_2, \dots, P_{49}\}$ . The competency framework  $G$  (Table 6.) was determined from surveys results which shows which task which employee:

- can execute:  $g_{k,i} = 1$ ,
- can execute if they gain the missing competences:  $g_{k,i} \in \{0, 1\}$ ,
- cannot execute and cannot gain appropriate competences:  $g_{k,i} = 0$ .

A large number of competences means that an employee can carry out many similar tasks - for example:  $Z_1$  - "assembly and tacking 1";  $Z_2$  - "welding in the vehicle 1";  $Z_6$  - "assembly and tacking 2";  $Z_{74}$  "welding in the vehicle 2";  $Z_{212}$  - "welding in the vehicle 3"; etc. Due to the requirements imposed by the General Data Protection Regulation, data pseudonymisation has been introduced.



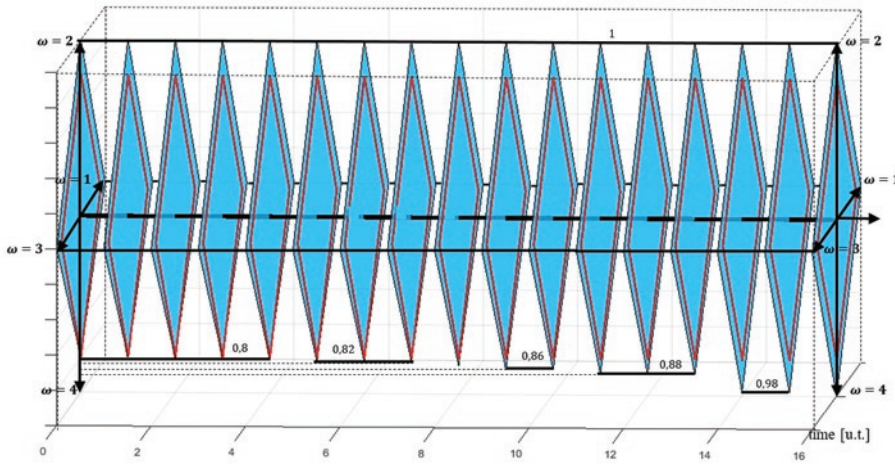


Fig. 12. Graphs of changes in  $R_P^Q(\omega, t)$  for the competency framework shown in Table 3

In addition, a lower ( $s_k$ ) and upper ( $z_k$ ) limit of working time assigned to each employee is collected in Table 7. Assignment of tasks  $X$  sufficient for completion of the given set of tasks  $Z$  following the schedule from Figure 13 is presented in Table 8 and meets the requirements assuming that:

- task  $Z_i$  can only be executed by a competent employee,

Table 5. Set of tasks  $Z$

$Z_i$	$y_i$ [days]	$l_i$ [hours]	$w_i$
$Z_1$	0	80	$\{Z_3, Z_5, Z_{12}, Z_{40}\}$
$Z_2$	10	25	$\{Z_4, Z_{10}, Z_{42}, Z_{50}, Z_{61}\}$
$Z_3$	0	60	$\{Z_1, Z_5, Z_{12}, \dots, Z_{92}\}$
$Z_4$	15	45	$\{Z_2, Z_{10}, Z_{42}, Z_{50}, Z_{61}, Z_{65}\}$
$Z_5$	9	100	$\{Z_1, Z_3, Z_{12}, Z_{40}\}$
...	...	...	...
$Z_{213}$	10	30	$\{Z_{192}, Z_{201}, Z_{207}, \dots, Z_{212}\}$
$Z_{214}$	6	30	$\{Z_{188}, Z_{199}, Z_{210}, Z_{211}\}$

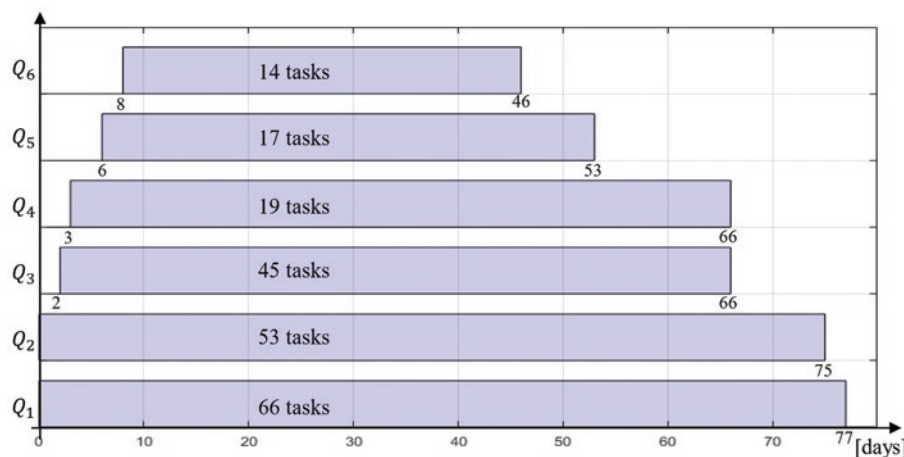


Fig. 13. Schedule of the project portfolio  $Q$

employee working time limits ( $s_k$  and  $z_k$ ) cannot be exceeded.

The above data was used to conduct the analysis of robustness competency framework  $G$  and to determine the competency frameworks protecting the company against selected types of disruptions.

### Analysis of robustness of competency framework $G$

The subject of the analysis is the evaluation of robustness level  $R_P^Q(\omega, t)$  of competency framework  $G$  (Table 6) while implementing a given project portfolio  $Q$  (Fig. 13) and disturbances resulting in simultaneous absence of one to four employees ( $\omega = 1 \dots 4$ ). What is the current value of robustness  $R_P^Q(\omega, t)$  of competency framework  $G$  to the absences of  $\omega = 1 \dots 4$  employees?

The obtained values of robustness levels  $R_P^Q(\omega, t)$  (GUROBI/Intel i7-4770, 8 GB RAM, 10s.) are illustrated in Figure 14. The company is not protected against the threat of employee absence from the entire project portfolio horizon (the expected values of robustness levels are not less than 0.8). Further, the figure shows that just as was the case in the experiments presented in subsection 5.1, robustness value  $R_P^Q(\omega, t)$  increases monotonically with the passage of time and the expected value of 0.8 is reached:

- for  $\omega = 1$  after 38 days of portfolio  $Q$  realization,
- for  $\omega = 2, 3$  after 50 days of portfolio  $Q$  realization (after job  $Q_6$  completion),
- for  $\omega = 4$  on the last day (day 77) of portfolio realization  $Q$ .

In an extreme case (from 1–12 days), robustness to simultaneous absence from four employees is below 0.15. This means that 85% of four employees' absence scenarios in the period from the first to the 12th day will result in non-compliance with accepted deadlines for implemented orders.

### Synthesis of competency frameworks robust to simultaneous absence of $\omega = 1, \dots, 4$ employees

Due to the fact that the expected value of robustness level  $R_P^Q(\omega, t) \geq 0.8$  has not been achieved since the beginning of the project portfolio  $Q$  realization, an attempt was made to synthesize the competence framework guaranteeing this value. Two robustness thresholds were adopted:  $*R_P^Q(\omega, t) = 0.8$  and  $*R_P^Q(\omega, t) = 1$ .

The answer to the following question was sought: Is it possible to enlarge the existing staff of employees (by hiring additional employees) in such a way as to build a redundant competency framework that guarantees an expected level of robustness across time horizon  $H$ ?

The obtained values of robustness levels  $R_P^Q(\omega, t)$  for both considered thresholds are illustrated in Figures 15 and 16. The presented robustness values are conditioned by the need to employ additional staff:

- for  $R_P^Q(\omega, t) \geq 0.8$  (Fig. 15) a team of six employees with 56 additional competences  $\{Z_3, Z_4, Z_{10}, Z_{22}, \dots, Z_{212}\}$  should be employed,

Table 6. Competency framework  $G$  of staff  $\mathcal{P}$

$G$	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$	$Z_6$	...	...	$Z_{74}$	...	$Z_{212}$	$Z_{213}$	$Z_{214}$
$P_1$ : Mills	1	1	0	0	0	0	...	...	0	...	1	0	0
$P_2$ : Garner	0	0	0	0	0	0	...	...	0	...	0	0	0
$P_3$ : Ray	0	0	0	0	0	0	...	...	0	...	0	0	0
$P_4$ : MacPherson	0	0	0	0	0	0	...	...	{0,1}	...	0	0	0
$P_5$ : Burnham	0	0	0	0	0	0	...	...	{0,1}	...	0	0	0
$P_6$ : Davis	0	0	0	0	0	0	...	...	0	...	0	0	{0,1}
$P_7$ : Crockett	{0,1}	{0,1}	{0,1}	0	0	0	...	...	{0,1}	...	0	1	{0,1}
$P_8$ : Hudson	{0,1}	{0,1}	{0,1}	0	0	0	...	...	0	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
$P_{18}$ : Roach	0	0	0	0	0	0	...	...	1	...	0	0	{0,1}
...	...	...	...	...	...	...	...	...	...	...	...	...	...
$P_{47}$ : Fox	0	0	0	1	0	0	...	...	0	...	0	0	0
$P_{48}$ : Porterfield	0	0	0	0	0	0	...	...	0	...	0	0	0
$P_{49}$ : Johnson	0	0	0	0	0	0	...	...	0	...	0	0	0

Table 7. Limit of hours of team  $\mathcal{P}$  members

$G$	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$	$Z_6$	...	...	$Z_{74}$	...	$Z_{212}$	$Z_{213}$	$Z_{214}$
$P_1$ : Mills	1	1	0	0	0	0	...	...	0	...	1	0	0
$P_2$ : Garner	0	0	0	0	0	0	...	...	0	...	0	0	0
$P_3$ : Ray	0	0	0	0	0	0	...	...	0	...	0	0	0
$P_4$ : MacPherson	0	0	0	0	0	0	...	...	{0,1}	...	0	0	0
$P_5$ : Burnham	0	0	0	0	0	0	...	...	{0,1}	...	0	0	0
$P_6$ : Davis	0	0	0	0	0	0	...	...	0	...	0	0	{0,1}
$P_7$ : Crockett	{0,1}	{0,1}	{0,1}	0	0	0	...	...	{0,1}	...	0	1	{0,1}
$P_8$ : Hudson	{0,1}	{0,1}	{0,1}	0	0	0	...	...	0	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
$P_{18}$ : Roach	0	0	0	0	0	0	...	...	1	...	0	0	{0,1}
...	...	...	...	...	...	...	...	...	...	...	...	...	...
$P_{47}$ : Fox	0	0	0	1	0	0	...	...	0	...	0	0	0
$P_{48}$ : Porterfield	0	0	0	0	0	0	...	...	0	...	0	0	0
$P_{49}$ : Johnson	0	0	0	0	0	0	...	...	0	...	0	0	0

Table 8. Assignment  $X$  determined by the schedule from Figure 13

$G$	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$	$Z_6$	...	...	$Z_{74}$	...	$Z_{212}$	$Z_{213}$	$Z_{214}$
$P_1$ : Mills	1	1	0	0	0	0	...	...	0	...	1	0	0
$P_2$ : Garner	0	0	0	0	0	0	...	...	0	...	0	0	0
$P_3$ : Ray	0	0	0	0	0	0	...	...	0	...	0	0	0
$P_4$ : MacPherson	0	0	0	0	0	0	...	...	{0,1}	...	0	0	0
$P_5$ : Burnham	0	0	0	0	0	0	...	...	{0,1}	...	0	0	0
$P_6$ : Davis	0	0	0	0	0	0	...	...	0	...	0	0	{0,1}
$P_7$ : Crockett	{0,1}	{0,1}	{0,1}	0	0	0	...	...	{0,1}	...	0	1	{0,1}
$P_8$ : Hudson	{0,1}	{0,1}	{0,1}	0	0	0	...	...	0	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
$P_{18}$ : Roach	0	0	0	0	0	0	...	...	1	...	0	0	{0,1}
...	...	...	...	...	...	...	...	...	...	...	...	...	...
$P_{47}$ : Fox	0	0	0	1	0	0	...	...	0	...	0	0	0
$P_{48}$ : Porterfield	0	0	0	0	0	0	...	...	0	...	0	0	0
$P_{49}$ : Johnson	0	0	0	0	0	0	...	...	0	...	0	0	0

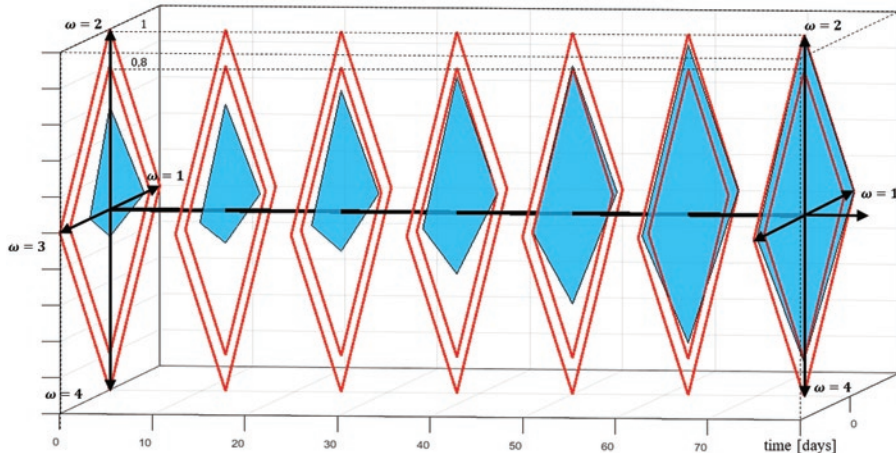


Fig. 14. Robustness  $R_P^O(\omega, t)$  of competency framework  $G$  from Table 6

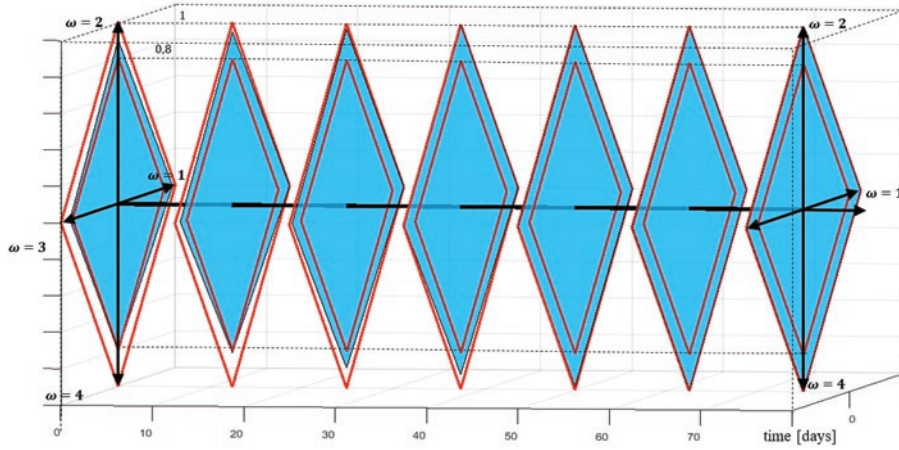


Fig. 15. Graphs of changes in  $R_P^O(\omega, t)$  for competency framework guaranteeing  $R_P^O(\omega, t) \geq 0.8$

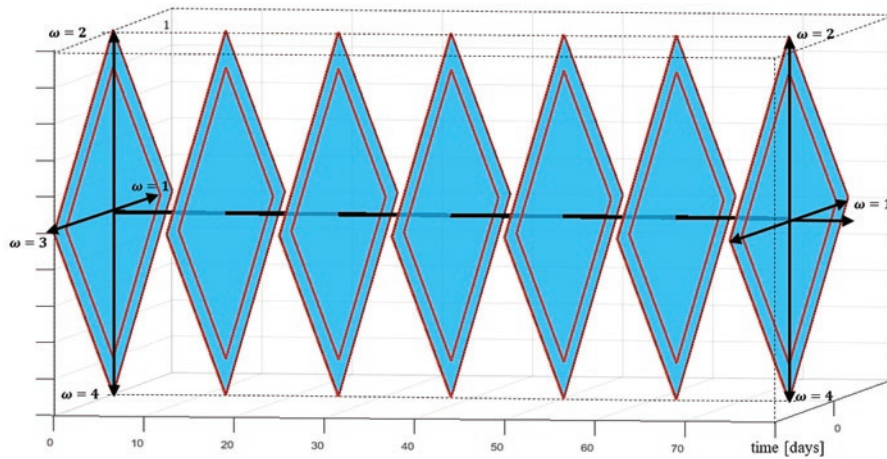


Fig. 16. Graphs of changes in  $R_P^O(\omega, t)$  for competency framework guaranteeing  $R_P^O(\omega, t) = 1$

- for  $R_P^O(\omega, t) = 1$  (Fig. 16) a team of 11 employees with 124 additional competences  $\{Z_6, Z_7, Z_9, Z_{32}, \dots, Z_{214}\}$  should be employed.

The above solutions are examples of selected variants from the set of acceptable solutions received.

According to the received solutions, securing the company against the effects of employee absence (absence of up to four employees at the same time) is conditioned by increasing the staff by six (robust to 80% of possible absence scenarios) and 11 (robust to 100% of possible absence scenarios) additional employees. In the considered case, the process of synthesizing resistant competency framework required more than four hours of calculation. Considering the scale of the project portfolio, this duration is acceptable to the company.

## 6. Conclusions

The proposed method allows one to plan the allocation of production jobs to resources in situations in which the disruptions are caused by employee absences. According to this method, it is necessary to determine which additional (redundant) competencies organizations need to possess in order to compensate for competencies lost as a result of employee absenteeism. The experiments have shown that the method can be effectively used in an online mode to solve small-scale problems in organizational units of up to 30 employees and 60 tasks. It may be possible to increase the scale of the problems solved by using hybrid methods [44] dedicated to models that use sparse data structures.

The conducted experiments were limited to a selected class of competencies occurring in the industrial environment. In general, the proposed model can also be used in other areas requiring management competencies, maintenance management competencies, software skills, and so on. Assessment of the possible implementation of the proposed approach in such areas will be the subject of further research.

The proposed approach can be implemented for example in Decision Support Systems (DSS), Enterprise Resource Planning (ERP) systems [30], used in the online task assignment. Our future work will focus on developing the computational module which can be used as a software overlay for commercially available decision support systems used in human resources management. The functionalities discussed are solutions falling within the scope of human resource controlling [9] aimed at effective staff management while creating transparent rules and procedures for planning, monitoring, and control. It is easy to notice that from the controlling perspective, our method can be used in a broader sense of a digital twin concept [24].

A topic worth considering in terms of the future modification of the model is the assessment of the cost and time consumption of changes in the competency framework. The presented model assumes that the cost/time of each acquired competence is the same. By introducing appropriate cost and time parameters, it will be possible to search for variants of competency frameworks that can also find their economic justification.

### Funding:

This research was funded by the National Science Centre, Poland, grant number 2019/33/N/HS4/00379.

## References

1. Akyol S.D, Baykasoglu A. ErgoALWABP: a multiple-rule based constructive randomized search algorithm for solving assembly line worker assignment and balancing problem under ergonomic risk factors. *J Intell Manuf* 2019; 30: 291-302, <https://doi.org/10.1007/s10845-016-1246-6>.
2. Antosz K. Maintenance - identification and analysis of the competency gap. *Eksplatacja i Niezawodnosc - Maintenance and Reliability* 2018; 20(3): 484-494, <https://doi.org/10.17531/ein.2018.3.19>.
3. Asensio-Cuesta S, Diego-Mas JA, Canos-Daros L, Andres-Romano C. A genetic algorithm for the design of job rotation schedules considering ergonomic and Competence criteria. *International Journal of Advanced Manufacturing Technology* 2012; 60: 1161-1174, <https://doi.org/10.1007/s00170-011-3672-0>.
4. Baladeh AE, Zio E. A Two-Stage Stochastic Programming Model of Component Test Plan and Redundancy Allocation for System Reliability Optimization. *IEEE Transactions on Reliability* 2020, <https://doi.org/10.1109/TR.2020.2974284>.
5. Banaszak Z, Bocewicz G. Declarative Modeling for Production Order Portfolio Scheduling. *Foundations of Management* 2014; 6(3): 7-24, <https://doi.org/10.1515/fman-2015-0014>.
6. Bänziger T, Kunz A, Wegener K. Optimizing human-robot task allocation using a simulation tool based on standardized work descriptions. *J Intell Manuf* 2020, <https://doi.org/10.1007/s10845-018-1411-1>.
7. Brucker P, Qu R, Burke E. Personnel Scheduling: Models and Complexity. *European Journal of Operational Research* 2011; 210(3): 467-473, <https://doi.org/10.1016/j.ejor.2010.11.017>.
8. Chan P, Weil G. Cyclical Staff Scheduling Using Constraint Logic Programming, in: *Practice and Theory of Automated Timetabling III*, Springer Berlin Heidelberg, Lecture Notes in Computer Science 2011; 2079: 159-175, [https://doi.org/10.1007/3-540-44629-X\\_10](https://doi.org/10.1007/3-540-44629-X_10).
9. Dugelova M, Strenitzerova M. How to select appropriate human resource controlling indicators. *CBU International Conference On Innovation, Technology Transfer And Education*, Prague, Czech Republic 2015, <https://doi.org/10.12955/cbup.v3.585>.
10. Dück V, Ionescu L, Kliewer N, Suhl L. Increasing stability of crew and aircraft schedules. *Transportation Research Part C: Emerging Technologies* 2012; 20(1): 47-61, <https://doi.org/10.1016/j.trc.2011.02.009>.
11. Ehrgott M, Ryan DM. Constructing robust crew schedules with bi-criteria optimization. *Journal of Multi-Criteria Decision Analysis* 2002; 11(3): 139-150, <https://doi.org/10.1002/mcda.321>.
12. Felberbauer T, Gutjahr WJ, Doerner KF. Stochastic project management: multiple projects with multi-skilled human resources. *Journal of Scheduling* 2019; 22: 271-288, <https://doi.org/10.1007/s10951-018-0592-y>.
13. Guo S, Wang X, Fu L, Liu Y. How Individual's proactive behavior helps construction sustainability: exploring the effects of project citizenship behavior on project performance. *Sustainability* 2019; 11(24): 6922, <https://doi.org/10.3390/su11246922>.
14. Hazir O, Haouari M, Erel E. Robust scheduling and robustness measures for the discrete time/cost trade-off problem. *European Journal of Operational Research* 2010; 207(2): 633-643, <https://doi.org/10.1016/j.ejor.2010.05.046>.
15. Ingels J, Maenhout B. Optimised buffer allocation to construct stable personnel shift rosters. *Omega* 2019; 82: 102-117, <https://doi.org/10.1016/j.omega.2017.12.006>.
16. Ingels J, Maenhout B. Employee substitutability as a tool to improve the robustness in personnel scheduling. *OR Spectrum* 2017, <https://doi.org/10.1007/s00291-017-0476-0>.
17. Ingels J, Maenhout B. The impact of reserve duties on the robustness of a personnel shift roster: an empirical investigation. *Computers & Operations Research* 2015; 61: 153-169, <https://doi.org/10.1016/j.cor.2015.03.010>.
18. Jafari H, Haleh H. Nurse scheduling problem by considering fuzzy modeling approach to treat uncertainty on nurses' preferences for working shifts and weekends off. *Journal of Optimization in Industrial Engineering* 2019, <https://doi.org/10.22094/joie.2019.576759.159>.
19. Jaśkowski P. Methodology for enhancing reliability of predictive project schedules in construction. *Eksplatacja i Niezawodnosc - Maintenance and Reliability* 2015; 17 (3): 470-479, <https://doi.org/10.17531/ein.2015.3.20>.
20. Kafabad ST, Zanjani MK, Nourelfath M. Integrated planning of operations and on-job training in maintenance logistics networks. *Reliability Engineering and System Safety* 2020; 199, <https://doi.org/10.1016/j.res.2020.106922>.
21. Klimek M, Lebkowski P. Robustness of schedules for project scheduling problem with cash flow optimisation. *Bulletin of the Polish Academy of Sciences. Technical Sciences* 2013; 61(4): 1005-1015, <https://doi.org/10.2478/bpasts-2013-0108>.
22. Korytkowski P. Competencies-based performance model of multi-skilled workers with learning and forgetting. *Expert Systems With Applications* 2017; 77: 226-235, <https://doi.org/10.1016/j.eswa.2017.02.004>.
23. Lecuona JR, Reitzig M. Knowledge worth having in 'excess': The value of tacit and firm-specific human resource slack. *Strategic Management Journal* 2014; 35(7): 954-973, <https://doi.org/10.1002/smj.2143>.
24. Lim K, Zheng P, Chen Ch-H. A state-of-the-art survey of Digital Twin: techniques, engineering product lifecycle management and business innovation perspective. *Journal of Intelligent Manufacturing* 2020; 31: 1313-1337, <https://doi.org/10.1007/s10845-019-01512-w>.
25. Liu Y, Huang HZ, Wang Z, Li YF, Zhang XL. Joint optimization of redundancy and maintenance staff allocation for multi-state series-parallel systems. *Eksplatacja i Niezawodnosc - Maintenance and Reliability* 2012; 14 (4): 312-318.
26. Liu Y, Huang HZ, Wang Z, Li Y, Yang Y. A Joint Redundancy and Imperfect Maintenance Strategy Optimization for Multi-State Systems. *IEEE Transactions on Reliability* 2013; 62(2): 368-378, <https://doi.org/10.1109/TR.2013.2259193>.
27. Moudani W, Mora-Camino F. Solving crew reserve in airlines using dynamic programming approach. *International Journal of Optimization: Theory, Methods and Applications* 2010; 2(4): 302-329.
28. Nielsen I, Wójcik R, Bocewicz G, Banaszak Z. Towards constraint programming driven methodology for online multi-project planning and control. *Information Systems Architecture and Technology*, Wroclaw: Wroclaw University of Technology 2014: 65-74.
29. Niknafs A, Denzinger J, Ruhe G. A Systematic Literature Review of the Personnel Assignment Problem. *Proc. of the International Multi-Conference of Engineers and Computer Scientists 2013 Vol II, MECS 2013*, Hong Kong, 1121-1128, ISSN: 2078-0966.
30. Patalas-Maliszewska J. Assessing the impact of ERP implementation in the small enterprises. *Foundations of Management* 2012; 4(2): 51-62, <https://doi.org/10.2478/fman-2013-0010>.
31. Pawar US, Hanchate DB. Literature review on personnel scheduling. *International Journal of Computer Engineering & Technology* 2013; 4(5), 312-324.
32. Pinedo ML. *Planning and Scheduling in Manufacturing and Services*. In: *Springer Series in Operations Research*, Eds.: Peter W. Glynn

- Stephen M. Robinson, Springer Science Business Media, Inc., New York, 2005, ISBN 0-387-22198-0.
33. Roblek M, Majec M, Benjamin U. Knowledge-based assignment model for allocation of employees in engineering-to-order production In: Knowledge Management Strategies and Applications, Muhammad Mohiuddin, Norrin Halilem, SM Ahasanul Kobir and Cao Yuliang, IntechOpen, 2017, <https://doi.org/10.5772/intechopen.70073>.
  34. Rocha M, Oliveira F, Carravilla M. Cyclic staff scheduling: optimization models for some real-life problems, *Journal of scheduling* 2013; 16(2): 231-242, <https://doi.org/10.1007/s10951-012-0299-4>.
  35. Rosenberger J, Schaefer A, Goldsman D, Johnson E, Kleywegt A, Nemhauser G. A stochastic model of airline operations. *Transportation Science* 2002; 36(4): 357-377, <https://doi.org/10.1287/trsc.36.4.357.551>.
  36. Samson S, Reneke JA, Wiecek MM. A review of different perspectives on uncertainty and risk and alternative modelling paradigm. *Reliability Engineering and System Safety* 2009; 94(2): 558-567, <https://doi.org/10.1016/j.ress.2008.06.004>.
  37. Supian S, Wahyuni S, Subiyanto J. Optimization of Personnel Assignment Problem Based on Traveling Time by Using Hungarian Methods: Case Study on the Central Post Office Bandung. 4th International Conference on Operational Research, IOP Conf. Series: Materials Science and Engineering 300 (2017), <https://doi.org/10.1088/1757-899X/300/1/012005>.
  38. Szwarc E, Bocewicz G, Banaszak Z, Wikarek J. Competence allocation planning robust to unexpected staff absenteeism. *Eksploatacja i Niezawodnosc - Maintenance and Reliability* 2019; 21(3): 440-450, <https://doi.org/10.17531/ein.2019.3.10>.
  39. Szwarc E, Bocewicz G, Bach-Dąbrowska I. Planning of teacher staff competence structure robust to unexpected personnel absence. *IFAC-PaperOnLine* 2019; 52 (13): 2740-2745, <https://doi.org/10.1016/j.ifacol.2019.11.622>.
  40. Topaloglu S, Selim H. Nurse scheduling using fuzzy modelling approach. *Fuzzy Sets and Systems* 2010; 161(11): 1543-1563, <https://doi.org/10.1016/j.fss.2009.10.003>.
  41. Yun-Chia L, Smith AE. An ant colony optimization algorithm for the redundancy allocation problem (RAP). *IEEE Transactions on Reliability* 2004; 53(3): 417-423, <https://doi.org/10.1109/TR.2004.832816>.
  42. Van den Bergh J, Beliën J, De Bruecker P, Demeulemeester E, Boeck L. Personnel scheduling: a literature review. *European Journal of Operational Research* 2013; 226: 367-385, <https://doi.org/10.1016/j.ejor.2012.11.029>.
  43. Volk R. Proactive-reactive, robust scheduling and capacity planning of deconstruction projects under uncertainty. Ph.D. Dissertation, Karlsruher Institut für Technologie, Fakultät für Wirtschaftswissenschaften, KIT Scientific Publishing, Karlsruhe, 2017.
  44. Wikarek J, Sitek P. A Data-Driven Approach to modelling and Solving Academic Teachers' Competencies Configuration Problem. *Intelligent Information and Database Systems* 2019, [https://doi.org/10.1007/978-3-030-14799-0\\_35](https://doi.org/10.1007/978-3-030-14799-0_35).
  45. Woodruffe C. What is meant by competency? In: Boam R., Sparrow P. (Eds.), *Designing and achieving competency*. New York, 1992.

## Induction motor fault classification via entropy and column correlation features of 2D represented vibration data

Indexed by:



Murat Basaran<sup>a</sup>, Mehmet Fidan<sup>a</sup>

<sup>a</sup>Vocational School of Transportation, Department of Motor Vehicles and Transportation Technologies, Eskisehir Technical University, 26140 Odunpazari, Eskisehir, Turkey

### Highlights

- Induction motor bearing defects are classified due to the defect type as well as depth.
- Normalised and partitioned 1D vibration signals are converted into 2D greyscale images.
- 2D discrete wavelet transform is applied to greyscale images to create four sub-images.
- Novel entropy and column correlation-based features are extracted from these sub-images.
- Comparative classification success of the proposed feature extraction method is shown.

### Abstract

Due to long-term use under challenging conditions, the sub-elements of induction motors may suffer certain defects over time. Such defects impair the vibration characteristics of the motors in different ways, depending on the type of defect. Therefore, the change in vibration characteristic provides indicators about the fault type and can be used in preventive maintenance strategies to ensure safe operation of the system. In this work, discrete-time vibration data were transformed into 2-dimensional grey-level images and decomposed into individual components by the Wavelet decomposition method. Features based on entropy and column correlation were extracted from these components and used to classify motor faults by using the Support Vector Machine method implemented by using the Sequential Minimal Optimisation algorithm. When the selected classifier is compared with other popular Machine Learning algorithms, it is observed that motor faults are more successfully classified, and these observations are presented in detail with comparative classification performance results.

### Keywords

entropy, fault diagnosis, support vector machines, wavelet transforms.

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

Electric motors consume nearly half of the electricity supply and may therefore be considered the workhorses of industry. Among all electric motors, induction motors — ranging from fractional horsepower to higher levels of industrial scale — are the most preferred electric motors in industrial applications due to their simple rugged construction, cost-effective price and ease of maintenance. Industrial companies are continuously pushed to continue their operations more reliably and efficiently, and therefore, applying predictive maintenance strategies such as vibration monitoring are essential to ensuring the continuity and quality of the production process.

Bearings contribute to the proper mechanical rotation of motors via their set of spherical or cylindrical rolling elements located between two circular rings called ‘races’ — one inner and one outer. Due to non-ideal operating conditions and ageing, incipient types of bearing defects may occur, which may further deteriorate and propagate on to the races and the rolling elements of the bearings.

The vibration frequencies caused by bearing-related defect types are formulated as functions of rotational frequency and bearing geometry metrics [34]. Bearing-related defects can be broadly classified as outer race defects (ORD), inner race defects (IRD), and ball defects (BD). The characteristic vibration frequencies related to these main defect types are:

$$f_{ORD} = (N_b / 2) f_r [1 - d_b \cos(\beta) / d_p] \quad (1)$$

$$f_{IRD} = (N_b / 2) f_r [1 + d_b \cos(\beta) / d_p] \quad (2)$$

$$f_{BD} = d_p f_r / 2 d_b [1 - (d_b \cos(\beta) / d_p)^2] \quad (3)$$

where  $f_r$  is the rotational speed,  $N_b$  is the number of balls between the races,  $\beta$  is the contact angle of the ball with the races, and  $d_b$  and  $d_p$  are the ball diameter and the pitch diameter, respectively, as depicted in Fig. 1.

These characteristic frequencies are widely used to construct models of impulse trains induced by bearing related defects and the vibration responses are used in various fault classification techniques. However, these characteristic frequencies derived by kinematical relationship based on simple rolling motion and smooth rolling assumptions. But in reality, in the presence of a load between a rolling ball and the races, the contact surface is formed and ball rotates relative to the deformed surface of the races. Due to the loading conditions, the impulse trains show stochastic character rather than being strictly

E-mail addresses: M. Basaran - [murath@eskisehir.edu.tr](mailto:murath@eskisehir.edu.tr), M. Fidan - [mfidan@eskisehir.edu.tr](mailto:mfidan@eskisehir.edu.tr)

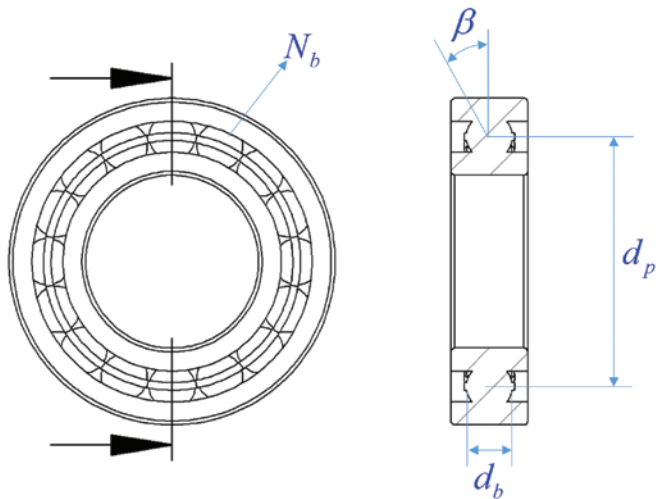


Fig. 1. Ball bearing geometry

periodical. Consequently, the characteristic defect frequencies are strongly depending on bearing metrics, loading conditions and lubrication level that has an important role in the contact angle condition and the relative slippage between balls and the raceway [26]. Therefore, more complex models not solely based on kinematic constraints but considering cyclostationary behaviour of the vibration signals is needed to enhance the effectiveness of the diagnosis.

Monitoring vibration signals is a widely used technique to detect incipient types of bearing defects. In the existing literature, numerous methods have been applied to detect and classify induction motor faults. In one of the first studies on this subject, Ocak and Loparo used the frequency of the fundamental harmonics of the vibration signals to predict motor fault type [27]. In another research study, Ocak and Loparo used the Hidden Markov Model (HMM) to separate motors with inner race defects, outer race defects and ball defects from healthy motors by using the same vibration data used in our research, and obtained maximally 100% accuracies for both class problems [28]. Nandi, Toliyat and Li investigated motor current harmonics instead of vibration, using Fast Fourier Transform (FFT) and clustering the motor faults by using this harmonic knowledge on an Artificial Neural Network (ANN); their results achieved accuracies as high as 93% with an uneasy current-measuring process [25]. Trajin, Regnier and Faucher compared the vibration-based motor fault diagnosis method with the motor stator current-based fault diagnosis method, and concluded that the vibration method is more appropriate under constant speed mode, while the current method is more appropriate for the variable speed mode of induction motors [40]. Immovilli, Bellini, Rubini and Tassoni also compared the vibration-based method with the current-based method, and concluded that although the current-based method is suitable only for low-frequency working conditions, the vibration-based method is suitable for both low- and high-frequency working conditions [17]. Lei and Meng proposed the Symplectic Entropy method with Radial Basis Function (RBF) classifier for vibration-based defect classification for four conditions: normal condition, outer race defect, rolling element defect and inner race defect, and achieved 99.8% test accuracy with 6,000 vibration samples [19]. In more recent research, Li, Wang, Si and Huang applied an entropy-based defect classification method to the same data used in our research, and achieved 98.75% accuracy for ball defect detection and 100% accuracy for inner and outer race defect detection [20]. In another recent study, Zhao, Liu and Meng proposed switchable normalisation semi-supervised generative adversarial networks (SN-SSGAN) for 1D representation of the same data used in our research, and achieved 99.93% test accuracy for four problem classes (normal, inner race defect, outer race defect and ball defect) while splitting data as 80% training to 20% testing [49]. Gan, Zhao and Chow also utilized vibration signals in electrical and mechanical motor fault detection under different frequency conditions

using genetic algorithms and achieved a maximum of 93.96% test accuracy for electrical faults under 35 Hz frequency and a maximum of 96.9% test accuracy for mechanical faults under 45 Hz frequency [9]. One of the most recent studies using vibration signals for diagnostics was presented by Tabaszewski and Szymański, which proposes a set of binary tree-based classification for three valve clearance classes, listed as: tight, optimum and excess [39]. In [39], the classification accuracy achieved was 99%.

Support Vector Machines (SVM) are also applied for motor fault classification problems [5, 10, 13, 23]. Banerjee and Das proposed a hybrid method of Short-Time Fourier Transform (STFT) and Linear SVM (LSVM) for motor fault detection with multiple sensors and achieved 95% test accuracy [5]. Glowacz used LSVM for the classification of healthy motors, motors with faulty rotor bar and motors with two faulty rotor bars by using acoustic data, and achieved a 96.66% total efficiency of recognition of acoustic signal (TEoRoAS) [13]. Gangsar and Tiwari applied one-vs-one multiclass SVM (MSVM) with RBF kernel for classification of nine different electrical and mechanical faults of an induction motor by using vibration and current data both separately and together, and achieved 98.3% test accuracy by using vibration and current data together under the condition of no-load and 20 Hz working frequency [10]. In a more recent study, Mao and Wang proposed an SVM method supported by Multi-Objective Particle-Swarm-Optimisation (MOPSO) and applied it to separate inner race defect data from normal condition data, outer race defect data from normal condition data, and ball defect data from normal condition data — and obtained 99.47%, 100% and 100% respective accuracies for the corresponding two-class problems [23].

One-dimensional time domain vibration signals can be converted into 2D via quantisation of actual values between 0 to 255 for greyscale image representation by using non-overlapping segments as rows of 2D matrices [11]. Representation in 2D has some significant benefits over the regular 1D signal representation [41]. Do and Chong proposed the Scale-Invariant Feature Transform (SIFT) algorithm to detect faults using 2D representation of the vibration signals, and achieved 98.1% accuracy for the eight-class classification problem, wherein the classes are listed as: angular misalignment, bowed rotor shaft, broken rotor bar, faulty bearing, motor unbalance, normal motor, parallel misalignment and phase unbalance [6]. If a priori knowledge of the classes to be recognised exists, texture-based methods can be applied for the classification of the patterns caused by indicators of vibration signals [18, 36, 48]. In one of these texture-based methods, Khan and Kim calculated Local Binary Patterns (LBP) from 2D representation of the vibration data used in our research, derived global histogram of these LBPs, and used the values in the global histogram as input of k-NN classifier to achieve an average classification accuracy of 99.74% [18]. 2D texture analysis based on local binary patterns (LBP) is also applied to vibration signals. Vibration data converted into 2D greyscale images are used as a medium to find discriminating texture features by employing an LBP operator [36]. In the other texture-based research, Zhang, Peng and Li used Convolutionary Neural Network (CNN) for the classification of motor faults from the 2D representation of the vibration data used in our research, and obtained 99.95% accuracy for separations of 30,000 training and 7,500 test data samples, and 98.17% accuracy for separations of 1,500 training and 7,500 test data samples, where each data sample contains 2,400 data points [48]. In one of the most recent studies, Sun and Cao integrated curvature filtering, Histogram of Gradients (HOG) and one-vs-one MSVM for classification of motor faults by using 2D representation of vibration data and achieved a maximum of 98.48% accuracy by using RBF kernel for MSVM [38]. In another recent study, Ma et al. applied Transfer Learning CNN (TLCNN) to 2D representation of the same data used in our research and achieved 99.71% accuracy [22]. Zimnickas, Vanagas, Dambrauskas and Kalvaitis built a test workbench for collecting vibration signals from several different induction motors whose conditions can be listed as: bad bearings, loose mounting, rotor eccentricity, lost phase to motor

and short circuit in stator winding, and they applied a hybrid method of Continuous Wavelet Transform (CWT) and CNN to 2D representations of their collected data [50]. In [50], 97.53% classification accuracy was achieved. Artificial Neural Network (ANN)-based bearing defect classification via vibration spectrum imaging is another application of greyscale to binary image conversion, showing the spectral contents of the translation-variant time-segmented vibration signal, transformed into a spectral image [3].

This article proposes a novel feature extraction method for column correlation and entropy features from Wavelet decomposition of 2D represented vibration. The vibration data of the induction motors used in the experiments are explained in detail in the introduction of Section 2. In Section 2.1, the construction of 2D greyscale images from 1D vibration data is explained. In Section 2.2, the 2D Haar Wavelet decomposition method is explained. Calculations of column correlation features and entropy features are explained in Section 2.3 and Section 2.4, respectively, and the combined feature vector representation is introduced in Section 2.5. In Section 3, the chosen MSVM classifier with Pearson VII (PUK) kernel is introduced. Section 4 presents comparative results for different classifiers using proposed features, and in Section 5 the presented comparative results are analysed.

## 2. Material and feature extraction method

In this work, the proposed feature extraction technique is tested on a publicly available seeded fault data set from Case Western Reserve University (CWRU) Bearing Data Center [1]. The test bench included a 2-horsepower (hp) induction motor, a torque transducer and a dynamometer used to acquire vibration signals using accelerometers, which were attached to the housing with magnetic bases. Single-point defects were deliberately introduced to the test bearings with three different defect diameters. In this work, 12 kHz drive-end vibration data of induction motors with defective bearings with defect diameters of 0.18, 0.36 and 0.54 millimetres of inner race defect, outer race defect located at the 6:00 position and ball defect bearing are used. Vibration data was taken under four different loading conditions: 0, 1, 2 and 3 hp motor loads. Class labels of defect types with different defect diameters are given in Table 1.

Table 1. Defect Descriptions for Classes

Class Label	Defect Description	Defect Depth Level
Class 1	Inner Race Defect	0.18 mm
Class 2	Inner Race Defect	0.36 mm
Class 3	Inner Race Defect	0.54 mm
Class 4	Ball Defect	0.18 mm
Class 5	Ball Defect	0.36 mm
Class 6	Ball Defect	0.54 mm
Class 7	Outer Race Defect	0.18 mm
Class 8	Outer Race Defect	0.36 mm
Class 9	Outer Race Defect	0.54 mm

All loading conditions of motors are included into the classification to propose a classification method that is effective regardless of loading.

Sample 1D vibration data related to inner race defect, ball defect and outer race defect with different defect diameters under 2 hp load are plotted in Figs. 2, 3 and 4, respectively.

### 2.1. Construction of 2D greyscale images

The first step of the proposed feature extraction method is to convert 1D discrete vibration data to 256-grey-level 2D images. The discrete 1D vibration data can be represented by the vector  $V_i$  as in (4) for the  $i^{th}$  experiment, where  $v_i$  denotes the  $j^{th}$  sample of the  $i^{th}$  data.

$$V_i = [v_i[0] \ v_i[1] \ v_i[2] \ \dots \ v_i[L-1]] \quad (4)$$

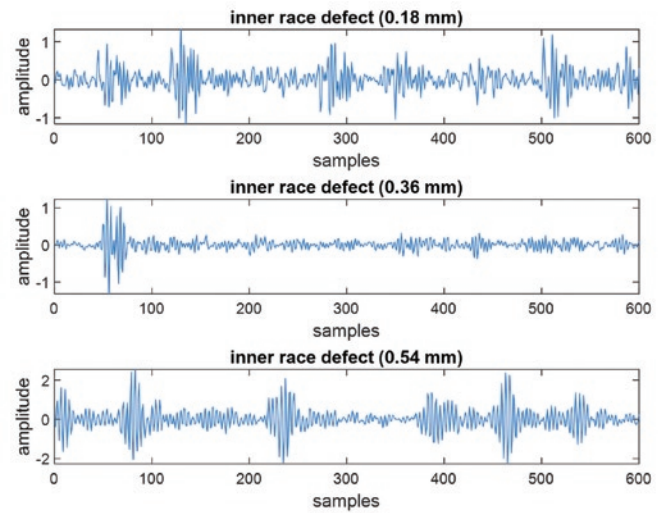


Fig. 2. Sample 1D inner race defect vibration data under 2 hp motor load

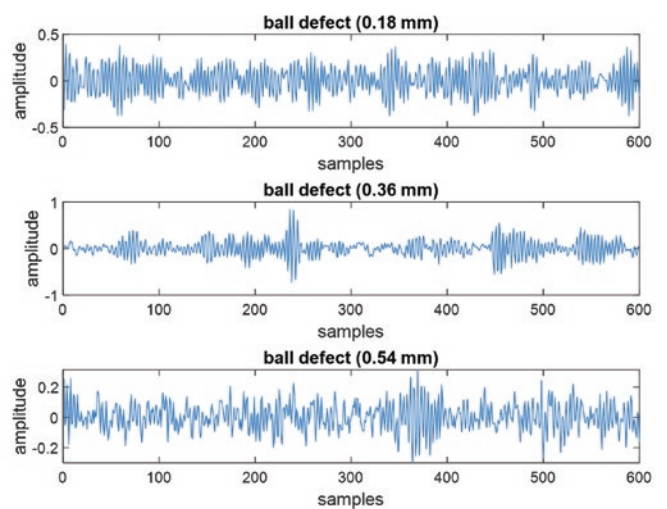


Fig. 3. Sample 1D ball defect vibration data under 2 hp motor load

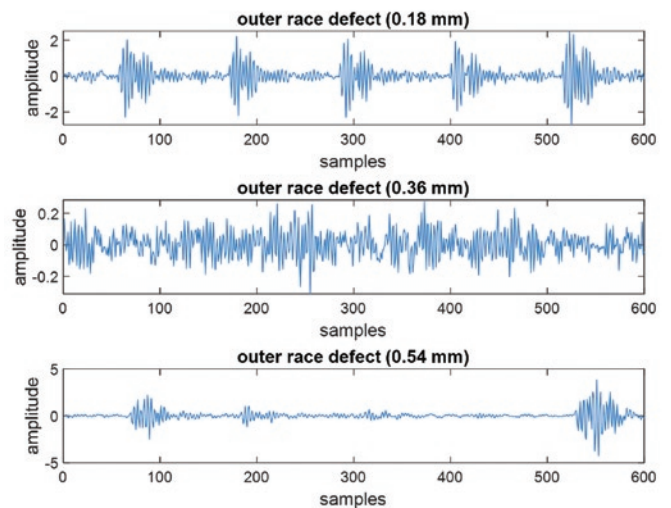


Fig. 4. Sample 1D outer race defect vibration data under 2 hp motor load



where:  $\hat{V}_i$  denotes the normalised vector for the  $i^{th}$  1D vibration data, which is calculated as in (5). The subtraction of minimum valued element of the vector from each sample guarantees that the values will be nonnegative, and the division of each subtraction result by the difference of maximum and minimum elements guarantees that the values will fall within the interval [0,1]:

$$\hat{V}_i = \frac{V_i - \min(V_i)}{\max(V_i) - \min(V_i)} \quad (5)$$

The size selection of the vectors is another crucial step in the construction of 2D greyscale images:

$$L = M \times N \quad (6)$$

In (6),  $L$  denotes the number of samples in a single measurement vector, which is taken as 6,000.  $M$  is the number of rows in the image representation, which is taken as 30, which correspond to 30 cycles.  $N$  is the number of columns in the image representation, which is taken as 200. Since the sampling frequency is 12 kHz, the length  $N$  is taken as 200. The related experiments were carried out at a network frequency of 60 Hz. Thus, each part of the vibration data comprising 200 samples will correspond to a single period of the AC voltage applied at the motor input. In other words, if the sampling frequency is 12 kHz, by setting  $M$  and  $N$  values to 30 and 200 respectively, each greyscale image represents a vibration data recorded for 30 cycles, corresponding to half seconds of data, and the pixels in each row belong to a single period of that part. The normalised vector is converted to 2D 8-bit grey-level images as in (7) and (8).

$$I_i[m, :] = \text{Round}\left(255 \cdot \hat{V}_i[(mN) : ((m+1)N - 1)]\right) \quad (7)$$

$$I_i = \begin{bmatrix} I_i[0, :] \\ I_i[1, :] \\ \vdots \\ I_i[(M-1), :] \end{bmatrix} \quad (8)$$

It should be noted that 60 Hz is the fundamental supply frequency in all experiments, and the  $f_{ORD}$ ,  $f_{IRD}$  and  $f_{BD}$  are the frequencies expressed in kinematical equations in (1), (2) and (3) respectively, are the approximate frequencies of the harmonics those may occur due to the relevant fault type. It should be noted that these harmonics do not always occur in a deterministic way and show stochastic behaviour due to external conditions. In addition, the depth of the defects may cause different harmonics in the vibration spectrum, as a result, the type of fault that cannot be explained solely by the characteristic equations related to fault types. For this reason, the  $N$  parameter used in 2D image representation should be calculated based on the fundamental frequency, not the approximate harmonic frequencies. The effect of harmonic frequencies depending on the fault type on the

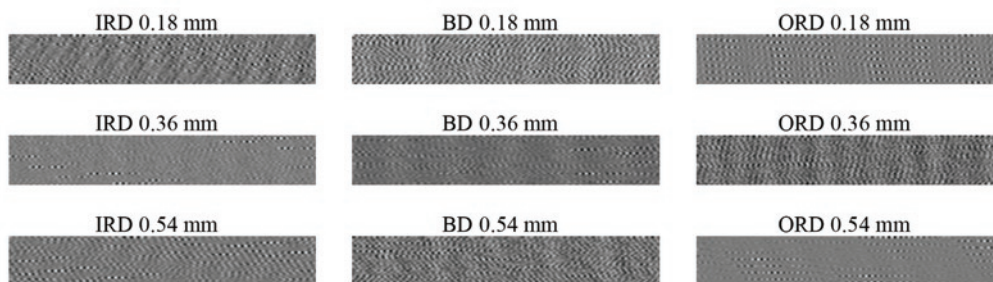


Fig. 5. Sample 2D image representations of vibration data

vibration data will be better understood with the texture structure that will be occurred specific to the fault type in 30x200 pixels 2D images. In this work, column correlation and entropy features which described detailly in Sections 2.3 and 2.4 respectively, suggested to employed together to detect texture type corresponds to fault type. Sample 2D images constructed from vibration data recorded under 2 hp motor loads for inner race defect, ball defect and outer race defect classes with different defect diameters are given in Fig. 5.

## 2.2. 2D Haar Wavelet Decomposition of constructed images

After constructing 20 images per data file downloaded from the CWRU database, single-level 2D Discrete Wavelet Transformation (DWT) is applied to these greyscale images in order to obtain the sub-band images needed for the proposed feature extraction technique [2, 4]. In DWT, High-Pass (HP) and Low-Pass filters (LP) and the down-sampling by two operations are applied along rows of 30 to 200 pixels images. Haar filters are selected for the Low-Pass and High-Pass filters [14, 37]. The High-Pass Haar filter is given in (9) and the Low-Pass Haar Filter is given in (10):

$$H_{Haar} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (9)$$

$$L_{Haar} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (10)$$

The same procedure is applied along the columns of the two sub-images obtained, with a nuance — this time, for the columns, the Low-Pass filter is applied before the High-Pass filter. The outputs at the second step are four greyscale images having sizes of 15 to 100 pixels, corresponding to single-level DWT of the image outputs. The output images are called the vertical, diagonal, approximate and horizontal sub-images, as shown in Fig. 6.

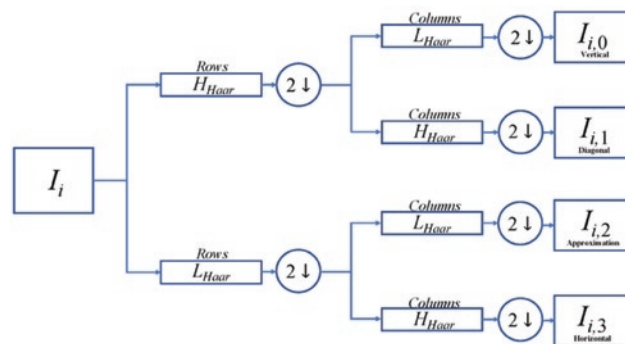


Fig. 6. 2D Haar wavelet decomposition

The order of the sub-images is determined by the strength of the information obtainable from the sub-image, with stronger information for the classification to the sub-images with weaker information.

Sample sub-images for the nine classes of vibration data are presented in Fig. 7 in order of vertical, diagonal, approximate and horizontal components. The approximate image resembles the original image, while remaining images contain

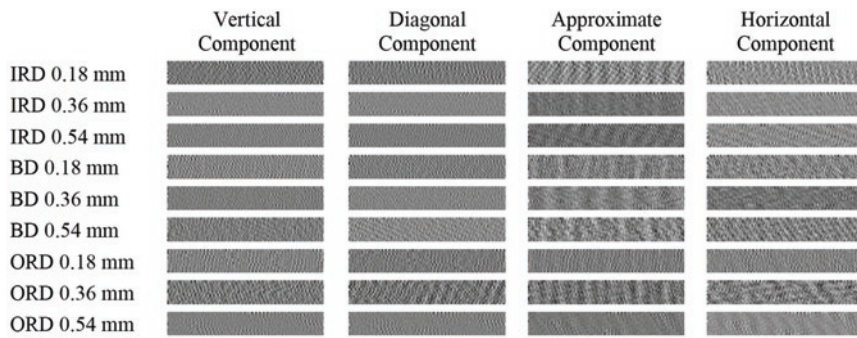


Fig. 7. Sample sub-images obtained for the nine classes of vibration data

detail textures — all of which have some close relationship with the indicators of motor faults.

### 2.3. Extraction of column correlation features

The correlations between two neighbouring columns are evaluated over the sub-images, and the means of these calculated values are used as column correlation features. The detailed analysis of the images obtained by wavelet decomposition reveals that the strongest information is captured in their columns, since each row corresponds to single period of vibration data. Consequently, there are insignificant variations between rows due to the periodicity. The column correlation information contains the potential distinguishing indicators of the fault types, which would be more informative. The  $n^{\text{th}}$  column of the  $j^{\text{th}}$  sub-image for the  $i^{\text{th}}$  image can be represented as in (11):

$$c_{i,j}[n] = I_{i,j}(0:(M_j - 1), n) \quad (11)$$

where  $M_j = M/2$ .

The Pearson correlation coefficient can be expressed as in (12), which can be computed by the covariance of two random variables and the variances of each random variable [30, 46]:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} \quad (12)$$

The column correlation is generated by the mean of Pearson correlations of the neighbouring columns in the sub-images, which can be expressed as in (13):

$$r_{\text{column}}(i, j) = \frac{\sum_{n=0}^{N_j-1} \rho(c_{i,j}[n], c_{i,j}[n+1])}{N_j - 1} \quad (13)$$

where  $N_j = N/2$  [12].

### 2.4. Extraction of entropy features

Entropy is a strong measure for the randomness and texture of the images [15, 45]. However, for the measurement of entropy, the data samples or the pixels of the images should take a countable number of values. After the wavelet transform, the pixels of the sub-images take floating-point values, thereby making each value unique, which is not conducive to entropy calculation. To overcome this problem, the pixels of wavelet sub-images are rounded to the nearest integer value as in (14):

$$Q_{i,j} = \text{Round}(I_{i,j} - \min(I_{i,j})) \quad (14)$$

After this, a histogram is constructed as in (15) and (16):

$$h_{i,j}(k) = \text{count}(s_{i,j}[k]) \quad (15)$$

$$H_{i,j} = [h_{i,j}(0) \ h(1) \cdots \ h(NS_{i,j} - 1)] \quad (16)$$

where  $s_{i,j}[k]$  is the  $k^{\text{th}}$  state in  $Q_{i,j}$  and  $NS_{i,j}$  is the total number of states observed in  $Q_{i,j}$ .

The histogram is an array of numbers of each state, which are generalised as in (17):

$$L_j = M_j \times N_j \quad (17)$$

The probabilities of each state are calculated by (18):

$$P(s_{i,j}[k]) = \frac{H_{i,j}[k]}{L_j} \quad (18)$$

According to the probability of each state, the entropy of each of the wavelet sub-images can be calculated as in (19):

$$\text{entropy}(i, j) = - \sum_{k=0}^{NS_{i,j}-1} [P(s_{i,j}[k]) \log_2(P(s_{i,j}[k]))] \quad (19)$$

### 2.5. Feature vector representation

After the calculation of column correlation and entropy features from sub-band images, features are combined to construct a feature vector of size eight. The overall feature extraction process can be summarised with the flowchart depicted in Fig. 8. For vibration data samples of size 6,000, 20 sets of feature vectors are extracted from each data file downloaded from the CWRU data set. Each set can be considered as a separate experiment and the feature vector for the  $i^{\text{th}}$  experiment can be represented by (20):

$$F_i = \begin{bmatrix} r_{\text{column}}(i, 0) \\ r_{\text{column}}(i, 1) \\ r_{\text{column}}(i, 2) \\ r_{\text{column}}(i, 3) \\ \text{entropy}(i, 0) \\ \text{entropy}(i, 1) \\ \text{entropy}(i, 2) \\ \text{entropy}(i, 3) \end{bmatrix} \quad (20)$$

Feature vectors are normalised as in (21)–(23):

$$\hat{F}_i[k] = \frac{F_i[k] - \min_k}{\max_k - \min_k} \quad (21)$$

where:

$$\min_k = \min([F_0[k] : F_{K-1}[k]]) \quad (22)$$

$$\max_k = \max([F_0[k] : F_{K-1}[k]]) \quad (23)$$

The scatter plots of selected normalised feature vectors for all experiments are presented in Fig. 9 for the Three-Class case, where

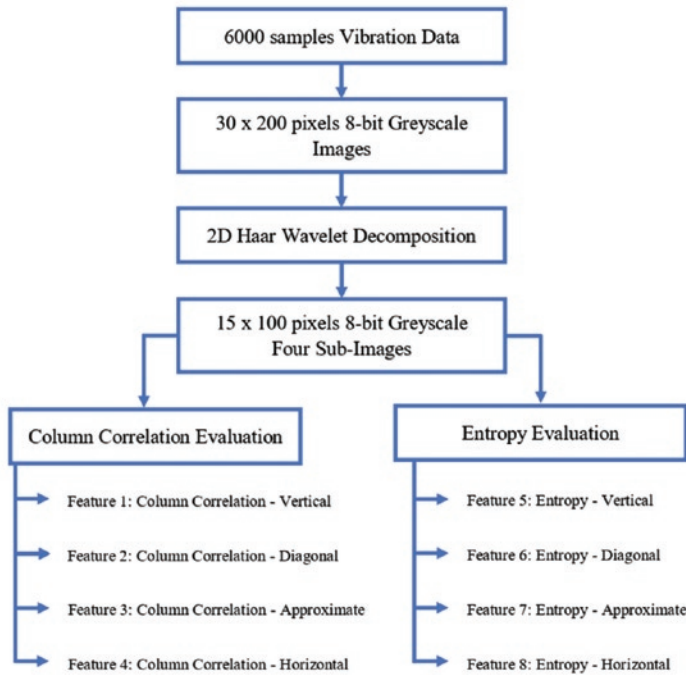


Fig. 8. Feature extraction flowchart

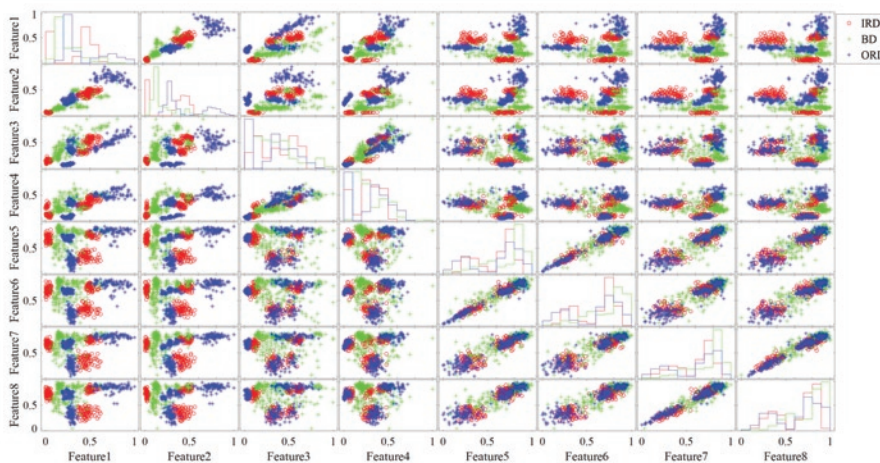


Fig. 9. Scatter plot of normalised features for Three-Class case

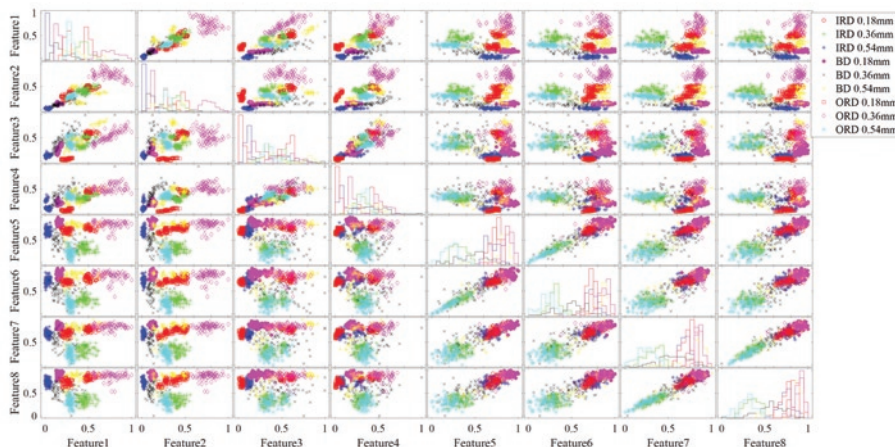


Fig. 10. Scatter plot of normalised features for Nine-Class case

defect diameters are ignored. Similarly, scatter plots of selected normalised feature vectors for all experiments for the Nine-Class case, where each defect diameter was considered as a separate class, as described in Table 1, are presented in Fig 10.

The contribution of column correlation features to the Three-Class problem can be observed by examining the 16 sub-figures in the upper-left quadrant of Fig. 9, which show that in particular, the column correlation coefficients group ball defect and outer race defect better than inner race defect. However, upon examining the 16 sub-figures in the upper-right quadrant and the 16 sub-figures in the lower-left quadrant of Fig. 9, it is seen that the inner race defect can be grouped significantly with the contribution of entropy features.

As seen in the 16 sub-figures in the upper-left quadrant of Fig. 10, it is understood that the most significant contributions of column correlation features to the Nine-Class problem are grouping with 0.18, 0.36, and 0.54 millimetres outer race defects and 0.54 millimetres inner race defects. When the 16 sub-figures in the lower-left quadrant are examined, it can be understood that 0.18, 0.36, and 0.54 millimetres ball defects can be grouped better when column correlation features are used together with entropy features.

When the 16 sub-figures in the lower-right quadrant of Fig. 10 are examined, it is seen that when entropy features alone are used, 0.18, 0.36, and 0.54 millimetres outer race defects can be largely distinguished from each other. However, entropy features can distinguish ball defects from inner race and outer race defects only with the help of column correlation features as seen in the 16 sub-figures in the upper-right quadrant of Fig. 10.

### 3. Support Vector Machine Classifier

The purpose of the Support Vector Machine (SVM) method is to obtain a suitable hyperplane in an N-dimensional space that distinctly separates the classes [35, 43]. The chosen hyperplane should have the maximum distance from both classes.

The original method was developed for two-class problems. However, the method has been adapted to multi-class problems using one of three potential strategies: one vs. one, one vs. all, and the non-heuristic method [24, 44]. In this study, the one vs. one strategy is chosen.

The algorithm is fastened by the Sequential Minimal Optimization (SMO) algorithm [32]. SVM is defined in terms of a quadratic programming (QP) problem. SMO decomposes this QP problem into several QP sub-problems and fastens the algorithm by choosing smallest possible optimisation problem at each step.

The classification accuracy of SVM is improved by the application of the Pearson VII Kernel (PUK) method [29, 31]. Üstün, Melssen and Buydens presented the accuracy of SVM with PUK on various test data for classification [42]. Zhang and Ge used SVM with PUK for the classification of halophilic and non-halophilic proteins [47]. The main reason for using PUK is the nonlinearity of the hyperplanes between classes. The PUK can be expressed as in (24) for two feature vectors:

$$K(\hat{F}_i, \hat{F}_j) = 1 / \left[ 1 + \left[ \left( 2\sqrt{|\hat{F}_i - \hat{F}_j|^2 \cdot \sqrt{2^{1/\omega} - 1}} \right) / \sigma \right]^2 \right]^\omega \quad (24)$$

$$\text{Precision}_i = \frac{TP_i}{\hat{P}_i} \cdot 100\% = \frac{TP_i}{TP_i + FP_i} \cdot 100\% \quad (29)$$

The best classification performance is achieved when  $\omega$  is taken as 1 and  $\sigma$  is taken as 0.5 in (24).

#### 4. Comparative results

For the implementation of the chosen classifiers, the 3.8.2 version of the Workbench for Machine Learning (WEKA) is used [16]. The performance of the proposed feature extraction method with the chosen classifiers is compared with the Bayesian Network classifier [8], Naïve Bayes classifier [33], Naïve Bayes classifier with Kernel Density Estimation (KDE) [21], K-Nearest Neighbours (KNN) Classifier [7] and SMO with Polynomial Kernel [35].

##### 4.1. Comparison metrics

Classification accuracies are compared using the following performance metrics: Accuracy, Macro-Precision, Macro-Recall and Macro-F1 Score.

These performance metrics are calculated from a confusion matrix. A confusion matrix for a multi-class problem can be generalised as in (25):

$$CM \begin{array}{c|cccc} & \hat{P}_1 & \hat{P}_2 & \dots & \hat{P}_N \\ \hline P_1 & TP_{11} & F_{12} & \dots & F_{1N} \\ P_2 & F_{21} & TP_{22} & \dots & F_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_N & F_{N1} & F_{N2} & \dots & TP_{NN} \end{array} \quad (25)$$

In (25),  $\hat{P}_i$  denotes the total number of samples predicted as  $i^{th}$  class and  $P_i$  denotes the exact number of samples in  $i^{th}$  class.  $TP_i$  denotes the number of correctly classified samples as  $i^{th}$  class (number of True Positive samples) and  $F_{ij}$  denotes the number of samples in  $i^{th}$  class that are wrongly classified as  $j^{th}$  class. The Accuracy metric is calculated from a confusion matrix as in (26):

$$\text{Accuracy} = \frac{\sum_{i=1}^N TP_i}{\sum_{i=1}^N P_i} \cdot 100\% \quad (26)$$

The calculations of Precision, Recall and F1 Score need False Positive ( $FP_i$ : the total number of samples that are wrongly classified as  $i^{th}$  class) and False Negative ( $FN_i$ : the total number of samples in  $i^{th}$  class that are wrongly classified as one of the other classes) numbers of each class. The calculation of  $FP_i$  is given in (27), and that of  $FN_i$  is given in (28):

$$FP_i = \sum_{\substack{j=1 \\ j \neq i}}^N F_{ji} \quad (27)$$

$$FN_i = \sum_{\substack{j=1 \\ j \neq i}}^N F_{ij} \quad (28)$$

The Precision of the classifier for a specific  $i^{th}$  class is the fraction of the number of samples correctly classified as  $i^{th}$  class out of the total number of samples classified as  $i^{th}$  class, which is defined by (29):

The Recall of the classifier for a specific  $i^{th}$  class is the fraction of the number of samples correctly classified as  $i^{th}$  class among the total number of samples in  $i^{th}$  class, which is defined by (30):

$$\text{Recall}_i = \frac{TP_i}{P_i} \cdot 100\% = \frac{TP_i}{TP_i + FN_i} \cdot 100\% \quad (30)$$

The F1 Score of the classifier for a specific  $i^{th}$  class is the geometric mean of the Precision and Recall metrics for that class as shown in (31):

$$F1_i = \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{(\text{Precision}_i + \text{Recall}_i)} \quad (31)$$

The data set used in experiments consists of equal numbers of samples belonging to each specific fault class. Therefore, the classification problem dealt with is a balanced multi-class problem. Since the classes are balanced, it is appropriate to use the macro average of the class-based metrics in measuring the overall classification performance.

The overall precision, recall and F1 score performances of the classifiers are measured by Macro-Precision, Macro-Recall and Macro-F1 Score metrics, which are equal to the arithmetic mean of the class-specific precision, recall and F1-scores as shown in (32), (33) and (34) respectively:

$$\text{Macro-Precision} = \frac{\sum_{i=1}^N \text{Precision}_i}{N} \quad (32)$$

$$\text{Macro-Recall} = \frac{\sum_{i=1}^N \text{Recall}_i}{N} \quad (33)$$

$$\text{Macro-F1} = \frac{\sum_{i=1}^N F1_i}{N} \quad (34)$$

##### 4.2. Results for Three-Class case

Table 2 shows overall the Three-Class classification success of the benchmarked classifiers for proposed features. Table 3 and Table 4 show detailed Three-Class performance of the benchmarked classi-

Table 2. Three-Class Classification Accuracy

Classifier	10-fold cross validation	80% split
Bayes Net with Naïve Bayes search	95.1389%	95.1389%
Naïve Bayes without kernel	51.9444%	58.3333%
Naïve Bayes with kernel	95.5556%	95.1389%
KNN	99.0278%	98.6111%
SMO with polykernel	70.5556%	77.7778%
SMO with PUK (Pearson VII kernel)	99.5833%	100.000%

Table 3. Three-Class Detailed Classification Performance Statistics with 10-Fold Cross Validation

Classifier	Macro-Precision	Macro-Recall	Macro-F1
Bayes Net with Naïve Bayes search	95.17%	95.14%	95.14%
Naïve Bayes without kernel	51.85%	51.94%	51.78%
Naïve Bayes with kernel	95.72%	95.56%	95.50%
KNN	99.04%	99.03%	99.02%
SMO with polykernel	69.53%	70.56%	69.59%
SMO with PUK (Pearson VII kernel)	99.58%	99.58%	99.58%

Table 4. Three-Class Detailed Classification Performance Statistics with 80% Split

Classifier	Macro-Precision	Macro-Recall	Macro-F1
Bayes Net with Naïve Bayes search	95.36%	95.15%	95.23%
Naïve Bayes without kernel	58.90%	58.93%	58.90%
Naïve Bayes with kernel	95.17%	95.54%	95.22%
KNN	98.55%	98.77%	98.64%
SMO with polykernel	76.77%	76.84%	75.88%
SMO with PUK (Pearson VII kernel)	100%	100%	100%

fiers according to the comparison metrics for 10-fold cross validation and a split of 80% training set and 20% test set, respectively.

### 4.3. Results for Nine-Class case

Table 5 shows overall Nine-Class classification success of the benchmarked classifiers for proposed features. Table 6 and Table 7 show detailed Nine-Class performance of the benchmarked classifiers according to the comparison metrics for 10-fold cross validation and a split of 80% training set and 20% test set, respectively.

Table 5. Nine-Class Classification Accuracy

Classifier	10-fold cross validation	80% split
Bayes Net with Naïve Bayes search	95.1389%	96.5278%
Naïve Bayes without kernel	95.2778%	97.9167%
Naïve Bayes with kernel	96.2500%	99.3056%
KNN	99.0278%	98.6111%
SMO with polykernel	91.6667%	91.6667%
SMO with PUK (Pearson VII kernel)	99.4444%	100.000%

### 4.4. Evaluation of the Results

When the literature is re-examined, there are similar studies that include feature extraction from the 2D representations of vibration data taken from the CWRU database or a self-designed test rig [3, 6, 18, 22, 36, 38, 48]. In the studies that use CWRU data, image representation of vibration signals differs in some aspects.

In [48], pixels of 60 to 40 sized images were directly used as 2400 inputs of a batch-normalized CNN and 99.95% accuracy for a separation of 30,000 training and 7,500 test data samples and 98.17% accu-

Table 6. Nine-Class Detailed Classification Performance Statistics with 10-fold Cross Validation

Classifier	Macro-Precision	Macro-Recall	Macro-F1
Bayes Net with Naïve Bayes search	95.28%	95.14%	95.11%
Naïve Bayes without kernel	95.46%	95.28%	95.34%
Naïve Bayes with kernel	96.48%	96.27%	96.29%
KNN	99.05%	99.03%	99.02%
SMO with polykernel	93.14%	91.67%	90.90%
SMO with PUK (Pearson VII kernel)	99.45%	99.44%	99.44%

Table 7. Nine-Class Detailed Classification Performance Statistics with 80% Split

Classifier	Macro-Precision	Macro-Recall	Macro-F1
Bayes Net with Naïve Bayes search	96.14%	96.76%	96.26%
Naïve Bayes without kernel	97.92%	98.36%	97.99%
Naïve Bayes with kernel	99.21%	99.47%	99.32%
KNN	98.25%	98.89%	98.50%
SMO with polykernel	91.34%	92.02%	90.19%
SMO with PUK (Pearson VII kernel)	100%	100%	100%

racy is achieved for a separation of 1,500 training and 7,500 test data samples, where each data sample contains 2400 data points.

In [18], average classification accuracy was obtained as 99.74% by LBP features, which is similar to the classification accuracy of our proposed method.

In [3], spectral imaging and filtering with appropriate threshold selection — which is more complicated than the proposed technique — was used to discriminate the same type of faults, and 96.9% average classification accuracy was obtained. However, it focused only on the fault types and neglected the defect depths.

In [38], the sizes of 2D representation of the vibration signal are 100 x 128, which has more data points than our proposed 2D representation (30 x 200) and obtained 98.48% accuracy by using one-vs-one MSVM with RBF kernel, which is less than the accuracy of our proposed method.

Ma et al. achieved 99.71% accuracy by using TLCNN [22]; however, in our proposed method 100% accuracy can be achieved for both Three-Class and Nine-Class problems without the complex structure of CNN and additional knowledge used for transfer learning.

Classifying bearing-related faults, as well as classifying the same type of faults with different defect depths, makes the classification process more challenging when compared to the studies using a self-designed test rig where the fault types are very distinct from each other. In [6], the classification accuracy of the SIFT algorithm, which yields feature vectors of size 128, remained at 98.1%. In [36], the LBP technique, which uses histogram bins of size 256 as feature vectors, was used to discriminate the same fault types as in [6], and 100% accuracy was measured by 4-fold cross validation. Even though the classification accuracy of [36] is similar to that of our proposed method, the feature vector size and computational effort are higher than in our method. On the other hand, we proposed only 8 features based on 2D Discrete Wavelet Transform to discriminate similar bearing-related defects with different depths and achieved remarkably high classification accuracies up to 100%.

## 5. Discussion

The Three-Class classification problem comprises challenge of the ball defect separation from the inner and outer race defects. The difficulty of this separation problem can be observed in Fig. 9. According to the figures, it can be emphasised that the ball defect data is spread over a larger area in the feature space than the data of the other two classes, which may potentially cause confusion. On the other hand, the inner race data bunch along three narrow regions in the feature space, and this situation expedites the distinction of the inner race defect from other defect types. The outer race data bunch along three other narrow regions different from the inner race regions; therefore, the outer race defect type is also distinct from the inner race defect type. The inner race and outer race defects can be separated by linear hyperplanes. However, the ball defect needs a nonlinear hyperplane. As also seen in (3), the characteristic vibration frequency of the ball defect is more complicated than that of other defect types. Therefore, it is understandable that a ball defect is harder to discriminate from other defect types.

According to the overall accuracies of the classifiers on Three-Class classification, the worst performance is observed for the Naïve Bayes without KDE. The simple structure of pure Naïve Bayes is sufficient to avoid both the confusion between ball defect and the other classification types, as well as the confusion between inner and outer race defects. This confusion problem is mostly solved by application of KDE to Naïve Bayes. After the application of KDE, the overall success of the Naïve Classifier increased from 58.33% to 95.14% for the 80% training – 20% testing strategy. In addition, confusion is observed only between ball defect and others after KDE. Bayesian Network shows similar performance with Naïve Bayes with KDE. SMO with simplest polynomial kernel with an exponent of 1 shows relatively worse performance than Naïve Bayes with KDE.

The SMO with polynomial kernel shows confusions mostly between inner race defect and the others. The cause of this confusion could be bunching along three narrow regions which cannot be separated by first-degree polynomial kernel. The performance of SMO is boosted by choosing PUK, which shows 100% sufficiency for Three-Class classification for the 80% train – 20% test strategy. The closest performance to SMO with PUK is observed for KNN classifier, which offers little confusion on ball defect.

## References

1. Case Western Reserve University Bearing Data Center Website. <https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>, (accessed 12-22-2019).
2. Akansu AN, Haddad R A. Multiresolution Signal Decomposition: Transforms, Subbands and Wavelets. Academic Press, 2001: 396-401.
3. Amar M., Gondal I., and Wilson C. Vibration Spectrum Imaging: A Novel Bearing Fault Classification Approach. IEEE Transactions on Industrial Electronics 2015; 62(1): 494-502, <https://doi.org/10.1109/tie.2014.2327555>.
4. Arivazhagan S, Ganesan L. Texture classification using wavelet transform. Pattern Recognition Letters 2003; 24(9-10): 1513-1521, [https://doi.org/10.1016/S0167-8655\(02\)00390-2](https://doi.org/10.1016/S0167-8655(02)00390-2).
5. Banerjee TP, Das S. Multi-sensor data fusion using support vector machine for motor fault detection. Information Sciences 2012; 217: 96-107, <https://doi.org/10.1016/j.ins.2012.06.016>.
6. Do V, Chong UP. Signal Model-Based Fault Detection and Diagnosis for Induction Motors Using Features of Vibration Signal in Two-Dimension Domain. Stroj Vestn-J Mech E 2011; 57(9): 655-666, <https://doi.org/10.5545/sv-jme.2010.162>.
7. Duda RO, Hart PE. Pattern Classification and Scene Analysis. The Library Quarterly 1973; 44(3): 258-259, <https://doi.org/10.1086/620282>.
8. Friedman N, Geiger D, Goldszmidt M. Bayesian Network Classifiers. Machine Learning 1997; 29(2-3): 131-163, <https://doi.org/10.1023/A:1007465528199>.
9. Gan Z, Zhao M-B, Chow T. W. Induction machine fault detection using clone selection programming. Expert Systems with Applications 2009; 36(4): 8000-8012, <https://doi.org/10.1016/j.eswa.2008.10.058>.
10. Gangsar P, Tiwari R. Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms. Mechanical Systems and Signal Processing 2017; 94: 464-481, <https://doi.org/10.1016/j.ymssp.2017.03.016>.
11. Gerek ON, Ece DG. 2-D analysis and compression of power-quality event data. IEEE Transactions on Power Delivery 2004; 19(2): 791-798, <https://doi.org/10.1109/tpwr.2003.823197>.
12. Germen E, Basaran M, Fidan M. Sound based induction motor fault diagnosis using Kohonen self-organizing map. Mechanical Systems and Signal Processing 2014; 46(1): 45-58, <https://doi.org/10.1016/j.ymssp.2013.12.002>.
13. Glowacz A. Recognition of acoustic signals of induction motor using FFT, SMOFS-10 and LSVM. Eksploatacja i Niezawodność -

If the Nine-Class performances of the classifiers with proposed feature extraction methods are analysed, the first pattern that stands out is that the confusions in the Three-Class classification are mostly avoided, especially for the worst classifiers in Three-Class classification, as Naïve Bayes without Kernel and SMO with polykernel. The one-vs-one hyperplanes in the Nine-Class problem have simple structures than the one-vs-one hyperplanes in Three-Class problem, because each level of the inner race and outer race defects are bunched at specific locations, which can be easily separated from other type of defects as seen in Fig. 10. According to Tables 5–7, despite the remarkable improvement in scores, the SMO with polykernel becomes the worst classifier in Nine-Class classification. Naïve Bayes without KDE achieves the success of Bayesian Network, and Naïve Bayes with KDE exceeds the success of Bayesian Network. The best performance in the Nine-Class classification is also observed in chosen SMO with PUK classifier, which is followed by KNN classifier.

If the success of the proposed features in both Three-Class and Nine-Class problems is evaluated over the precision, recall and F1 metrics, there is no significant difference between the precision and recall values, as can be seen in Tables 3, 4, 6 and 7. Because of the closeness of precision and recall performances, the F1 scores are remarkably close to the accuracy scores. The precision, recall and F1 metrics prove that the proposed features are sufficient to obtain consistent and reliable fault classification results.

## 6. Conclusions

To sum up, the proposed feature extraction technique provides strong information, not only for the Nine-Class fault classification problem, but also for the Three-Class fault classification problem. In addition, if these features are supported by a kernel-based classifier with a suitable kernel, they provide great improvement — up to 100% correct classification. The type of kernel is crucial, as observed in the comparison of polykernel and PUK.

In future work, the method can be improved upon in two different ways. First, some additional features can be proposed, especially for separating ball defects from other types of defects. Alternatively, a different kernel can be proposed that is much more suitable for the hyperplanes between separated classes.

- Maintenance and Reliability 2015; 17(4): 569-574, <https://doi.org/10.17531/ein.2015.4.12>.
14. Haar A. Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen* 1910; 69(3): 331-371, <https://doi.org/10.1007/bf01456326>.
  15. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics* 1973; SMC-3(6): 610-621, <https://doi.org/10.1109/tsmc.1973.4309314>.
  16. Holmes G, Donkin A, Witten IH. Weka: A machine learning workbench. Second Australia and New Zealand Conference on Intelligent Information Systems 1994, In: *Proceedings of ANZIIS '94*, <https://doi.org/10.1109/ANZIIS.1994.396988>.
  17. Immovilli F, Bellini A, Rubini R, Tassoni C. Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison. *IEEE Transactions on Industry Applications* 2010; 46(4): 1350-1359, <https://doi.org/10.1109/TIA.2010.2049623>.
  18. Khan SA., Kim J-M. Automated Bearing Fault Diagnosis Using 2D Analysis of Vibration Acceleration Signals under Variable Speed Conditions. *Shock and Vibration* 2016; 2016: 1-11, <https://doi.org/10.1155/2016/8729572>.
  19. Lei M, Meng G, Dong GM. Fault Detection for Vibration Signals on Rolling Bearings Based on the Symplectic Entropy Method. *Entropy* 2017; 19(11), <https://doi.org/10.3390/e19110607>.
  20. Li Y, Wang X, Si S, Huang S. Entropy Based Fault Classification Using the Case Western Reserve University Data: A Benchmark Study. *IEEE Transactions on Reliability* 2019: 1-14, <https://doi.org/10.1109/tr.2019.2896240>.
  21. Liu B, Yang Y, Webb GI, Boughton J. A Comparative Study of Bandwidth Choice in Kernel Density Estimation for Naive Bayesian Classification. *Lecture Notes in Computer Science* 2009; 5476: 302-313, [https://doi.org/10.1007/978-3-642-01307-2\\_29](https://doi.org/10.1007/978-3-642-01307-2_29).
  22. Ma P, Zhang H, Fan W, Wang C, Wen G, Zhang X. A novel bearing fault diagnosis method based on 2D image representation and transfer learning-convolutional neural network. *Measurement Science and Technology* 2019; 30(5): 055402, <https://doi.org/10.1088/1361-6501/ab0793>.
  23. Mao W, Wang L, Feng N. A New Fault Diagnosis Method of Bearings Based on Structural Feature Selection. *Electronics* 2019; 8(12), <https://doi.org/10.3390/electronics8121406>.
  24. Mayoraz E, Alpaydm E. Support Vector Machines for Multi-Class Classification. *IWANN99* 1999, In: *Engineering Applications of Bio-Inspired Artificial Neural Networks*, *Lecture Notes in Computer Science* 1607, <https://doi.org/10.1007/BFb0100551>.
  25. Nandi S, Toliyat HA, Li XD. Condition monitoring and fault diagnosis of electrical motors - A review. *IEEE Transactions on Energy Conversion* 2005; 20(4): 719-729, <https://doi.org/10.1109/Tec.2005.847955>.
  26. Niu L, Cao H, He Z, Li Y. A systematic study of ball passing frequencies based on dynamic modeling of rolling ball bearings with localized surface defects. *Journal of Sound and Vibration* 2015; 357: 207-232, <https://doi.org/10.1016/j.jsv.2015.08.002>.
  27. Ocak H, Loparo KA. Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data. *Mechanical Systems and Signal Processing* 2004; 18(3): 515-533, [https://doi.org/10.1016/S0888-3270\(03\)00052-9](https://doi.org/10.1016/S0888-3270(03)00052-9).
  28. Ocak H, Loparo KA. HMM-Based Fault Detection and Diagnosis Scheme for Rolling Element Bearings. *Journal of Vibration and Acoustics* 2005; 127(4): 299-306, <https://doi.org/10.1115/1.1924636>.
  29. Pearson K. Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 1895; 186: 343-414, <https://doi.org/10.1098/rsta.1895.0010>.
  30. Pearson K. Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 1896; 187: 253-318, <https://doi.org/10.1098/rsta.1896.0007>.
  31. Pearson K. Mathematical Contributions to the Theory of Evolution. X. Supplement to a Memoir on Skew Variation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 1901; 197(287-299): 443-459, <https://doi.org/10.1098/rsta.1901.0023>.
  32. Platt JC. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Microsoft, MSR-TR-98-14, 1998.
  33. Rish I. An Empirical Study of the Naive Bayes Classifier. *IJCAI Workshop on Empirical Methods in AI*. 2001.
  34. Schoen RR, Lin BK, Habetler TG, Schlag JH, Farag S. An unsupervised, on-line system for induction motor fault detection using stator current monitoring. *IEEE Transactions on Industry Applications* 1995; 31(6): 1280-1286, <https://doi.org/10.1109/28.475698>.
  35. Schölkopf B, Smola AJ. *Learning with Kernels*. The MIT Press, Cambridge, Massachusetts, London, England 2001, <https://doi.org/10.7551/mitpress/4175.001.0001>.
  36. Shahriar MR, Ahsan T, Chong U. Fault diagnosis of induction motors utilizing local binary pattern-based texture analysis. *Eurasip J Image Vide* 2013, <https://doi.org/10.1186/1687-5281-2013-29>.
  37. Stanković RS, Falkowski BJ. The Haar wavelet transform: its status and achievements. *Computers & Electrical Engineering* 2003; 29(1): 25-44, [https://doi.org/10.1016/s0045-7906\(01\)00011-8](https://doi.org/10.1016/s0045-7906(01)00011-8).
  38. Sun W, Cao X. Curvature enhanced bearing fault diagnosis method using 2D vibration signal. *Journal of Mechanical Science and Technology* 2020; 34(6): 2257-2266, <https://doi.org/10.1007/s12206-020-0501-0>.
  39. Tabaszewski M, Szymański G. Engine valve clearance diagnostics based on vibration signals and machine learning methods. *Eksplotacja i Niezawodność - Maintenance and Reliability* 2020; 22(2): 331-339, <https://doi.org/10.17531/ein.2020.2.16>.
  40. Trajin B, Régnier J, Faucher J. Comparison between vibration and stator current analysis for the detection of bearing faults in asynchronous drives. *IET electric power applications* 2010; 4(2): 90-100, <https://doi.org/10.1049/iet-epa.2009.0040>.
  41. Tuceryan M, Jain AK. *Texture Analysis*. The Handbook of Pattern Recognition and Computer Vision. World Scientific Publishing Co., 1998: 207-248.
  42. Üstün B, Melsse WJ, Buydens LMC. Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. *Chemometrics and Intelligent Laboratory Systems* 2006; 81(1): 29-40, <https://doi.org/10.1016/j.chemolab.2005.09.003>.
  43. Vapnik V. The Support Vector method. *ICANN'97* 1997, In: *Artificial Neural Networks*, *Lecture Notes in Computer Science* 1327: 261-271, <https://doi.org/10.1007/BFb0020166>.
  44. Weston J, Watkins C. Multi-class Support Vector Machines. Royal Holloway University of London, Egham, Surrey TW20 0EX, England, Technical Report CSD-TR-98-04, 1998.
  45. Weszka JS, Rosenfeld A. An application of texture analysis to materials inspection. *Pattern Recognition* 1976; 8(4): 195-200, [https://doi.org/10.1016/0031-3203\(76\)90039-x](https://doi.org/10.1016/0031-3203(76)90039-x).
  46. Wright S. Correlation and Causation. *Journal of Agricultural Research* 1921; 20(7): 557-585.

47. Zhang G. and Ge H. Support vector machine with a Pearson VII function kernel for discriminating halophilic and non-halophilic proteins. *Comput Biol Chem* 2013; 46: 16-22, <https://doi.org/10.1016/j.compbiolchem.2013.05.001>.
48. Zhang W, Peng G, Li C. Bearings Fault Diagnosis Based on Convolutional Neural Networks with 2-D Representation of Vibration Signals as Input. *MATEC Web of Conferences* 2017; 95, <https://doi.org/10.1051/mateconf/20179513001>.
49. Zhao D, Liu F, Meng H. Bearing fault diagnosis based on the switchable normalization SSGAN with 1-D representation of vibration signals as input. *Sensors* 2019; 19(9): 2000, <https://doi.org/10.3390/s19092000>.
50. Zimmnickas T, Vanagas J, Dambrauskas K, Kalvaitis A. A Technique for Frequency Converter-Fed Asynchronous Motor Vibration Monitoring and Fault Classification, Applying Continuous Wavelet Transform and Convolutional Neural Networks. *Energies* 2020; 13(14): 3690, <https://doi.org/10.3390/en13143690>.



## Planning and management of aircraft maintenance using a genetic algorithm

Mirosław Kowalski<sup>a</sup>, Mariusz Izdebski<sup>b</sup>, Jolanta Żak<sup>b</sup>, Paweł Gołda<sup>a</sup>, Jerzy Manerowski<sup>a</sup>

<sup>a</sup>Air Force Institute of Technology, ul. Księcia Bolesława 6, 01-494 Warsaw, Poland

<sup>b</sup>Warsaw University of Technology, Faculty of Transport, ul. Koszykowa 75, 00-662 Warsaw, Poland

Indexed by:



### Highlights

- Aircraft maintenance is a multi-criteria problem.
- Planning and management of aircraft maintenance is a phase process.
- Aircraft maintenance may be planned by genetic algorithm.
- Maintenance management is the assignment of personnel to a maintenance facility.

### Abstract

The aim of the article was to develop a tool to support the process of planning and managing aircraft (ac) maintenance. Aircraft maintenance management has been presented for scheduled technical inspections resulting from manufacturers' technical documentation for ac. The authors defined the problem under investigation in the form of a four-phase decision-making process taking into account assignment of aircraft to airports and maintenance stations, assignment of crew to maintenance points, setting the schedules, i.e. working days on which aircraft are directed to maintenance facilities. This approach to the planning and management of aircraft maintenance is a new approach, unprecedented in the literature. The authors have developed a mathematical model for aircraft maintenance planning and management in a multi-criteria approach and an optimisation tool based on the operation of a genetic algorithm. To solve the problem, a genetic algorithm was proposed. The individual steps of the algorithm construction were discussed and its effectiveness was verified using real data.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>) aircraft operation, maintenance, multi-criteria optimization, genetic algorithm.

## 1. Introduction

The process of planning and management of aircraft maintenance is one of the key processes determining the effective functioning of an organization (both military and civil) dealing with air transport. Effective planning [14,20] and management of ac maintenance [23,40,47] is aimed at ensuring the continuity of the process of transport tasks.

Aircraft maintenance planning at work has been defined as the determination of maintenance schedules, i.e. the days of directing of the ac to the maintenance facility together with the indication of the facility carrying out the maintenance, while maintenance management has been defined as the assignment of crew (personnel) to the maintenance facility. Maintenance may be performed at the airport where the aircraft is operated or at specialist maintenance facilities located outside the stand of the aircraft. Both the crew assignment and the identification of the maintenance facility is a decision making issue depending on technical factors such as the distance of the airport from the maintenance facility, or economic factors such as the cost of repair at the facility or the cost of maintenance carried out by personnel.

While an aircraft is performing its tasks, it is used [10]. The duration of the ac use depends, among other things, on type of ac and the number of tasks performed, which are determined by the number of flights from a specific airport to which ac is assigned. Each type

of aircraft is limited by the permissible duration of use. Therefore, the comprehensive approach to solving the problem of the aircraft maintenance planning and management proposed by the authors of the article is presented in the form of four phases, Fig. 1.

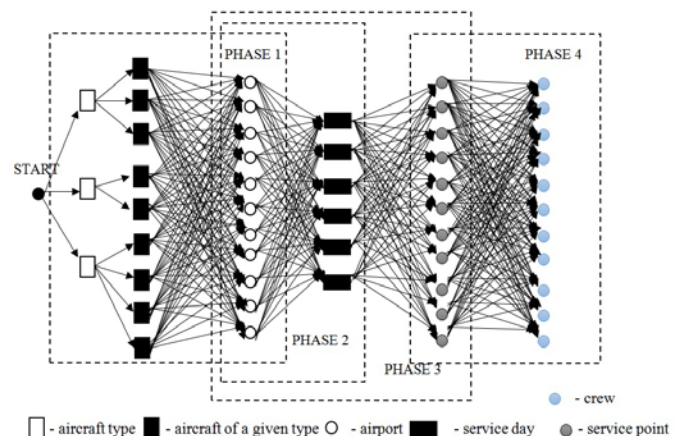


Fig. 1. Decision-making process for aircraft maintenance planning and management

Source: own study

E-mail addresses: M. Kowalski - [miroslaw.kowalski@itwl.pl](mailto:miroslaw.kowalski@itwl.pl), M. Izdebski - [mariusz.izdebski@pw.edu.pl](mailto:mariusz.izdebski@pw.edu.pl), J. Żak - [jolanta.zak@pw.edu.pl](mailto:jolanta.zak@pw.edu.pl), P. Gołda - [pawel.golda@itwl.pl](mailto:pawel.golda@itwl.pl), J. Manerowski - [jerzy.manerowski@itwl.pl](mailto:jerzy.manerowski@itwl.pl)

The phases shown can be described as follows:

- Phase 1 - Assignment of a certain number of aircraft of a given type to a given airport. All ac must be assigned to the airports. The limit is the number of ac at a given airport.
- Phase 2 - Decision to direct the aircraft for maintenance based on the maintenance schedule.
- Phase 3 - Assigning the aircraft to relevant maintenance facilities.
- Phase 4 - Assigning crews to maintenance facilities.

In phase 1 of the aircraft maintenance planning and management, aircraft must be allocated to airports in such a way that the difference between the ac's permissible operating hours and the operating hours resulting from the tasks assigned is maximum. This guarantees a long service life of the aircraft and thus fewer decision-making situations resulting from directing the ac to the maintenance facility. The assignment of ac to airports with different flight demand intensity affects the service life of a particular aircraft and thus the moment/time when maintenance is required (Phase 2). This emphasizes the fact that the phases shown in Figure 2 are closely interrelated and the decisions taken in each phase influence the decisions in the subsequent phases. The main factor influencing the process of ac maintenance planning is the time (day, month) of making a decision to direct the ac to the maintenance facility. This decision may be made within the permissible time of the ac use, in case of exceeding the calendar operating life, maintenance must be carried out absolutely. In the case of ac maintenance planning and management, it is important to indicate such time, e.g. the date on which the aircraft is being directed to the maintenance facility. The decision to direct the aircraft to the maintenance facility when it is fit for performance of tasks is essential. The ac must be directed to the maintenance facilities in such a way as to ensure that a minimum number of aircraft capable of completing all the tasks assigned are available each day.

The number of crews assigned to a maintenance facility depends on the number of stations intended for performance of ac maintenance. The crews may serve different maintenance facilities depending on the maintenance demand. The function of the criterion in the assessment of the crews assigned to the maintenance facility is the cost of the maintenance depending on the type of maintenance performed.

The mathematical model for the management of ac maintenance developed in point 3 supports the assignment of the aircraft type to an airport by minimising its intensity of use, the assignment of aircraft for repair, the assignment of the repair crew and the determination of optimum moments at which maintenance is to be performed. In order to determine ac maintenance schedules, an optimization tool based on the operation of a genetic algorithm was developed, determining the analysed problem in a multi-criteria approach. Genetic algorithms are representatives of heuristic algorithms. The necessity of applying heuristic algorithms in the analysed problem is dictated by the fact that the computational complexity of the problem of planning and managing aircraft maintenance increases exponentially, non-linearly (introduction of an additional airport, additional maintenance facility, additional crew generates exponential increase of decision variables solving the given mathematical model). For this reason, classical methods and algorithms solving linear decision-making problems do not work in this problem. The exponential nature of the problem classifies it into problems described in the literature as NP-Hard[41].

The presented innovative optimization tool based on a genetic algorithm solves the examined problem in a multi-criteria approach according to the adopted global criterion function consisting of partial criterion functions. The application of classical multi-criteria optimization algorithms is difficult due to two aspects [5]: the problem is NP-Hard and in the presented mathematical model the decision variables are variables of different types (binary and integer). It is therefore advisable to develop new optimization tools, adequate to the solution of the problem under investigation.

The following thematic issues were distinguished in the article: a thorough analysis of the state of the art on the issue of ac maintenance

planning and management was carried out, a mathematical model was presented with defined input data, variable decisional constraints and criterion functions, a genetic algorithm was described and its correctness was verified on real data.

## 2. Literature review

In the context of optimisation problems, aircraft maintenance management is a complex decision-making issue consisting of the problem of allocation of resources to tasks [32], which may concern different transport areas [10, 17, 19, 18] and the problem of scheduling repair actions for individual aircraft over a certain period of its operation [36]. In the case of aircraft maintenance, the issue of allocation concerns the assignment of aircraft to airports where transport tasks are carried out [3,8], the assignment of aircraft to maintenance facilities where maintenance tasks are carried out [9, 31, 38] and the assignment of crew to maintenance facilities [1]. The issue of assignment of resources to tasks as well as construction of work schedules are optimization problems widely discussed in operational research or graph theory. In general, the assignment of resources to tasks consists in such assignment of resources, e.g. equipment, employees to the commissioned tasks, that the profit resulting from the assignment of these resources to tasks is optimal. Profit is usually defined as the cost or time of task performance. Limitations result from the number of resources allocated or tasks performed. The problem with schedule preparation, on the other hand, is to arrange the tasks for individual contractors in such a way that the profit also aims at optimal value. It is therefore crucial to determine the start and end points of the tasks. These points can be defined by hours, days or months.

The problem of ac maintenance planning and management in the literature is mainly analysed in the context of problems connected with planning and organization of the aircraft operation and assigning it to particular tasks in situations of failure of the aircraft performing current transport tasks [37, 39]. Maintenance planning is defined as the determination of actions necessary to restore the aircraft to operational condition such as ordering missing parts, assigning staff to maintenance operations [36]. Maintenance of aircraft is often determined on the basis of reliability models of damage of individual components leading to the shutdown of the aircraft. At work [46, 51] the authors have developed a model for assessing the reliability of aircraft analysing the periods in which the aircraft will operate without failure. Further models assessing the reliability of aircraft are described in [42, 43]. The linear model for damage forecasting and maintenance schedule setting is presented in publications [11, 12] and a stochastic model of damage forecasting for aircraft with a non-linear target function taking into account the relationship between repair costs and failure costs is presented in the publication [6]. A risk analysis and a study of the impact of the maintenance organisation on task delays, flight cancellations and safety during combat are presented in [28, 29, 35]. An important aspect of aircraft maintenance planning is to organise maintenance in such a way as to be able to maintain continuity in the performance of tasks, so it is important to take into account the availability of resources in the planning process [22, 26, 37]. In the publication [34] the authors have developed an algorithm for aircraft routing, taking into account reliability aspects, in order to mitigate the impact of unexpected events, which is important for planning and organising aircraft maintenance. Maintenance planning is one of the stages in the development of flight schedules that take into account the time of departure and arrival of the aircraft at the airport, the time of transfer of the crew, rest times as well as the time of possible repair or maintenance [38, 39]. Aircraft maintenance involves assigning personnel to specific maintenance activities [2, 44]. The quality of maintenance depends on human errors, maintenance management models in terms of identifying staff errors are described in [25]. The optimisation of personnel planning for aircraft operation using integer programming models is presented in [1]. A model for the allocation of technical personnel to maintenance services, taking into account their

qualifications, their maintenance certificates, their skills, is presented in [48, 49]. The application of multi-criteria support in the selection of the decision to allocate service crews to maintenance services using the AHP method is described in [4], while the use of evolutionary algorithms to determine the multi-criteria assignment problem is described in [33]. Computer simulations are carried out to model the aircraft maintenance [1, 13], which are an effective tool for assessing the quality of the maintenance provided.

After analysing the literature, it can be concluded that maintenance publications mostly focus on the stage of determining aircraft performance reliability, i.e. determining the probability of failure, the maintenance organisation stage is reduced to the procurement of materials and spare parts only [7, 21] and the assignment of staff to the maintenance without taking into account the mobility of maintenance staff to other maintenance facilities.

Having analysed the literature, it can also be concluded that, most of the publications concern random maintenance events, the management of maintenance resulting from scheduled technical inspections as defined in the aircraft manufacturer's technical documentation is not analysed.

Many authors analyse the issues of damage reliability and spare parts order planning using artificial neural networks [27, 24]. The use of neural networks in aircraft maintenance planning and management may be the subject of further research.

### 3. Mathematical model for aircraft maintenance planning and management

In order to develop a mathematical model, the following determinations of data and parameters of the research problem were introduced:

- $S = \{1, \dots, s, \dots, S\}$  - a set of aircraft types,
- $N(s)$  - a set of a given aircraft type numbers,
- $n$  - number of the subsequent aircraft of the type,  $n=1, 2, \dots, N(s)$ , where  $N(s)$  number of aircraft of type  $s$ ;
- $O = \{1, \dots, o, \dots, O\}$  - a set of maintenance types;
- $V = \{1, \dots, v, \dots, V\}$  - a set of airports;
- $V1 = \{1, \dots, v1, \dots, V1\}$  - a set of aircraft maintenance facilities;
- $Z = \{1, \dots, z, \dots, Z\}$  - a set of crew numbers;
- $t$  - day of operation of the aircraft,  $t \in T$  ;
- $te(s, n)$  - the permissible operating time of the aircraft type as specified by the manufacturer (number of flight hours);
- $tp(s, n)$  - the lower limit of the aircraft operation after which the aircraft can be directed for maintenance (number of flight hours);
- $tk(s, n)$  - the upper limit of the aircraft operation before which the aircraft should be directed for maintenance (number of flight hours);
- $a(s, n, v, t)$  - number of flight hours (flight intensity) to be performed by the aircraft of a given type on a given day of operation at a given airport, value determined on the basis of the schedule of planned flights (number of flight hours);
- $N1(s, v)$  - the number of aircraft of the type necessary for the tasks to be performed at the airport, this number shall include the surplus resulting from the decommissioning of aircraft at the expense of maintenance;
- $N2(v1, o)$  - number of maintenance stations at a given maintenance facility intended for a given maintenance service;
- $N3(s, v, t)$  - required number of aircraft on a given day of operation at a given airport;

- $d((s, n), v, t)$  - the average distance travelled on a given working day by the aircraft of a given type assigned to a given airport;
- $d1(v, v1)$  - the distance between the airport and the maintenance facility;
- $ks(s, n)$  - unit cost of fuel consumption for given aircraft;
- $ko(s, o, v1)$  - the cost of operating a given maintenance station for a given type of vessel carrying out a given service;
- $to(o, s, n)$  - the maintenance time for aircraft of a given type, expressed in days, assumed by a given airport to which the aircraft is assigned; this time is needed to establish the aircraft's continuous operation schedules for each working day;
- $ta(s, v1, o)$  - the time of the aircraft's stay at a particular maintenance facility providing the maintenance service, this time is determined by the maintenance facility;
- $Ddop(v, v1)$  - the accepted distance of allocation to a given maintenance facility from a given airport;
- $kz(z, v1, o)$  - the cost of the execution of a given maintenance service by a given crew at a given maintenance facility;

The problem is, as already pointed out in item 1 of the Article, planning and managing aircraft maintenance in such a way that aircraft of a particular type can be assigned to a particular airport with a view to minimising the intensity of its operation and identify the maintenance facilities for aircraft with the assignment of the maintenance crew and the determination of the optimum times for maintenance to be carried out. For such a problem three types of binary variables and one type of natural variable, i.e. variable forms, were defined:

- $X1 = [x1((s, n), v) : x1((s, n), v) \in \{0, 1\}; s \in S, n \in N(s), v \in V]$  - a binary variable specifying the assignment of ac of a particular type to a particular airport,  $x1((s, n), v) = 1$  when ac of a certain type is assigned to a particular airport,  $x1((s, n), v) = 0$  otherwise;
- $X2 = [x2((s, n), o, v, v1, t) : x2((s, n), o, v, v1, t) \in \{0, 1\}; s \in S, n \in N(s), o \in O, v \in V, v1 \in V1, t \in T]$  - a binary variable specifying the assignment of ac of a particular type to a maintenance facility at a particular airport on a particular working day for the provision of a service,  $x2((s, n), o, v, v1, t) = 1$  when a given ac type is assigned to a maintenance facility that carries out the service of that type on a specific working day,  $x2((s, n), o, v, v1, t) = 0$  otherwise;
- $TPOCZ = [tpocz(s, n) : tpocz(s, n) \in \mathbb{N}, s \in S, n \in N(s)]$  - a natural type variable specifying the day in the schedule, e.g. the first, second day of directing the aircraft of a particular type to a maintenance facility;
- $X3 = [x3(z, v1, t, o) : x3(z, v1, t, o) \in \{0, 1\}, t \in T, o \in O]$  - a binary variable specifying the assignment of the crew to a maintenance facility on a particular day, performing the particular type of the maintenance service,  $x3(z, v1, t, o) = 1$  when the crew has been assigned on a given working day to a maintenance facility carrying out a particular type of service,  $x3(z, v1, t, o) = 0$ , otherwise.

The limitations and boundary conditions of a solution to a decision-making problem concern:

- the allocation of the number of aircraft of a given type to a given airport:

$$\forall s \in S, \forall v \in V \quad \sum_{n \in N(s)} x1((s, n), v) = N1(s, v) \quad (1)$$

- particular aircraft that can only be assigned to one airport:

$$\forall s \in \mathbf{S}, \forall n \in \mathbf{N}(s) \quad \sum_{v \in \mathbf{V}} x1((s, n), v) = 1 \quad (2)$$

- the maintenance must be carried out within a given period of time:

$$\forall s \in \mathbf{S}, \forall n \in \mathbf{N}(s)$$

$$\sum_{t=1}^{tpocz(s,n)} \alpha(s, n, v, t) \geq tp(s, n) \quad (3)$$

$$\sum_{t=1}^{tpocz(s,n)} \alpha(s, n, v, t) \leq tk(s, n) \quad (4)$$

- the allocation of the number of aircraft to a given maintenance facility carrying out the type of service:

$$\forall v \in \mathbf{V}, \forall v1 \in \mathbf{V}1, \forall t \in \mathbf{T}, \forall o \in \mathbf{O} \\ \sum_{s \in \mathbf{S}} \sum_{n \in \mathbf{N}(s)} x2((s, n), o, v, v1, t) \leq N2(v1, o) \quad (5)$$

- the distance from the allocation of aircraft to a given maintenance facility:

$$\forall v \in \mathbf{V}, \forall s \in \mathbf{S}, \forall v1 \in \mathbf{V}1, \forall v \in \mathbf{V}, \forall n \in \mathbf{N}(s), \forall o \in \mathbf{O}, \forall t \in \mathbf{T} \\ x2((s, n), o, v, v1, t) \cdot d1(v, v1) \leq Ddop(v, v1) \quad (6)$$

- the maintenance time assumed by the airport must be greater than or equal to the maintenance time offered by the maintenance facility:

$$\forall v \in \mathbf{V}, \forall s \in \mathbf{S}, \forall v1 \in \mathbf{V}1, n \in \mathbf{N}(s), o \in \mathbf{O}, t \in \mathbf{T} \\ x2((s, n), o, v, v1, t) \cdot ta(s, v1, o) \leq to(o, s, n) \quad (7)$$

- the required number of aircraft carrying out tasks on a given working day:

$$\forall v \in \mathbf{V} \\ \forall v \in \mathbf{V}, \forall s \in \mathbf{S}, \forall v1 \in \mathbf{V}1, \forall t \in \mathbf{T}, \forall o \in \mathbf{O} \\ N1(s, v) - \sum_{n \in \mathbf{N}(s)} x2((s, n), o, v, v1, t) \geq N3(s, v, t) \quad (8)$$

- the required number of crews at a given maintenance facility:

$$\forall t \in \mathbf{T}, \forall v1 \in \mathbf{V}1, \forall o \in \mathbf{O} \\ \sum_{z \in \mathbf{Z}} x3(z, v1, t, o) \leq N2(v1, o) \quad (9)$$

- the assignment of crews to aircraft, the number of aircraft subject to maintenance at a given facility shall be equal to the number of assigned crews:

$$\forall t \in \mathbf{T}, \forall v1 \in \mathbf{V}1, \forall v \in \mathbf{V}, \forall o \in \mathbf{O} \\ \sum_{z \in \mathbf{Z}} x3(z, v1, t, o) = \sum_{s \in \mathbf{S}} \sum_{n \in \mathbf{N}(s)} x2((s, n), o, v, v1, t) \quad (10)$$

- assignment of crews to aircraft maintenance facilities, crews may be assigned to only one maintenance facility on a given working day:

$$\forall o \in \mathbf{O}, \forall t \in \mathbf{T} \\ \sum_{v1 \in \mathbf{V}1} \sum_{z \in \mathbf{Z}} x3(z, v1, t, o) = 1 \quad (11)$$

The values of decision variables are therefore sought for the global function of the character criterion:

$$FGLOBA = (F1(X1), F2(X1), F3(X2), F4(X3)) \quad (12)$$

to take the extreme, with the individual partial functions of the criterion defined as follows:

1. Minimising aircraft operating time at all airports, the greater the difference between the permitted operating time of an aircraft type and the operating time at a given facility, the longer the operating time of individual aircraft:

$$F1(X1) = \sum_{s \in \mathbf{S}} \sum_{n \in \mathbf{N}(s)} \sum_{v \in \mathbf{V}} x1((s, n), v) \cdot \left[ te(s, n) - \sum_{t \in \mathbf{T}} \alpha(s, n, v, t) \right] \rightarrow \max \quad (13)$$

2. Fuel consumption:

$$F2(X1) = \sum_{s \in \mathbf{S}} \sum_{n \in \mathbf{N}(s)} \sum_{v \in \mathbf{V}} \sum_{t \in \mathbf{T}} x1((s, n), v) \cdot ks(s, n) \cdot d((s, n), v, t) \rightarrow \min \quad (14)$$

3. Costs associated with the operation of aircraft, i.e. the cost of transport plus the costs associated with the commissioning of the maintenance station:

$$F3(X2) = \sum_{s \in \mathbf{S}} \sum_{n \in \mathbf{N}(s)} \sum_{v1 \in \mathbf{V}1} \sum_{o \in \mathbf{O}} \sum_{t \in \mathbf{T}} x2[(s, n), o, v, v1, t] \cdot [ks(s, n) \cdot 2d1(v, v1) + ko(s, o, v1)] \rightarrow \min \quad (15)$$

4. The cost of assigning the crew to a maintenance facility:

$$F4(X3) = \sum_{v1 \in \mathbf{V}1} \sum_{o \in \mathbf{O}} \sum_{t \in \mathbf{T}} \sum_{z \in \mathbf{Z}} x3(z, v1, t, o) \cdot kz(z, v1, o) \rightarrow \min \quad (16)$$

## 4. Genetic algorithm for aircraft maintenance planning and management

### 4.1. General assumptions

Genetic algorithms have been applied as a practical optimization tool in complex decision-making problems, e.g. in the problem of a travelling salesman [30, 45], assignment of tasks to contractors [15] setting the work schedules [50] or in the problems of location [16]. The main advantages of genetic algorithms that dominate over other optimization algorithms are the ability to search for the optimal point not from a single location in the search plane, but from multiple locations and to rely on information determined by the adaptation function rather than derivatives. Based on function values, criterion is a valuable advantage of genetic algorithms. It should be stressed that the genetic algorithm is one of the heuristic methods that provide near-optimal solutions. Despite these inconveniences, genetic algorithms are successfully used in optimization problems. The genetic algorithm is an iterative algorithm, so the generated solution improves with each passing to the next iteration. Each individual in the population is assessed according to its adaptation function. The stages of the genetic algorithm developed for aircraft maintenance optimisation can be presented as:

- Stage 1. Determining the structure processed by the algorithm. A matrix structure was used to represent the chromosome in allocation problems in transport issues. Such a structure of input data works successfully in complex decision-making problems, e.g. in the problem of a travelling salesman. The matrix structure determines all decision variables developed in the mathematical model. The matrix structure of the chromosome was randomly generated according to developed algorithms taking into account all the limitations of the mathematical model. The initial population consists of a certain number of matrix structures determined at the beginning of the algorithm.
- Stage 2. Determining the adaptation function. The adaptation function is a function that evaluates a given structure. Based on the adaptation function, the final solution is determined.
- Stage 3. Selection. The selection process in the genetic algorithm consists of selecting the best individuals (chromosomes) from the initial population to the next generation. The developed algorithm uses the roulette method, in which linear scaling was applied to counteract premature algorithm convergence in the initial iterations. Scaling coefficient  $C=2.0$  has been adopted.
- Stage 4. Crossover. The algorithm for selecting chromosomes for crossover takes into account the whole process of selecting chromosomes for crossover, in the case of chromosome oddity it randomly selects the chromosome to pair, randomly pairs the two chromosomes, randomly selects the cutting points of the chromosomes and activates the crossover algorithm (exchanging structures between two chromosomes) adequate to the proposed matrix structure.
- Stage 5. Mutation. The mutation process involves the exchange of genes in a given chromosome structure.

Stages 1-2 are one-off stages, occurring at the beginning of the algorithm's operation, stages 3-5 are repeated with a certain number of iterations until the stop condition is reached. The stop condition is a certain number of iterations. The assessment of individuals is carried out on the basis of the adaptation function, which is developed based on the function of criteria.

The crossover and mutation algorithms occur with a certain probability defined as input data. The final effect of the genetic algorithm is a generated population that determines the optimal setting of decision variables in the mathematical model. The final structure is selected from the structures generated in the last iteration of the algorithm and has the highest value of the adaptation function of all structures.

In order to obtain correct results generated by the genetic algorithm, the sensitivity of the algorithm to changes in the input parameters, i.e. the likelihood of crossover and mutation, must first be analysed. The sensitivity analysis is carried out on the basis of a fixed set of input parameters, which are entered into the genetic algorithm. The quality of generated solutions depends on properly selected parameters. The maximum value generated by a genetic algorithm in a sensitivity analysis indicates the parameters of the algorithm that have determined this value. In the second place, the verification process of the genetic algorithm should be carried out. The process of verifying the genetic algorithm was carried out on the basis of comparison of the algorithm solutions with those obtained by means of a random algorithm. The number of comparisons is determined at the beginning of the algorithm verification.

If a random algorithm determines a better solution, the sensitivity step should be performed for a different set of parameters. If a genetic algorithm has generated a better result in each comparison with a random algorithm, the verification process is completed. The

initial population of the algorithm was generated in a random way according to the assumptions presented when creating the chromosome structure.

## 4.2. Chromosome structure

The chromosome structure is a matrix structure and has been divided into four parts. Each of them is a representative of the decision variable of the mathematical model. It should be noted that the structure developed refers to a specific type of maintenance. The aircraft and crew are selected for assignment to a particular maintenance facility for the type of service already specified. Graphic interpretation of the chromosome structure processed by the genetic algorithm for the weekly aircraft operation schedule is shown in Figure 2.

In the first part, aircraft of different types are assigned to airports, the decision variable  $X1$  is determined. The number of rows determines the number of aircraft of all types, while the number of columns determines the number of airports. Binary values are taken at random assuming that one aircraft can be assigned to one airport (formula 2), and a specific number of aircraft of a given type can be assigned to each airport (formula 1). Cells with a value of "1" indicate the assignment of a given type of aircraft to a given airport.

In the second part of the chromosome structure, the date of the aircraft's directing to a given maintenance facility is determined, the TPOCZ decision variable is determined. The lines of the structure have an interpretation of aircraft, the columns indicate the days of the week in the airport operating schedule. Binary values are taken at random, subject to the limit of checking the time interval when the aircraft is allowed to be directed for maintenance (Model 3, 4). Time windows are created in the schedule, in which the aircraft should be assigned to a particular maintenance facility. The creation of a given time window is optional and depends on whether the aircraft reaches the lower limit of use in a given week of the schedule. One binary value "1" is required in each time window. The size of the time window depends on the type of aircraft, the intensity of its use on a given working day. After selecting the moment when the aircraft is directed to the maintenance facility, its maintenance time is determined (blue genes in Fig. 2). For each set of aircraft assigned to airports a limitation of the required number of aircraft remaining at the airport and performing the assigned tasks is checked (Model 8).

In the third part of the structure, the allocation of aircraft to maintenance facilities is determined, the decision variable  $X2$  is determined. The lines of the structure define the aircraft and the columns define the maintenance facilities. The binary values are adopted at random taking into account the limitations of the assignment of the number of aircraft to a particular maintenance facility (formula 5). Aircraft are assigned to such maintenance facilities which comply with the limitations of distance from airports (formula 6) and maintenance duration (formula 7). The number of substructures referring to the third part depends on the number of aircraft to be directed for maintenance in the second part. Each subsequent substructure for the third part applies only to those aircraft which are maintained in the overlapping period of time or which are maintained consecutively from one day to the next, e.g. in Figure 2 directing of aircraft no. 8 on the seventh day would introduce a new substructure referring to the third part and consequently another substructure referring to the fourth part.

In the fourth part, crews are assigned to the maintenance facilities concerned, the decision variable  $X3$  is determined. The lines of the structure define the crews and the columns define maintenance facilities. Binary values are adopted at random taking into account the limitations of the required number of crews at a given maintenance facility (formula 9), the assignment of crews to aircraft (formula 10) and the assignment of crews to maintenance facilities (formula 11). It should be taken into account that the sum of all ones in the fourth part must be equal to the sum of all ones in the third part according to the limitation that the number of aircraft maintained in a given facility is equal to the number of crews assigned (formula 10).



values of the substructures are exchanged for each chromosome pair. Graphic interpretation of the crossover process is shown in Fig. 3a, Fig. 3b, Fig. 3c.

Since Parts II, III and IV are interrelated and the results in Part II determine the settings of the subsequent parts of the structure, only those lines of the structure that were drawn in Part II are selected for the crossover process for Parts III and IV. The cutting points for the columns are selected randomly. After the crossover process, a situation may occur where an incorrect structure is created that does not take into account all the limitations of the mathematical model, e.g. in Fig. 3b for Part Four in chromosome I the limitation (10) is not fulfilled because crew no. 6 is assigned to the maintenance facility to which no aircraft has been assigned in Part Three. The repair algorithm for Parts Three and Four checks the individual columns of these substructures for the limitation (5) for Part Three, i.e. the number of aircraft assigned to the maintenance facility, and in the event of an exceeding, it randomly eliminates the value "1" indicating the assignment of the aircraft to the maintenance facility until the limitation is met (5). In addition, the limitations of the distance from these facilities to the airports (formula 6) and the time of maintenance (7) are checked. If these limitations are not met, the algorithm removes the assignment. In Part Four, the sum of ones for the individual columns corresponding to maintenance facilities is to be equal to the sum of ones for the individual columns of Part Three according to the limitation (10). When the number of ones in Part IV is greater than in Part III, the algorithm randomly eliminates them until both values are equal. Where the number of ones is less, the algorithm will randomly assign the crews to a particular maintenance facility, taking into account the limitation to the required number of crews at the particular maintenance facility (9) and the assignment of crews to aircraft maintenance facilities (11). The structure of the chromosome after the repair process is shown in Fig. 3c (red values in Part IV are the values modified by the repair algorithm). The repair algorithm for Part I of the structure, when the limitation (1) for the required number of aircraft of a given type at a given airport is not met, adds another aircraft at random to the airport taking into account the limitation (2) that one aircraft is assigned to one airport. If the limitation (1) is exceeded, it randomly removes the selected assignment also taking into account the limitation (2). The repair algorithm for Part II comes down to correcting the time window limitations (3,4) and determining at random a new date of service if these limitations are not met. In addition, the algorithm removes the day on which a aircraft is assigned for maintenance if the aircraft is not assigned to any airport in Part I.

The mutation process consists in random selection of the part of the chromosome where the mutation is to take place and then the gene to mutate and random conversion from "0" to "1" or vice versa. Graphic interpretation of the mutation process is shown in Figure 4a and Figure 4b. In the case of an erroneous structure generated after the mutation process, a repair algorithm is started which works identically to the crossover process.

Before mutation

	1	2	3	4	1	2	3	4	5	6	7	1	2	3	4	1	2	3	4	
1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
3	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	3	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	1	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0
6	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	6	0	0	0	0
7	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0
8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	8	0	1	0	0

Fig. 4a. Chromosome before mutation

After mutation

	1	2	3	4	1	2	3	4	5	6	7	1	2	3	4	1	2	3	4	
1	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
3	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	3	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	1	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0
6	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	6	0	0	0	0
7	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0
8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	8	0	1	0	0

Fig. 4b. The chromosome after mutation

## 5. Verification of the genetic algorithm

### 5.1. Input data

In order to verify the genetic algorithm determining the decision variables presented in the mathematical model, it has been assumed that there is one type of maintenance for one type of aircraft. Twelve aircraft are considered for use at four airports, two ac are used at each airport on each working day, one ac is used as an auxiliary in case the other one is directed for maintenance. It was assumed that there are five crews and six maintenance facilities available. There are two maintenance stations at each maintenance facility. The permissible service life of the ac is shown in Table 1. Data on the operation of the ac on each day of its operation were extracted from the aircraft continuing airworthiness information system database providing information on aircraft operations. A flight schedule of fifteen working days has been adopted, the number of flight hours of aircraft at each airport is shown in Table 2. It was assumed that in the analysed schedule of flights on each working day there is a possibility to send each ship for maintenance. The distance covered by individual aircraft is shown in Table 3, to simplify the recording of data, this distance is assumed to be the same for each day of the schedule (Table 3 shows only distances for the first day of the aircraft operation). The unit combustion cost of aircraft of a given type is presented in Table 4. The distances between the airport and the maintenance facility are shown in Table 5. It is assumed that all maintenance facilities are active and aircraft can be directed to them. The maintenance time forecast by the airport is 3 days.

The cost of commissioning a given maintenance facility and the maintenance time are shown in Table 6. The costs associated with hiring a given crew to perform the maintenance at a given maintenance facility are shown in Table 7.

### 5.2. Sensitivity analysis and verification of the genetic algorithm

The genetic algorithm has been implemented in programming language C#. The first step in performing a sensitivity analysis of the algorithm is to find a set of the best parameters that generate the best solution. The parameters taken for analysis included the probability of crossover  $p_{kryz}$ , the probability of mutation  $p_{mut}$ . As a result of the experimental studies, the population number of 50 individuals and the number of iterations of 200 were determined. The algorithm parameters to be tested include the crossover operator assuming five values  $p_{kryz} = 0.2; 0.4; 0.6; 0.8; 1$  and the mutation operator assuming three values  $p_{mut} = 0.01; 0.03; 0.05$ . The following combinations were tested to determine the best parameter settings, Table 8. The best results generated in each parameter setting are shown in Table 9.

The sensitivity analysis showed that the best results were generated taking into account the following parameters: crossover parameter – 0.8 and mutation parameter 0.05. An example of the operation of a genetic algorithm for selected settings of crossover and mutation input parameters is shown in Fig. 5 respectively. The best solutions are generated with a crossover parameter of more than 0.8, due to the complexity of the research problem and the need to search for new space of solutions that can generate optimal or suboptimal solutions. The lower probability of crossover generates worse solutions, because there is no competition between individuals in the population and thus no search for new areas generating suboptimal solutions. In addition, with small values of the crossover parameter, there is a risk of passing from the local optimum of worse quality than the algorithm determined in the initial iterations (Fig. 5 c and 5). The presented examples emphasize the fact that the key role in generating suboptimal or optimal results is played by the process of selection of parameters of the genetic algorithm. These parameters are selected in an experimental way. The moment of convergence of the algorithm to the optimum depends also on the proper selection of

Table 1. Permissible service life of aircraft [h]

te(1,1)=420	te(1,2)=520	te(1,3)=330	te(1,4)=310	te(1,5)=430
te(1,6)=570	te(1,7)=600	te(1,8)=550	te(1,9)=590	te(1,10)=510
te(1,11)=650	te(1,12)=550	-	-	-

Table 2. Number of hours to fly [h]

$\alpha(1,1,1)=7$	$\alpha(1,1,2)=7$	$\alpha(1,1,3)=7$	$\alpha(1,1,4)=7$	$\alpha(1,1,5)=7$
$\alpha(1,1,6)=7$	$\alpha(1,1,7)=7$	$\alpha(1,1,8)=7$	$\alpha(1,1,9)=7$	$\alpha(1,1,10)=7$
$\alpha(1,1,11)=7$	$\alpha(1,1,12)=7$	$\alpha(1,1,13)=7$	$\alpha(1,1,14)=7$	$\alpha(1,1,15)=7$
$\alpha(1,2,1)=10$	$\alpha(1,2,2)=10$	$\alpha(1,2,3)=10$	$\alpha(1,2,4)=10$	$\alpha(1,2,5)=10$
$\alpha(1,2,6)=10$	$\alpha(1,2,7)=10$	$\alpha(1,2,8)=10$	$\alpha(1,2,9)=10$	$\alpha(1,2,10)=10$
$\alpha(1,2,11)=10$	$\alpha(1,2,12)=10$	$\alpha(1,2,13)=10$	$\alpha(1,2,14)=10$	$\alpha(1,2,15)=10$
$\alpha(1,3,1)=8$	$\alpha(1,3,2)=8$	$\alpha(1,3,3)=8$	$\alpha(1,3,4)=8$	$\alpha(1,3,5)=8$
$\alpha(1,3,6)=8$	$\alpha(1,3,7)=8$	$\alpha(1,3,8)=8$	$\alpha(1,3,9)=8$	$\alpha(1,3,10)=8$
$\alpha(1,3,11)=8$	$\alpha(1,3,12)=8$	$\alpha(1,3,13)=8$	$\alpha(1,3,14)=8$	$\alpha(1,3,15)=8$
$\alpha(1,4,1)=11$	$\alpha(1,4,2)=11$	$\alpha(1,4,3)=11$	$\alpha(1,4,4)=11$	$\alpha(1,4,5)=11$
$\alpha(1,4,6)=11$	$\alpha(1,4,7)=11$	$\alpha(1,4,8)=11$	$\alpha(1,4,9)=11$	$\alpha(1,4,10)=11$
$\alpha(1,4,11)=11$	$\alpha(1,4,12)=11$	$\alpha(1,4,13)=11$	$\alpha(1,4,14)=11$	$\alpha(1,4,15)=11$

Table 3. Length of flight routes for individual aircraft [km]

$d((1,1),1,1)=4000$	$d((1,2),1,1)=3500$	$d((1,1),2,1)=4500$	$d((1,2),2,1)=5000$	$d((1,1),3,1)=3800$
$d((1,2),3,1)=3900$	$d((1,1),4,1)=4100$	$d((1,2),4,1)=4500$	-	-

Table 4. Specific fuel consumption per aircraft [litres /100 km]

ks(1,1)=3,5	ks(1,2)=4,4	ks(1,3)=4,1	ks(1,4)=4,3	ks(1,5)=4,7
ks(1,6)=4,0	ks(1,7)=4,2	ks(1,8)=3,7	ks(1,9)=4,4	ks(1,10)=4,2
ks(1,11)=3,7	ks(1,12)=3,9	-	-	-

Table 5. Distance between airport and maintenance facility [km]

$d1(1,1)=70$	$d1(1,2)=30$	$d1(1,3)=35$	$d1(1,4)=45$	$d1(1,5)=35$	$d1(1,6)=25$
$d1(2,1)=45$	$d1(2,2)=100$	$d1(2,3)=20$	$d1(2,4)=35$	$d1(2,5)=15$	$d1(2,6)=25$
$d1(3,1)=60$	$d1(3,2)=15$	$d1(3,3)=25$	$d1(3,4)=25$	$d1(3,5)=40$	$d1(3,6)=30$
$d1(4,1)=35$	$d1(4,2)=40$	$d1(4,3)=15$	$d1(4,4)=50$	$d1(4,5)=20$	$d1(4,6)=30$

Table 6. Cost and time of maintenance at a given maintenance facility [PLN thousand, days]

$ko(1,1,1)=10$	$ko(1,1,2)=1.5$	$ko(1,1,3)=1.2$	$ko(1,1,4)=1.6$	$ko(1,1,5)=1.7$	$ko(1,1,6)=2.0$
$ta(1,1,1)=3$	$ta(1,1,2)=2$	$ta(1,1,3)=2$	$ta(1,1,4)=1$	$ta(1,1,5)=2$	$ta(1,1,6)=3$

Table 7. Costs of maintenance performed by the crew [thousand PLN]

$kz(1,1,1)=3.0$	$kz(1,2,1)=3.6$	$kz(1,3,1)=2.4$	$kz(1,4,1)=2.2$	$kz(1,5,1)=2.1$	$kz(1,6,1)=2.6$
$kz(2,1,1)=3.5$	$kz(2,2,1)=3.0$	$kz(2,3,1)=3.1$	$kz(2,4,1)=3.4$	$kz(2,5,1)=3.1$	$kz(2,6,1)=2.4$
$kz(3,1,1)=4.0$	$kz(3,2,1)=2.9$	$kz(3,3,1)=3.9$	$kz(3,4,1)=3.1$	$kz(3,5,1)=3.8$	$kz(3,6,1)=2.8$
$kz(4,1,1)=4.2$	$kz(4,2,1)=2.3$	$kz(4,3,1)=2.1$	$kz(4,4,1)=3.2$	$kz(4,5,1)=2.3$	$kz(4,6,1)=2.9$
$kz(5,1,1)=3.4$	$kz(5,2,1)=3.7$	$kz(5,3,1)=2.4$	$kz(5,4,1)=3.2$	$kz(5,5,1)=2.7$	$kz(5,6,1)=3.1$

Table 8. Test input parameters of the genetic algorithm

$P_{krzyz}$	$P_{mut}$	$P_{krzyz}$	$P_{mut}$	$P_{krzyz}$	$P_{mut}$
0.2	0.01	0.4	0.01	0.6	0.01
0.2	0.03	0.4	0.03	0.6	0.03
0.2	0.05	0.4	0.05	0.6	0.05
0.8	0.01	1	0.01	-	-
0.8	0.03	1	0.03	-	-
0.8	0.05	1	0.05	-	-

Tables 4-8 source: own study

the algorithm input parameters, which can be observed in Fig. 5 a and b. (convergence to optimum in 150 iterations – (a) and 50 iterations – (b)).

In order to verify the correctness of the genetic algorithm (AG), its results generated for the parameters: crossover – 0.8 and mutation – 0.05 were compared with random values generated by a random algorithm (AL). Random values are the values of chromosomes generated by the genetic algorithm before the crossover and mutation process according to the assumption that the initial population was generated at random. The value of the adaptation function for each algorithm is shown in Table 10. By analysing the results below, it can be concluded that the genetic algorithm in each comparison generated a better result than the random algorithm. Genetic algorithms belong to probabilistic algorithms, so they generate different, approximate final results. It is therefore recommended that several final solutions are identified and a maximum solution is selected. The adaptation function with the highest value is generated by decision variables for which the functions of the mathematical model criterion reach optimal values in



Table 9. Sensitivity analysis of the genetic algorithm

pkrzyż,	pmut	Results	pkrzyż,	pmut	Results	pkrzyż,	pmut	Results
0.2	0.01	1.23	0.4	0.01	1.63	0.6	0.01	2.87
0.2	0.03	1.26	0.4	0.03	1.72	0.6	0.03	2.91
0.2	0.05	1.29	0.4	0.05	1.83	0.6	0.05	2.80
0.8	0.01	3.31	1	0.01	3.05	-	-	-
0.8	0.03	3.38	1	0.03	3.22	-	-	-
0.8	0.05	3.42	1	0.05	3.12	-	-	-

Source: own study

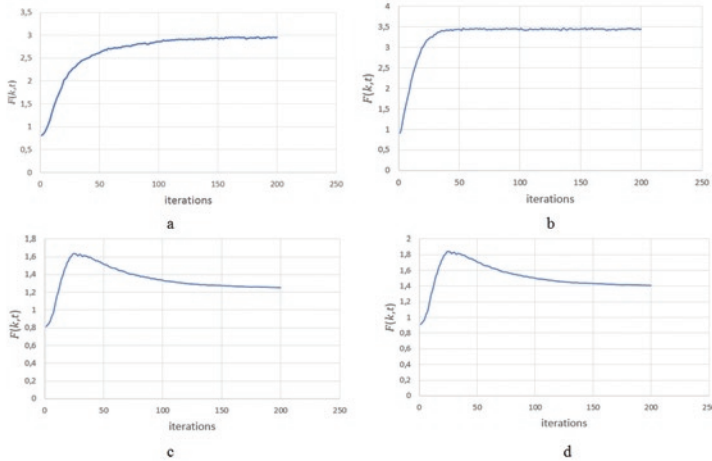


Fig. 5. Operation of the genetic algorithm, a) crossover 0.6; mutation 0.02, b) crossover 0.8; mutation 0.05; c) crossover 0.4; mutation 0.01, d) crossover 0.4; mutation 0.05

	1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
7	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
10	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	1	2	3	4	5	6	1	2	3	4	5	6
1	0	0	0	1	0	0	1	0	1	0	0	0
2	0	0	0	0	0	0	2	0	0	0	1	0
3	0	0	0	0	0	0	3	0	0	0	0	0
4	0	0	0	1	0	0	4	0	0	0	1	0
5	0	0	0	0	0	0	5	0	0	1	0	0
6	0	1	0	0	0	0						
7	0	0	0	0	0	0						
8	0	0	0	0	0	0						
9	0	0	0	0	0	1						
10	0	0	0	0	0	0						
11	0	0	0	0	0	0						
12	0	0	0	0	0	0						

Fig. 6. Chromosome structure for the analysed example  
Source: own study

Table 10. Verification of the genetic algorithm

AG	AL	AG	AL	AG	AL
3.47	1.67	3.60	1.46	3.44	1.54
3.44	1.89	3.56	1.22	3.65	1.72
3.57	1.55	3.54	1.45	3.46	2.01
3.43	2.12	3.45	1.89	3.47	2.11
3.41	2.24	3.39	1.54	3.57	1.98
3.45	2.16	3.54	1.72	3.66	1.09
3.47	1.88	3.48	1.32	3.42	1.22
3.66	1.34	3.44	1.58	3.55	1.67
3.44	1.53	3.55	1.82	3.57	1.89
3.43	1.78	3.56	1.88	3.44	2.11
3.46	1.33	3.50	1.49	3.65	2.32
3.44	1.25	3.45	1.77	3.49	1.55
3.45	1.57	3.50	1.89	3.66	1.34
3.46	1.32	3.66	1.56	3.45	1.76
3.40	1.56	3.57	1.46	3.45	1.34

the sense of the Pareto solution. In the analysed example, the adaptation function reaches a value of 3.66 four times, which may indicate a suboptimal value.

An example of the chromosome structure in graphic form is shown in Fig. 6. According to the structure presented, three aircraft are assigned to each airport. Aircraft number 1, 4, 6 and 9 have been assigned to maintenance. These aircraft are assigned to maintenance stations as follows: aircraft no. 1 to maintenance station no. 3, aircraft no. 4 to maintenance station no. 4, aircraft no. 6 to maintenance station no. 2, aircraft no. 9 to maintenance station no. 5. The following crews have been assigned to maintenance facilities: crew no. 1 to maintenance facility no. 2, crew no. 2 to maintenance facility no. 4, crew no. 4 to maintenance facility no. 5, crew no. 5 to maintenance facility no. 3.

## 6. Conclusions

In the article the problem of aircraft maintenance planning and management for scheduled maintenance specified in the manufacturer's technical documentation is presented in a comprehensive way taking into account not only the phase of assignment of aircraft to maintenance stations located in different facilities but also the phase of assignment of aircraft to airports, the phase of maintenance scheduling (days) as well as the phase of assignment of crews to maintenance of the aircraft types. The multi-criteria approach to the presented mathematical model additionally emphasizes the complexity of the analysed issue.

The effectiveness of the proposed genetic algorithm was verified by means of a random algorithm. The sensitivity analysis emphasized the fact that the process of selecting the input parameters characterizing the genetic algorithm is an experimental process. The number of iterations, population size has also been experimentally determined.

Both the mathematical model and the genetic algorithm are new approaches of analysing a given research problem, so it is not possible to compare it with results generated by other optimisation algorithms. The mathematical model and the proposed optimisation algorithm can serve as a test function for other algorithms.

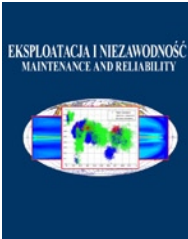
In the genetic algorithm, the classic roulette method was used to select individuals for the next generation. Further research may include testing of subsequent selection methods with the use of other selection algorithms described in the literature. The initial population was generated at random. Experiments are carried out to generate the initial

population with other heuristic algorithms, e.g. ant colony optimization algorithm, which can improve the effectiveness of the genetic algorithm and the quality of generated solutions. In addition, the further direction of research may comprise the inclusion of random events in the model in the form of sudden failures and unplanned repairs.

## References

1. Adamides E. D, Stamboulis Y. A, Varelis A. G. Model-based assessment of military aircraft engine maintenance systems. *Journal of the Operational Research Society* 2004; 55(9): 957-967, <https://doi.org/10.1057/palgrave.jors.2601756>.
2. Beliën J, Demeulemeester E, De Bruecker P, Van den Bergh J, Cardoen B. Integrated staffing and scheduling for an aircraft line maintenance problem. *Computers & Operations Research* 2013; 40(4): 1023-1033, <https://doi.org/10.1016/j.cor.2012.11.011>.
3. Burkard R, Dell'Amico M, Martello S. *Assignment problems*. Society for Industrial and Applied Mathematics, Philadelphia 2009.
4. Cheung A, Ip W. H, Lu, D. Expert system for aircraft maintenance services industry. *Journal of Quality in Maintenance Engineering* 2005; 11(4): 348-358, <https://doi.org/10.1108/13552510510626972>.
5. Ehrgott M. *Multicriteria Optimization*. Springer: Berlin/Heidelberg, Germany 2005.
6. Fard N. S, Melachrinoudis E. Maintenance scheduling for critical parts of aircraft. *Annual Reliability and Maintainability Symposium* 1991: 44-47.
7. Ghobbar A. A, Friend C. H. The material requirements planning system for aircraft maintenance and inventory control: a note. *Journal of Air Transport Management* 2004; 10(3): 217-221, <https://doi.org/10.1016/j.jairtraman.2003.10.001>.
8. The Application of Ant Algorithm in the Assignment Problem of Aircrafts to Stops Points on the Apron, in: *Journal of KONES* 2018; 25(4): 479-488, <https://doi.org/10.5604/01.3001.0012.7845>.
9. Gołda P, Kowalski M, Wasser C, Dygnatowski P, Szporka A. Elements of the model positioning of aircraft on the apron. *Archives of Transport* 2019; 51(3): 101-108, <https://doi.org/10.5604/01.3001.0013.6166>.
10. Gołda P. Selected decision problems in the implementation of airport operations. *Scientific Journal of Silesian University of Technology. Series Transport* 2018; 101: 79-88, <https://doi.org/10.20858/sjsutst.2018.101.8>.
11. Gustavsson E, Patriksson M, Strömberg A.-B, Wojciechowski A, Önnheim M. Preventive maintenance scheduling of multi-component systems with interval costs. *Computers & Industrial Engineering* 2014; 76: 390-400, <https://doi.org/10.1016/j.cie.2014.02.009>.
12. Hahn R. A, Newman A. M. Scheduling united states coast guard helicopter deployment and maintenance at Clearwater Air Station. *Computers & Operations Research* 2012; 35(6): 1829-1843, <https://doi.org/10.1016/j.cor.2006.09.015>.
13. Horning S, Leung P, Fitzgerald A, Mrad N. Operational readiness simulator: Optimizing operational availability using a virtual environment. *International Journal of Aerospace Engineering* 2012; 1-8, <https://doi.org/10.1155/2012/425075>.
14. Izdebski M, Jacyna M. The organization of municipal waste collection: The decision model, w: *Annual Set The Environment Protection* 2018; 20: 919-933.
15. Izdebski M, Jacyna-Gołda I, Markowska K, Murawski J. Heuristic algorithms applied to the problems of servicing actors in supply chains. *Archives of Transport* 2017; 44(4): 25-34, <https://doi.org/10.5604/01.3001.0010.6159>.
16. Izdebski M, Jacyna-Gołda I, Wasiak M, Jachimowski R, Kłodawski M, Pyza D, Żak J. The application of the genetic algorithm to multi-criteria warehouses location problems on the logistics network, *Transport* 2018; 33(3): 741-750, <https://doi.org/10.3846/transport.2018.5165>.
17. Jacyna M, Izdebski M, Szczepański E, Gołda P. The task assignment of vehicles for a production company. *Symmetry* 2018; 11(10): 1-19, <https://doi.org/10.3390/sym10110551>.
18. Jacyna Marianna, Jachimowski Roland, Szczepański Emilian [and others] : Road vehicle sequencing problem in a railroad intermodal terminal – simulation research, in: *Bulletin of the Polish Academy of Sciences, Technical Sciences*, vol. 68, no. 5, 2020, ss. 1135-1148.
19. Jacyna-Gołda I, Izdebski M, Podvieszko A. Assessment of efficiency of assignment of vehicles to tasks in supply chains: A case study of a municipal company. *Transport* 2017; 32(3): 243-251, <https://doi.org/10.3846/16484142.2016.1275040>.
20. Jacyna-Gołda I, Izdebski M, Szczepański E, Gołda P. The assessment of supply chain effectiveness. *Archives of Transport* 2018; 45(1): 43-52, <https://doi.org/10.5604/01.3001.0012.0966>.
21. Joo S. J, Min H. A multiple objective approach to scheduling the preventive maintenance of modular aircraft components. *International Journal of Services and Operations Management* 2011; 9(1): 18-31, <https://doi.org/10.1504/IJSOM.2011.040319>.
22. Kalton A, Richards R. A. Advanced scheduling technology for more efficient (shorter) resource constrained schedules, *IEEE Aerospace Conference* 2008: 1-9, <https://doi.org/10.1109/AERO.2008.4526690>.
23. Knopik L, Migawka K. Multi-state model of maintenance policy. *Eksplotacja i niezawodność – Maintenance and Reliability* 2018; 20 (1): 125-130, <https://doi.org/10.17531/ein.2018.2.16>.
24. Kosicka E, Kozłowski E, Mazurkiewicz D. Intelligent Systems of Forecasting the Failure of Machinery Park and Supporting Fulfilment of Orders of Spare Parts. In *International Conference on Intelligent Systems in Production Engineering and Maintenance Springer Cham* 2017: 54-63.
25. Latorella K. A, Prabhu P. V. A review of human error in aviation maintenance and inspection. *International Journal of Industrial Ergonomics* 2000; 26(2): 133-161, [https://doi.org/10.1016/S0169-8141\(99\)00063-3](https://doi.org/10.1016/S0169-8141(99)00063-3).
26. Le M, Wu C, Zhan C, Sun L. Airline recovery optimization research: 30 years' march of mathematical programming-a classification and literature review, in *2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE) 2011*: 113-117, <https://doi.org/10.1109/TMEE.2011.6199160>.
27. Mazurkiewicz D. Artificial neural networks in efficiency analysis of brittle materials cutting. *Eksplotacja i Niezawodność – Maintenance and Reliability* 2002; 1(13): 25-29, <https://doi.org/10.17531/ein.2018.2.16>.
28. Mohaghegh Z, Kazemi R, Mosleh A. Incorporating organizational factors into Probabilistic Risk Assessment (PRA) of complex socio-technical systems: a hybrid technique formalization. *Reliability Engineering and System Safety* 2009; 94(5): 1000-1018, <https://doi.org/10.1016/j.res.2008.11.006>.
29. Ostrom L. T, Wilhelmsen C. A. Developing risk models for aviation maintenance and inspection. *International Journal of Aviation Psychology*

- 2008; 18(1): 30-42, <https://doi.org/10.1080/10508410701749407>.
30. Otman A, Jaafar A. A Comparative Study of Adaptive Crossover Operators for Genetic Algorithms to Resolve the Travelling Salesman Problem. *International Journal of Computer Applications* 2012; 31(11): 49-57.
  31. Papakostas N, Papachatzakis P, Xanthakis V, Mourtzis D, Chryssolouris G. An approach to operational aircraft maintenance planning. *Decision Support Systems* 2010; 48(4): 604-612, <https://doi.org/10.1016/j.dss.2009.11.010>.
  32. Pentico D. Assignment problems: A golden anniversary survey. *European Journal of Operational Research* 2007; 176(2): 774-793, <https://doi.org/10.1016/j.ejor.2005.09.014>.
  33. Quan G, Greenwood G. W, Liu D, Hu, S. Searching for multiobjective preventive maintenance schedules: Combining preferences with evolutionary algorithms. *European Journal of Operational Research* 2007; 177(3): 1969-1984, <https://doi.org/10.1016/j.ejor.2005.12.015>.
  34. Reiners T, Pahl J, Maroszek M, Rettig C. Integrated aircraft scheduling problem: an auto-adapting algorithm to find robust aircraft assignments for large flight plans. *45th International Conference on System Sciences* 2012; 4-7; 1267-1276, <https://doi.org/10.1109/HICSS.2012.330>.
  35. Sachon M, Paté-Cornell E. Delays and safety in airline maintenance. *Reliability Engineering and System Safety* 2000; 67(3): 301-309, [https://doi.org/10.1016/S0951-8320\(99\)00062-9](https://doi.org/10.1016/S0951-8320(99)00062-9).
  36. Samaranayake P, Kiridena S. Aircraft maintenance planning and scheduling: an integrated framework. *Journal of Quality in Maintenance Engineering* 2012; 18(4): 432-453, <https://doi.org/10.1108/13552511211281598>.
  37. Sarac A, Batta R, Rump C. M. A branch-and-price approach for operational aircraft maintenance routing. *European Journal of Operational Research* 2006; 175(3): 1850-1869, <https://doi.org/10.1016/j.ejor.2004.10.033>.
  38. Sriram C, Haghani A. An optimization model for aircraft maintenance scheduling and re-assignment. *Transportation Research Part A: Policy and Practice* 2003; 37(1): 29-48, [https://doi.org/10.1016/S0965-8564\(02\)00004-6](https://doi.org/10.1016/S0965-8564(02)00004-6).
  39. Stojković G, Soumis F, Desrosiers J, Solomon M. M. An optimization model for a real-time flight scheduling problem. *Transportation Research Part A: Policy and Practice* 2002; 36(9): 779-788, [https://doi.org/10.1016/S0965-8564\(01\)00039-8](https://doi.org/10.1016/S0965-8564(01)00039-8).
  40. Świderski A, Józwiak A, Jachimowski R. Operational quality measures of vehicles applied for the transport services evaluation using artificial neural networks. *Eksploatacja i Niezawodność – Maintenance and Reliability* 2018; 20 (2): 292-299, <https://doi.org/10.17531/ein.2018.2.16>.
  41. T'kindt V, Billaut J. Ch. *Multicriteria scheduling Theory, Models and Algorithms*, Springer, Berlin 2006.
  42. Tiassou K, Kanoun K, Kaâniche M, Seguin C, Papadopoulos C. Modeling aircraft operational reliability, in Flammini, F., Bologna, S., and Vittorini, V. (eds.) *Computer safety, reliability, and security*. Heidelberg, Germany: Springer-Verlag 2011: 157-170.
  43. Tiassou K, Kanoun K, Kaâniche M, Seguin C, Papadopoulos, C. Aircraft operational reliability: a model-based approach and a case study. *Reliability Engineering & System Safety* 2013; 120: 163-176, <https://doi.org/10.1016/j.ress.2013.07.008>.
  44. Van den Bergh J, De Bruecker P, Beliën J, De Boeck L., Demeulemeester, E. A three-stage approach for aircraft line maintenance personnel rostering using MIP, discrete event simulation and DEA. *Expert Systems with Applications* 2013; 40(7): 2659-2668, <https://doi.org/10.1016/j.eswa.2012.11.009>.
  45. Viotti P, Poletti A, Pomi R, Innocenti C. Genetic algorithms as a promising tool for optimisation of the MSW collection routes. *Waste Management Res.* 2014; 21: 292-298, <https://doi.org/10.1177/0734242X0302100402>.
  46. Warrington L, Jones J. A, Davis N. Modelling of maintenance, within discrete event simulation. *Annual Reliability and Maintainability Symposium* 28-31 Jan. 2002. Seattle, WA: 260-265, <https://doi.org/10.1109/RAMS.2002.981652>.
  47. Wasiak M, Jacyna-Gołda I, Jachimowski R, Kłodawski M, Izdebski M. The use of a supply chain configuration model to assess the reliability of logistics processes, *Eksploatacja i niezawodność – Maintenance and Reliability* 2019; 3: 367-374, <https://doi.org/10.17531/ein.2019.3.2>.
  48. Yan S, Yang T.-H, Chen H.-H. Airline short-term maintenance manpower supply planning. *Transportation Research Part A: Policy and Practice*; 38(9-10): 615-642, [10.1016/j.tra.2004.03.005](https://doi.org/10.1016/j.tra.2004.03.005).
  49. Yang T.-H, Yan S, Chen H.-H. An airline maintenance manpower planning model with flexible strategies. *Journal of Air Transport Management*; 9(4): 233-239, [https://doi.org/10.1016/S0969-6997\(03\)00013-9](https://doi.org/10.1016/S0969-6997(03)00013-9).
  50. Yu E, Sung K. S. A genetic algorithm for a university weekly courses timetabling problem. *International Transactions in Operational Research* 2002; 9: 703-717, <https://doi.org/10.1111/1475-3995.00383>.
  51. Zieja M, Krutkow A, Iwaniuk, M, Gołda P. Model niezawodności i bezpieczeństwa lotów w systemie utrzymania zdadności do lotu statków powietrznych. *Prace Naukowe Politechniki Warszawskiej, Transport* 2017; 116: 337-346.



Article citation info:

Zhiyong G, Jiwu L, Rongxi W. Prognostics uncertainty reduction by right-time prediction of remaining useful life based on hidden Markov model and proportional hazard model. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 154–164, <http://dx.doi.org/10.17531/ein.2021.1.16>.

## Prognostics uncertainty reduction by right-time prediction of remaining useful life based on hidden Markov model and proportional hazard model

Gao Zhiyong<sup>a</sup>, Li Jiwu<sup>a</sup>, Wang Rongxi<sup>a,\*</sup>

<sup>a</sup>State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University; Xi'an 710049, China

Indexed by:



### Highlights

- Reducing the prognostic uncertainty by making full use of operating data.
- Extracting degradation characteristics from long-term running data.
- Using degradation characteristics as input variables to obtain the survival function.
- Targeted modeling for the last degradation state.

### Abstract

Uncertainty is a key problem in remaining useful life (RUL) prediction, and measures to reduce uncertainty are necessary to make RUL prediction truly practical. In this paper, a right-time prediction method is proposed to reduce the prognostics uncertainty of mechanical systems under unobservable degradation. Correspondingly, the whole RUL prediction process is divided into three parts, including offline modelling, online state estimating and online life predicting. In the offline modelling part, hidden Markov model (HMM) and proportional hazard model (PHM) are built to map the whole degradation path. During operation, the degradation state of the object is estimated in real time. Once the last degradation state reached, the degradation characteristics are extracted, and the survival function is obtained with the fitted PHM. The proposed method is demonstrated on an engine dataset and shows higher accuracy than traditional method. By fusing the extracted degradation characteristics, the obtained survival function can be basis for optimal maintenance with lower uncertainty.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

RUL, uncertainty, right-time prediction, PHM, HMM.

## 1. Introduction

With the increasing requirement of safety and economy of equipment operation, more and more attention is paid to the prediction of remaining useful life (RUL). Scholars in various fields have conducted extensive research on this issue, including aerospace field [1, 17], electronics [28], mechanical industrial [16, 22], and many other fields [7, 20]. Most researchers devoted to improve the accuracy of the prediction results, and have investigated many effective methods for RUL prediction, including various neural networks (NN) [41], support vector regression (SVR) [34], stochastic process [13, 36], and other methods [3, 21]. Ramezani, Moini and Riahi [27] gives a comprehensive summary to various methods.

However, the uncertainty of RUL prediction is still an inevitable problem remaining to be solved. For most mechanical systems, the actual degradation level is not directly observable and can only be inferred by running data. Inescapably, uncertainty is infused to the resulting degradation status with the inference process, and makes RUL prediction result under the inferred degradation status lack credibility in maintenance guiding. Considering this problem, scholars have studied and put forward some contributing methods [14, 24,

38]. Sankararaman and Goebel [30] suggests that the perfectly and precisely prediction of engineering systems behavior is not possible in practical engineering applications due to prognostics uncertainty, and divided the sources of uncertainty into four categories: present state uncertainty, future uncertainty, modeling uncertainty and prediction method uncertainty. Some researchers believe that the single value prediction of RUL is meaningless, even deceptive, what is really needed for decision-making is the prediction interval or probability distribution [11, 29, 32]. To reduce the uncertainty of RUL prediction, Sun, Zuo, Wang and Pecht [33] proposed a state space degradation model based method with combining online monitoring data. To reduce the impact of prediction method uncertainty, Baraldi, Mangili and Zio [2] fused the predicted results of two methods with belief function theory. Actually, the present state uncertainty and future uncertainty is time dependent. Engel, Gilmartin, Bongort and Hess [9] have demonstrated that with the reduction of residual life, the accuracy of the prediction increases and the uncertainty decreases. To reduce the uncertainty of later prediction, Deng, Bucchianico and Pechenizkiy [6] applied surrogate Wiener propagation model to RUL prediction to make later prediction more accurate by sacrificing the early prediction accuracy.

(\*) Corresponding author.

E-mail addresses: G. Zhiyong - [zhygao@mail.xjtu.edu.cn](mailto:zhygao@mail.xjtu.edu.cn), L. Jiwu - [lijiwu@stu.xjtu.edu.cn](mailto:lijiwu@stu.xjtu.edu.cn), W. Rongxi - [rongxiwang@163.com](mailto:rongxiwang@163.com)

Most studies divide the RUL prediction process into two parts: offline training and online predicting[43], and deal with the prognostics uncertainty under the method level. However, the source of uncertainty is the lack of known operational information. In fact, the entire degradation process of the equipment can be divided into several stages. In the early stage of degradation, the device can run stably for a long time, and little information is available for RUL prediction. Therefore, it is neither accurate nor meaningful to predict the RUL in the early degradation stage. Once the device gets into the later stage of degradation, the current state and degradation pattern can be revealed by the monitoring data or inspection data with in a long-term operation. This is vital for reducing the present state uncertainty and future uncertainty, and improving the accuracy of prediction results.

The proportional hazard model (PHM) is proposed by Cox in 1972 [5], and has been widely used in biology and medicine to study the association between a patient's survival time and one or more variables [39, 42]. The uncertainty of survival problem can be felicitously expressed by PHM with the form of survival probability. In the field of reliability, PHM has also been used for reliability analysis and life prediction. Tran, Hong, Yang and Tan [35] proposed a method for RUL prediction by using support vector machine (SVM) and PHM. Man and Zhou [23] modeled the monitoring indicator value with a Wiener process, and took the predicted indicator value as the input variable of PHM for RUL prediction. Hu and Chen [12] proposed a preventive maintenance strategy with RUL predicted by Wiener process and PHM. Zhou, Son, Zhou, Mao and Salman [43] used both the inspection data and event data as the input variable of PHM to predict RUL. Qiu, Gu and Chen [25] established a health indicator for bearing, and predicted the RUL of bearing with support vector regression (SVR) and a Weibull PHM. Du, Wu, Zhou and Makis [8] modeled the RUL of lubrication with PHM. Equeter, Ducobu, Rivière-Lorphèvre, Serra and Dehombreux [10] took the transformed operation parameter as the input variable of PHM, and predicted the RUL of cutting tool. You, Li, Meng and Ni [40] believes that modelling the whole degradation process with just a single model is inaccurate, and he proposed a two-zone PHM for RUL prediction by dividing the device operation process into stable state and unstable state. Lin, Sun, Xu and Zhang [19] propose a multi-zone PHM by establishing the PHM for each degradation state. The above studies generally take monitoring indicator or synthetic health indicator as the input variable of PHM, and the indicator value over the rest life of device have to be predicted firstly before the RUL prediction. However, the uncertainty of indicator value predicted by regression algorithm or stochastic process will inevitably increasing the uncertainty of RUL prediction. Moreover, the whole degradation process of a device is changeable. It is hardly to get accurate prediction results by covering the whole degradation process with just one model.

In this paper, the concept of right-time RUL prediction is put forward with the objective to reduce the prognostics uncertainty of mechanical systems under unobservable degradation. Through a relatively rough status assessment, an appropriate prediction moment can be selected, and all the operating data accumulated before the prediction moment can be used for RUL prediction, so as to reduce the data uncertainty with a bigger data size. After the prediction moment, the established life model with smaller modeling scope can be modeled to reduce the model uncertainty. Concretely, the whole degradation process of a mechanical system is divided into several states, name the last degradation state as critical state, other states as stable state, and the RUL is predicted only after the critical state reached with the accumulated long-term operation data. Correspondingly, a right-time RUL prediction method based on hidden Markov model (HMM) and PHM is proposed. Different from the existing methods, the proposed method divides the whole prediction process into three parts: offline training, online state estimating and online RUL predicting. In the offline training process, HMM is used for states dividing. For each training sample, a HMM is fitted to express the specific degradation path of the sample. By clustering these HMMs and synthesizing the

parameters of HMMs under each cluster, several general models can be obtained to represent some typical degradation paths. Using the degradation characteristics of stable state extracted from the training data by general HMMs as the input variable, the PHM specific to the critical state is modeled and fitted with the duration of the critical state of each training sample. In the online state estimating process, the trained general model is used to estimate the device state continuously during the whole stable state. It is considered that the device can still operation stably with a long time till the critical state reached, and there is no need to predict RUL in stable state. Once the critical state is judged to be reached, the degradation characteristics of the mechanical system are extracted and RUL is predicted. In the online RUL predicting process, the survival function of the critical state can be obtained by taking the extracted degradation characteristics as the input variable of the fitted PHM, and RUL can be derived from the survival function.

The right-time RUL prediction put forward in this paper can better fit the PHM to reduce prediction uncertainty for the following reasons:

- (1) The PHM is established only for the critical state of degradation. The local degradation process is relatively simple to model, and it is more appropriate to express by the baseline function of PHM.
- (2) During the critical state, the degradation path of the device is more clear with the long run data. The characteristics reflecting the specific degradation path of the device can be extracted from the existing data as the input of the PHM. The extracted characteristics can reflect the holistic degradation process and avoid the uncertainty from indicator prediction, is preferable to be the input variable of PHM which has an effect on the whole survival function.

The rest of this paper is organized as follows: Section 2 presents the proposed HMM and PHM based right-time RUL prediction method. In Section 3, a demonstration with the turbofan engine simulation data is conducted, and the performance of the proposed method is compared with a traditional SVR-PHM method. Finally, Section 4 concludes this paper.

## 2. Right-time RUL prediction method

In order to reduce the uncertainty of RUL prediction, a right-time RUL prediction method based on HMM and PHM is proposed, which divides the whole RUL prediction process into three parts: offline modeling, online state estimating and online RUL predicting. The procedure of the proposed method is shown in Figure. 1, which is mainly included in the following three sections.

### 2.1. Offline modelling

In the offline modeling stage, the whole life indicator data of training samples, which can be monitoring data or inspection data, is needed for fitting the HMMs and PHM. For each training sample, the individual optimal HMM is fitted. In order to get more general models, all this individual optimal HMMs are clustered, and the models under each cluster are synthesized into a general model. With these general models, the degradation states of each training sample are divided, and the degradation characteristics of stable state are extracted. Taking the extracted degradation characteristics as input variables, the PHM specific to the critical state is established.

#### 2.1.1. General HMM pool establishing

HMM is a probabilistic model with several hidden states  $\mathbf{S} = \{S_1, S_2, \dots, S_N\}$  and a set of discrete or continuous observations [26]. Figure. 2 gives the construction of HMM with  $i_t$  represents the hidden state at time  $t$  and  $o_t$  represents the observation at time  $t$ .

There are two assumptions in HMM:

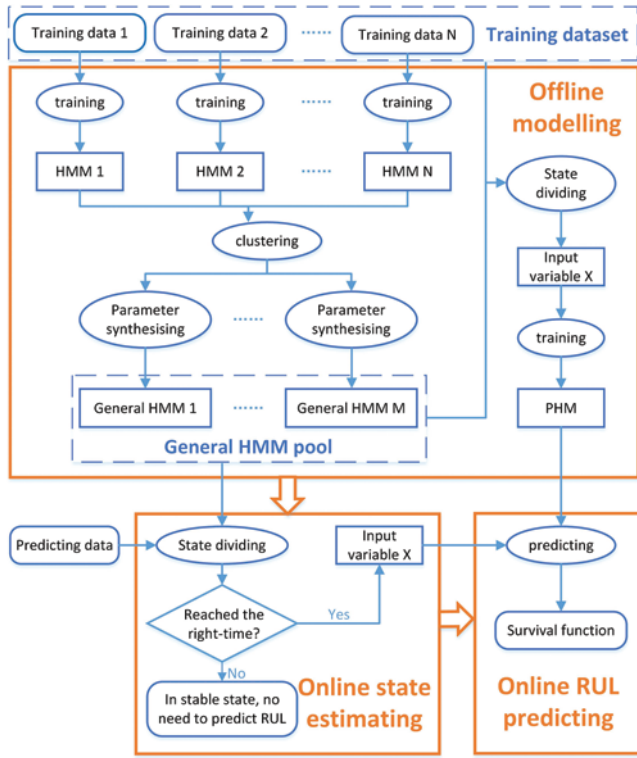


Fig. 1. The procedure of the proposed right-time RUL prediction method

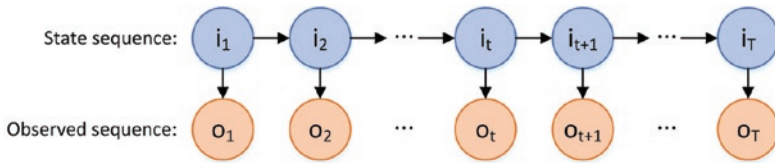


Fig. 2. Construction of HMM

The observation at any given moment is only relevant to the state at that moment, that is:

$$P(o_t|i_t, o_T, i_{T-1}, o_{T-1}, \dots, i_{t+1}, o_{t+1}, i_t, o_t, \dots, i_1, o_1) = P(o_t|i_t), t = 1, 2, \dots, T. \quad (1)$$

The state at any given moment is only related to the state at the last moment, that is:

$$P(i_t|i_{t-1}, o_{t-1}, \dots, i_1, o_1) = P(i_t|i_{t-1}), t = 1, 2, \dots, T. \quad (1)$$

For a standard HMM, the following parameters are defined:

An initial state probability distribution, represents the probability of the object in each state at the initial time,  $\pi = \{\pi_i\}$ , where:

$$\pi_i = P(i_1 = S_i), 1 \leq i \leq N, \sum_i \pi_i = 1. \quad (3)$$

A state transition probability distribution, represents the probability of an object transforms from one state to another,  $\mathbf{A} = \{a_{ij}\}$ , where:

$$a_{ij} = P(i_{t+1} = S_j | i_t = S_i), 1 \leq i, j \leq N, \sum_j a_{ij} = 1. \quad (4)$$

An observation probability distribution related to the hidden states,  $\mathbf{B} = \{b_j(\cdot)\}$ .

For convenience, the above parameters can be abbreviated as a triplet:

$$\lambda = (\mathbf{A}, \mathbf{B}, \pi). \quad (5)$$

According to the above definition, the parameters can be solved with:

$$\lambda = \arg \max_{\lambda} \sum_I P(O|I, \lambda) \cdot P(I|\lambda), \quad (6)$$

where  $I = (i_1, i_2, \dots, i_T)$  and  $O = (o_1, o_2, \dots, o_T)$ .

In the process of general HMM pool establishing, HMMs is established and fitted for each training sample to represent the specific degradation path of the sample. For obtaining some more general degradation paths and reducing the computation load, clustering is applied to these fitted HMMs by the Euclidean distance between parameters of different HMMs. The HMMs in the same cluster represent a similar degradation path, and can be integrated to a general model. For each cluster, a general model is obtained to represent a typical degradation path by synthesizing the parameters of HMMs with:

$$\lambda_m = \frac{\sum_{i=1}^{K_m} \lambda_{m,i} \cdot (d_{m,max} - d_{m,i})}{d_{m,max} \cdot K_m - \sum_{i=1}^{K_m} d_{m,i}}, m = 1, 2, \dots, M, \quad (7)$$

where  $M$  is the number of clusters,  $K_m$  is the number of models in cluster  $m$ ,  $\lambda_m$  is the parameter of general model  $m$ ,  $d_{m,max}$  is the maximum distance of models in cluster  $m$  to the clustering center of cluster  $m$ ,  $d_{m,i}$  is the distance between model  $i$  in cluster  $m$  and the clustering center of cluster  $m$ , and  $\lambda_{m,i}$  is the parameter of model  $i$  in cluster  $m$ .

Consisting of these integrated models, the general HMM pool can be used to estimate the degradation state of the object more generally, and make the state estimation result less affected by model uncertainty.

### 2.1.2. Survival function modelling

In the field of survival analysis, representing the  $m$  variables that have effect on survival probability as:

$$\mathbf{X} = \{x_1, x_2, \dots, x_m\}, \quad (8)$$

then the risk function with the given input variable  $\mathbf{X}$  can be represented as:

$$h(t, \mathbf{X}) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t | T > t, \mathbf{X})}{\Delta t}, \quad (9)$$

where  $T$  is the survival time of the object.

The PHM is a semi-parametric model with a fixed baseline hazard function and the input variable  $\mathbf{X}$  that effects on the baseline hazard function:

$$h(t, \mathbf{X}) = h_0(t) \cdot \exp(\beta \cdot \mathbf{X}), \quad (10)$$

where  $h_0(t)$  is the baseline hazard function,  $\mathbf{X}$  is the input variable and  $\beta$  is the coefficient of  $\mathbf{X}$ .

The maximum likelihood function of the PHM is:

$$L(\beta) = \prod_{i=1}^N \frac{\exp(\beta \cdot X_i)}{\sum_{j:t_j \geq t_i} \exp(\beta \cdot X_j)}, \quad (11)$$

where  $N$  is the number of events,  $X_i$  is the input variable corresponding to event  $i$ ,  $t_i$  is the time that event  $i$  happened.

In this paper, the degradation characteristics extracted from the training samples by the general HMM pool are used as the input variable of PHM, including the duration of each stable state and the mean value of indicator data in each stable state:

$$X = \{D_1, D_2, \dots, D_{N-1}; M_1, M_2, \dots, M_{N-1}\} \quad (12)$$

where  $N$  is the number of states that a whole degradation process contains.

The duration of each stable state reflects the degradation characteristics of the current equipment in the time dimension. The longer duration of stable state indicates that the overall degradation process of the equipment is proceeding at a slower speed, which can provide a reference for the degradation process of critical state. The duration of state  $i$  can be represented as:

$$D_i = t_{i+1} - t_i, 1 \leq i < N, \quad (13)$$

where  $t_i$  is the time that state  $i$  reached.

The mean value of indicator data in each state can reflect the degradation characteristic of the current equipment in the indicator dimension, and the variation of the indicator data in each stable state can also reflect the overall degradation process. The mean value of indicator data in state  $i$  can be represented as:

$$M_i = \frac{\sum_{t=t_i}^{t_{i+1}} I(t)}{D_i}, 1 \leq i < N, \quad (14)$$

where  $I(t)$  is the indicator value at time  $t$ .

When the device degrades to the critical state, its overall degradation path has already appeared. The duration of the previously experienced states and the mean value of the indicator data of these states can well reflect the presented degradation path, and act on the baseline function as the input variable of PHM. Given the input variable, the parameters can be solved by maximizing equation (11) with training samples.

## 2.2. Online state estimating

In the process of online state estimating, the indicator data is input into the general HMM pool to realize continuous status monitoring. For  $M$  HMMs in model pool, the HMM with the maximal log-likelihood is considered as the model with the most similar degradation path to the device, and the estimation result of this HMM is regarded as the current state of the device. The next step of the proposed method depends on the estimated current state. If the critical state not reached, the device is considered to still have a long running time, and there is no need to predict the RUL. Otherwise, the duration of the previously experienced states and the mean value of the indicator data of these states are extracted for RUL prediction in the next step. The procedure of online state estimating is shown in Figure. 3.

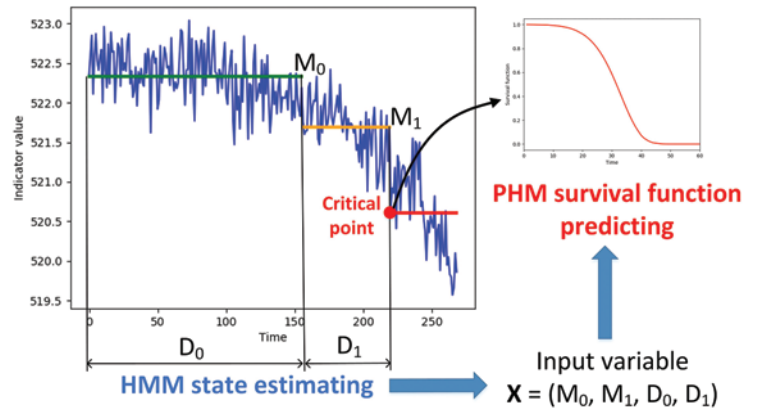


Fig. 3. procedure of online state estimating

## 2.3. Online RUL predicting

Once the critical state reached, the hazard function during the whole critical state can be obtained by substituting the extracted variable into the fitted PHM. With the hazard function, the survival function can be derived by:

$$S(t) = \exp\left(-\int_{t_c}^t h(t) dt\right), \quad (15)$$

where  $t_c$  is the current time and  $t$  is the future time.

The computational load of the calculation process of survival function is low, which can completely meet the requirements of online prediction.

In fact, before the real failure of the device, it can never be determined when the device will fail. Therefore, the probabilistic representation of survival function is more suitable for the uncertainty of RUL. Moreover, most maintenance decisions are made based on the probability distribution rather than the exact RUL value [4, 15, 18].

Although the probability representation of RUL is more meaningful, the method proposed in this paper can still give the value of RUL through expectation or determined probability threshold for decision reference. Examples will be given in the following method demonstration section.

## 3. Method demonstration

As a typical and important mechanical system, aircraft engine is a very active research area for prognostics. To demonstrate the method proposed in this paper, a case study on turbofan engines is conducted. The main components of the aircraft turbine engine include fan, low-pressure compressor (LPC), high-pressure compressor (HPC), low-pressure turbine (LPT), high-pressure turbine (HPT), combustor and nozzle. In this section, the proposed method is demonstrated with the turbofan engine simulation data set FD001 provided by NASA [31]. The data set is generated by the C-MAPSS tool and simulates 21 sensor measurements during engine operation, as detailed in Table 1. The training set contains the simulated data of 21 sensors of 100 engines in the full life cycle. The prediction set has the same data form as the training set but truncated at any time. The RUL of each engine in the prediction set is also given for result comparison. In the last part of this section, the results of the proposed right-time prediction method are analysed, and the superiority is verified by comparing with a traditional SVR-PHM.

Table 1. Description of the sensor measurements for the aircraft turbine engine

Index	Symbol	Description	Units
1	T2	Total temperature at fan inlet	°R
2	T24	Total temperature at LPC outlet	°R
3	T30	Total temperature at HPC outlet	°R
4	T50	Total temperature at LPT outlet	°R
5	P2	Pressure at fan inlet	psia
6	P15	Total pressure in bypass-duct	psia
7	P30	Total pressure at HPC outlet	psia
8	Nf	Physical fan speed	rpm
9	Nc	Physical core speed	rpm
10	EPr	Engine pressure ratio (P50/P2)	--
11	Ps30	Static pressure at HPC outlet	psia
12	Phi	Ratio of fuel flow to Ps30	pps/psi
13	NRf	Corrected fan speed	rpm
14	NRc	Corrected core speed	rpm
15	BPR	Bypass Ratio	--
16	farB	Burner fuel-air ratio	--
17	htBleed	Bleed Enthalpy	--
18	Nf_dmd	Demanded fan speed	rpm
19	PCNfR_dmd	Demanded corrected fan speed	rpm
20	W31	HPT coolant bleed	lbm/s
21	W32	LPT coolant bleed	lbm/s

### 3.1. Model training

In the offline model training process, all training samples in training set are used for HMM modelling and PHM modelling.

Firstly, the variable 3 and 12 among all 21 variables is selected for modelling according to Wang [37]. With assuming that three implicit degradation states exist in the whole degradation process and each state has a specific degree and rate of degradation, three-state HMM is used in this demonstration.

For each training sample, a HMM is fitted. The parameters of the fitted HMM is considered to contain information of the specific degradation path of the corresponding sample. Note that all the modelling, fitting and predicting processes for HMM is achieved by the `hmmlearn` package in Python. The model used in this paper is Gaussian HMM, and the calculation process of model parameters is achieved by expectation maximization (EM) algorithm.

For a better generality and lower computation load, these 100 HMMs fitted with the 100 samples is clustered into five clusters for model pool establishing. For Gaussian HMM used in this demonstration, the observation probability distribution  $\mathbf{B}$  is expressed as a mean value matrix  $\mathbf{U}$  and a covariance matrix  $\mathbf{V}$ . The Euclidean distance between mean value matrix  $\mathbf{U}$  of different HMMs is used as the basis for clustering. After clustering, the HMMs under each cluster are synthesized with equation (7) to get a more general model. Consisting of the five integrated HMMs, general HMM pool has the ability to represent some different but common degradation paths.

Before the PHM established, the input variables from equation (12) have to be extracted from the state estimation results of the training samples. To be consistent with the online prediction, the general HMM pool is used for the state estimation of these training samples. For each sample, the estimation result from the general model with maximum

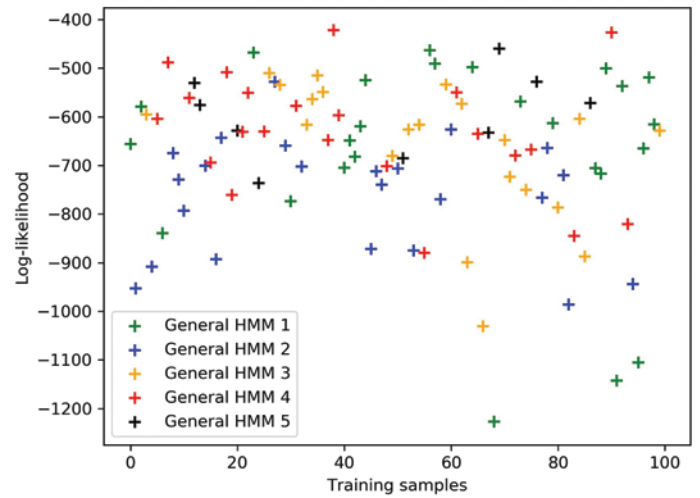


Fig. 4. Log-likelihood and corresponding general model for each sample

log-likelihood is considered to be the most credible. Figure. 4 showed general model with maximum log-likelihood of each training sample and the calculated log-likelihood value. With the general HMM pool, the fittest model specific to each sample rather than a constant model can be used for state estimating with a lower uncertainty.

Figure. 5 shows the proportion of estimated stable state and critical state to the whole life of each training sample. For most in these 100 samples, the critical state shown in red accounts for 10% to 20% of the total life span. Once the critical state reached, the holistic degradation path of the device can be shown from the long-term operation

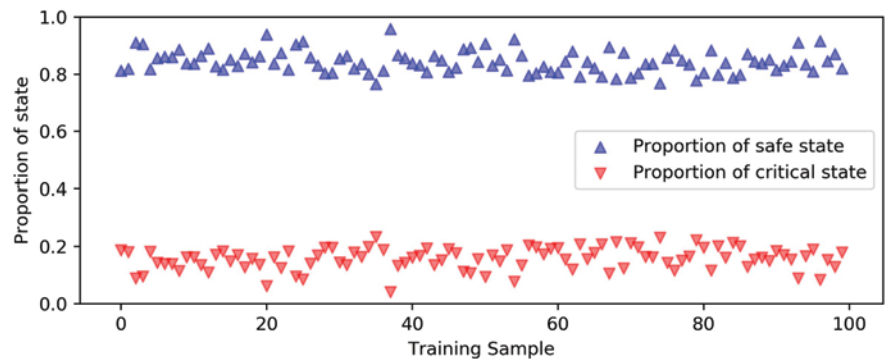


Fig. 5. State estimating result

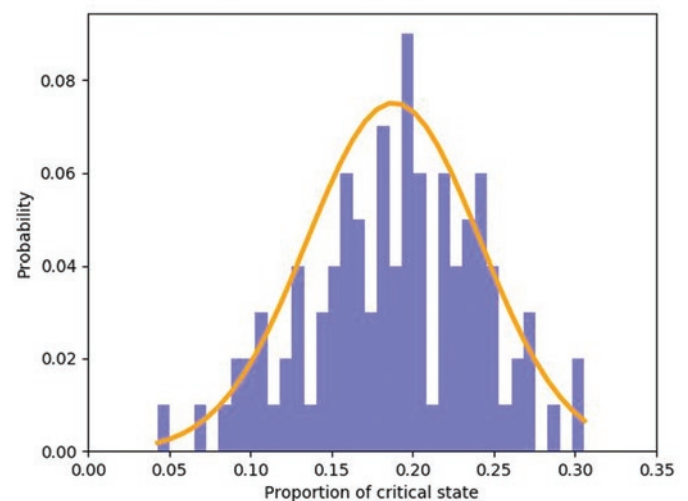


Fig. 6. Distribution of the proportion of critical state in total life



Table 2. Coefficients of PHM

	$D_0$	$D_1$	$M_{0,3}$	$M_{0,12}$	$M_{1,3}$	$M_{1,12}$
Coefficient	-0.03	0.03	-0.24	3.92	0.95	-1.74

data. The right-time RUL prediction with these existing degradation characteristics can not only reduce the uncertainty of prediction results, but also leave enough time for RUL-based decision-making.

In order to quantify the results of state estimating more accurately, the fitted distribution of the proportion of critical state in total life of 100 samples is shown in Figure. 6. The histogram shows the true distribution of samples, and the curve is a fitted normal distribution with mean value of 0.19 and variance of 0.05. Kolmogorov-Smirnov test is performed on the fitting results, with a D value of 0.05 and a P value of 0.98, which is considered to be in line with the fitted normal distribution. The 95% bilateral tolerance interval of the fitted distribution is [0.08,0.29], that is, under the 95% probability, the critical state duration is from 8% to 29% of the total life, and the RUL prediction is made at the moment in 71-92% of the total life. This also proves that the proposed right-time RUL prediction method is reliable for device safety.

With the extracted input variables, the PHM specific for the critical state can be established. Note that all the modeling, fitting, and predicting processes of PHM in this paper is implemented through the lifeline package of Python, in which the baseline hazard function is determined with Breslow’s Method and the model is fitted with Efron’s method for ties and Newton-Rhaphson algorithm.

Then the established PHM is fitted with the real RUL at the time of critical state reached of the training samples. The fitted coefficients of input variables are listed in

Table 2, in which  $D_0$  and  $D_1$  represent the duration of the first state and second state estimated by the three-state HMM respectively,  $M_{0,3}$

and  $M_{1,3}$  represent the mean value of variable 3 in the first state and second state respectively, similarly,  $M_{0,12}$  and  $M_{1,12}$  is the mean value of variable 12 in the two state.

Figure. 7 presents a sensitivity analysis of the fitted PHM to the input variables. For each input variables, five values are equably selected within the variation range of all training samples. Changing of survival function caused by changing of single variable is shown, and the changed survival function is compared with the baseline survival function. The picture shows that the variation of each input variable has a significant impact on the survival function, so it is necessary to obtain the most appropriate input variable for the sample through multiple models in the model pool.

### 3.2. State estimating and RUL predicting

In the online state estimating process, the general HMM pool is used to continuously monitor the equipment state according to the input indicators. The state estimating for all 100 predicting samples in the data set is carried out. According the state estimating results, 19 samples have reached the critical state and the RUL of these samples need to be predicted. The remaining samples are still in a stable state and are considered to have a long running time and do not need RUL prediction. Figure. 8 shows the results of the state estimation of the 19 samples. The blue line is the input monitoring indicator with abscissa the time, and ordinate the indicator value. The green line, the yellow line, and the red line are three states estimated by general HMM pool, respectively. The states represented by the green line and the yellow line are taken as stable state, and the state represented by the red line is critical state.

For these 19 samples which reached the critical state, the degradation characteristics in equation (12) are extracted from the existing degradation data, and taken as the input of the fitted PHM to obtain the hazard rate of the remaining life time. Then the survival function

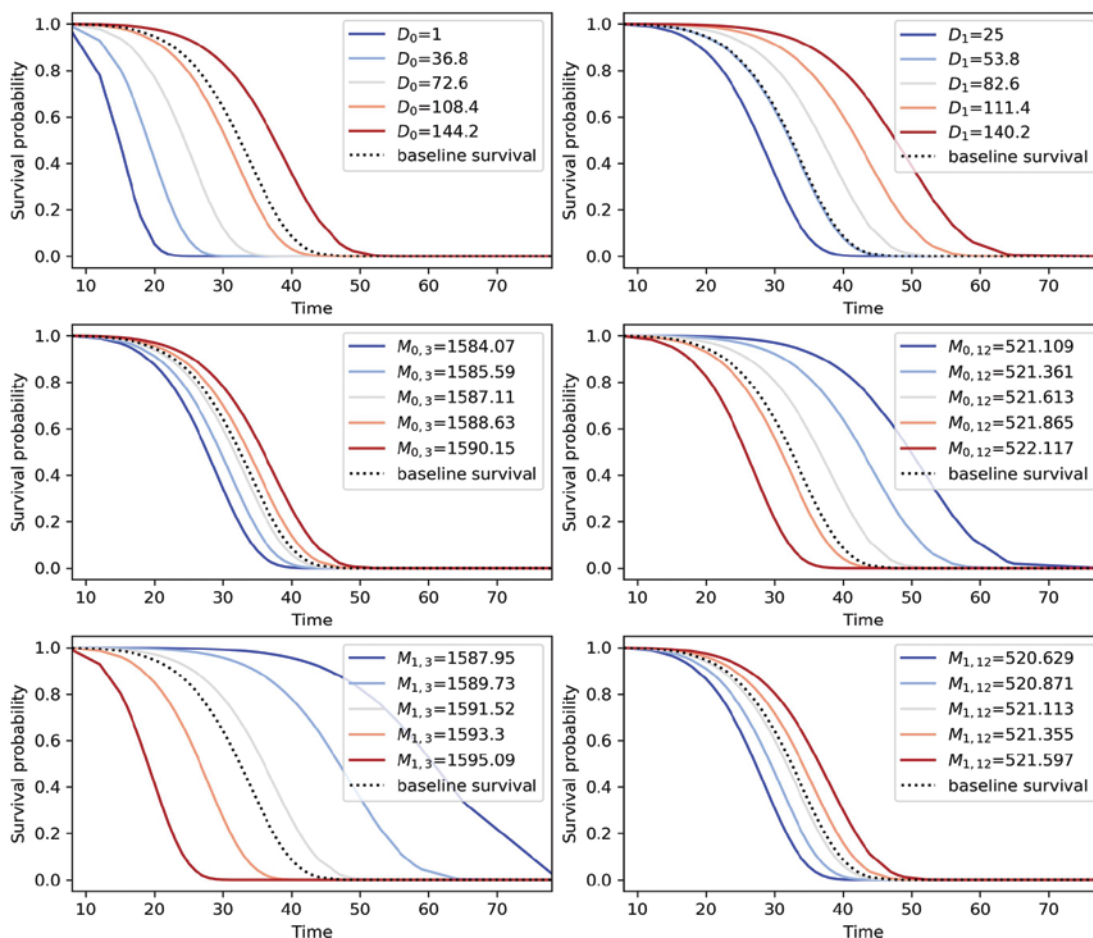


Fig. 7. Sensitivity of the fitted model to input variables

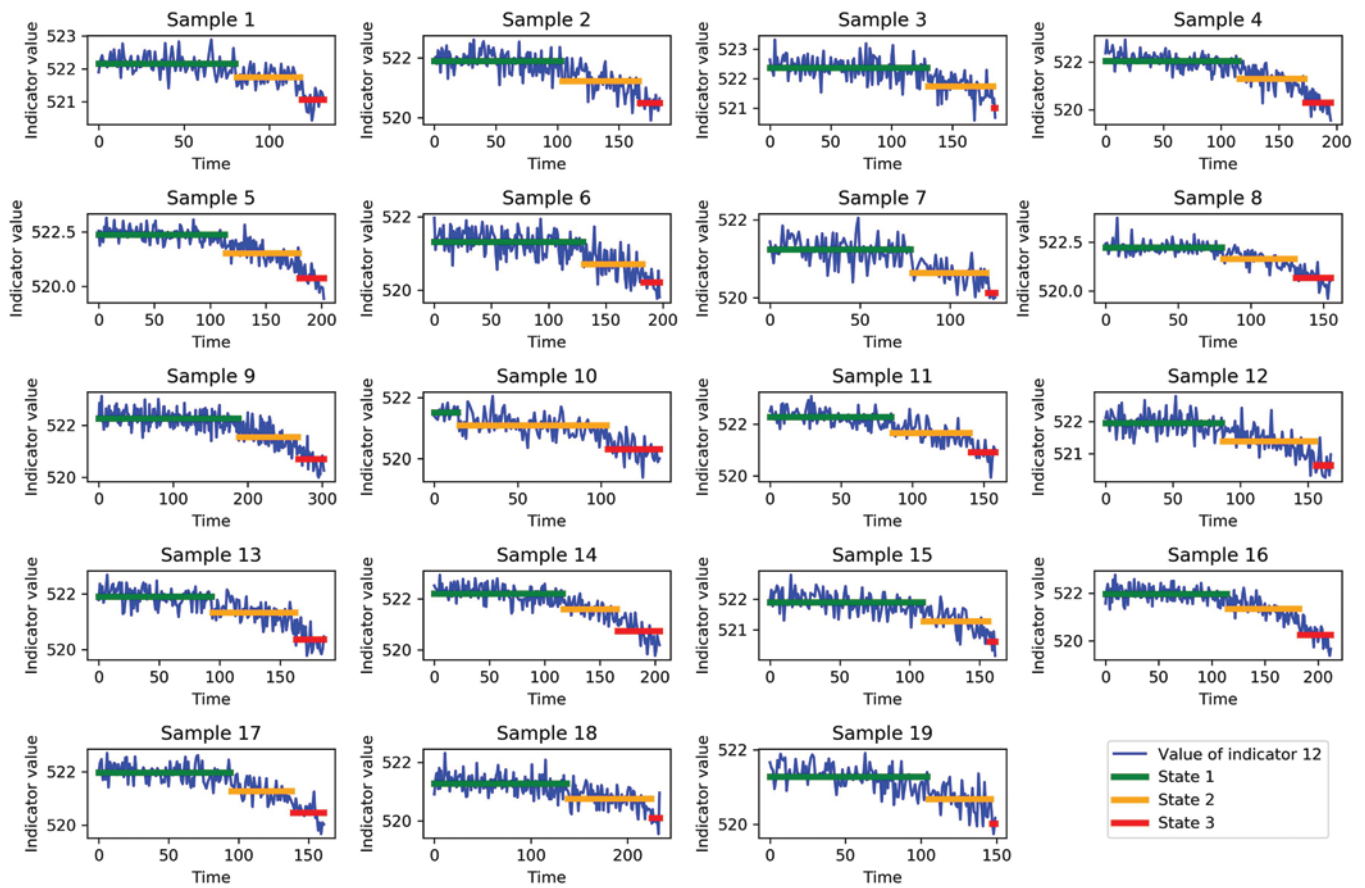


Fig. 8. State estimating results of the 19 samples

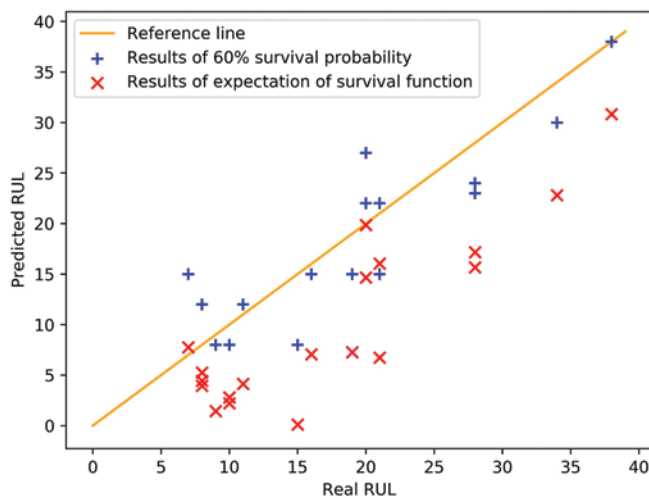


Fig. 9. Survival function of the 19 samples

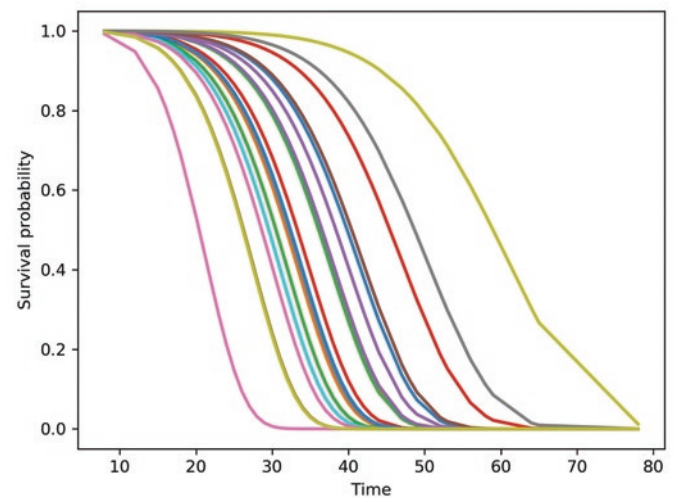


Fig. 10. RUL predicted with survival function

is obtained according to equation (15). The survival function from the moment that critical state reached of the 19 samples is shown in Figure. 9. The survival curves of different samples are clearly differentiated, indicating that the extracted degradation characteristics can reflect the unique degradation path of the sample and act on the baseline function, so that the obtained survival curve can adapt to different degradation processes through different input variables and be more compatible with the real degradation path.

The survival function reflects the survival probability of the device at different moments and can be used as reference for maintenance or replacement decision-making. Besides, the survival function can be used for predicting RUL value by expectation or a given threshold of

survival probability. Figure. 10 shows the predicted RUL value from the survival function in Figure. 9. The abscissa is the real life of the sample, and the ordinate is the corresponding predicted RUL. The yellow line in the figure is the reference line with slope 1. The degree of closeness between the prediction results and the reference line can indicate the prediction accuracy. The blue dots in the figure is the predicted RUL of 19 samples with a survival probability of 0.6 as the threshold. All these predicted points nearby the reference line, indicating that the prediction results are accurate. When the expectation of the survival function is taken as RUL value, the predicted results are shown as red dots in the figure. Most of the predicted points are below the reference line, indicating that the predicted results are generally

smaller than the real RUL, which can provide a reference for a conservative decision-making.

### 3.3. Results analysis and comparison

In this part, the proposed right-time prediction method is compared with a traditional method based on regression model and PHM. In the traditional method, the future value of indicators has to be predicted firstly, and then the predicted value is taken as the input variable of a PHM to obtain the survival function. There are two processes in traditional method, offline training and online prediction. In the training process, the total lifetime data of the 100 training samples is taken as the input variable of PHM for model fitting. To make an effective comparison, the prediction samples is selected the same as the 19 samples used in the right-time prediction. In the prediction process, the future indicator value is predicted by on SVR firstly, and then the predicted indicator value is taken as the input variable of PHM to obtain the survival function. Figure. 11 shows the survival curve of the 19 samples predicted by SVR-PHM from truncation time.

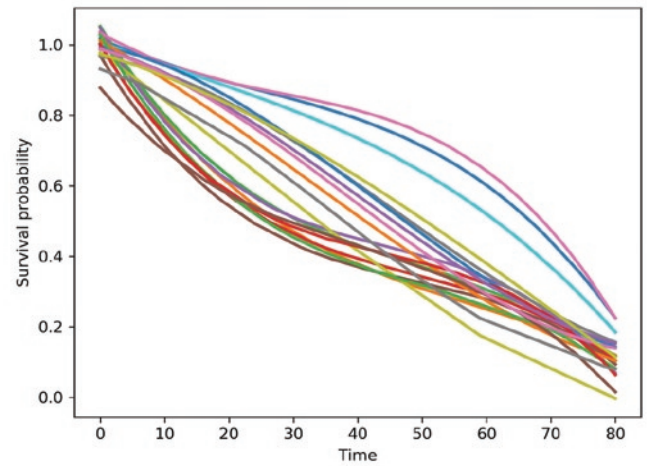


Fig. 11. Survival function of 19 samples by SVR-PHM

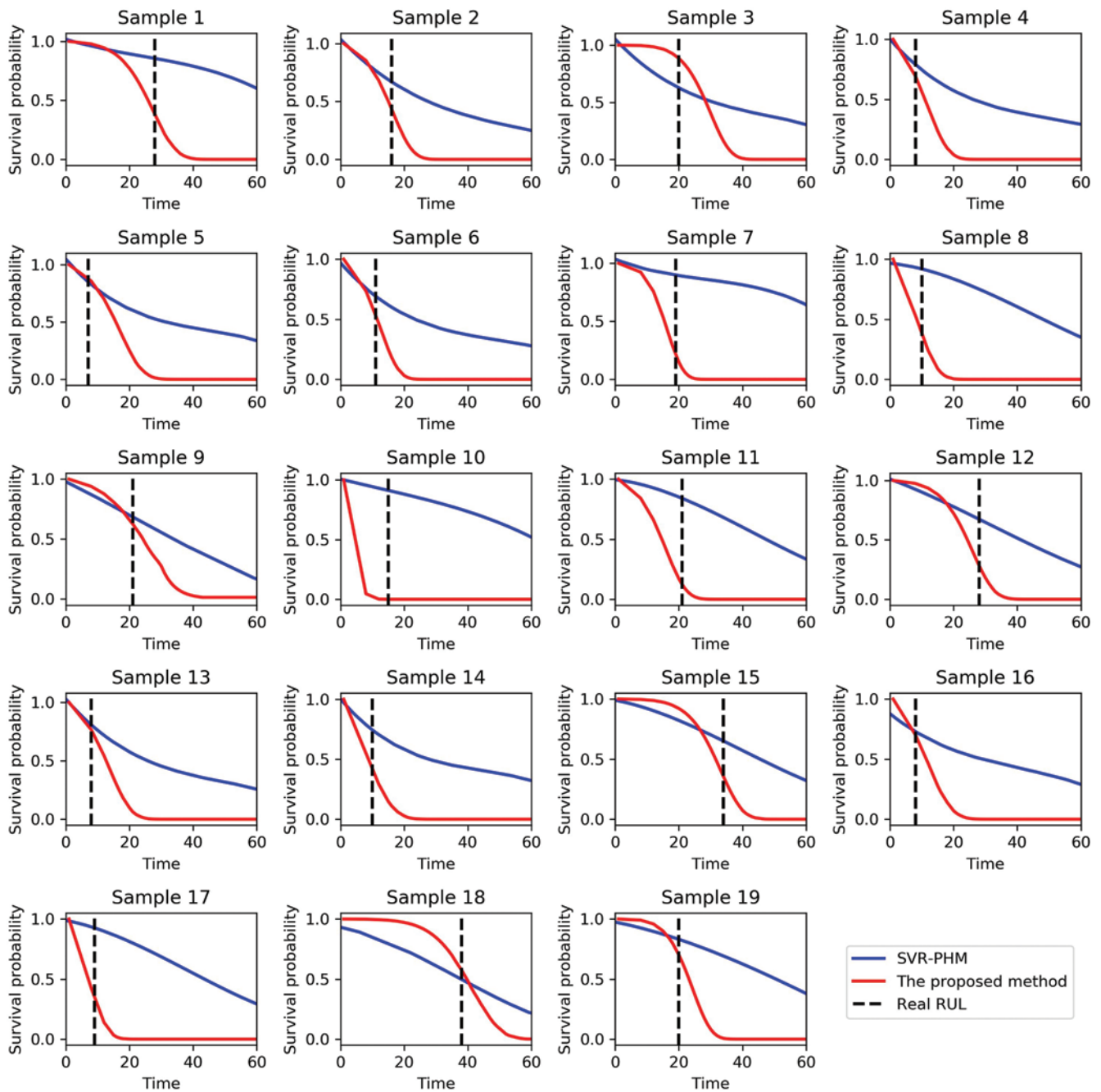


Fig. 12. Comparison of the predicted survival function

Table 3. Predicted survival probability at failure time

	Sample 1	Sample 2	Sample 3	Sample 4
SVR-PHM	0.85	0.66	0.65	0.78
The proposed method	0.33	0.41	0.87	0.56
	Sample 5	Sample 6	Sample 7	Sample 8
SVR-PHM	0.83	0.68	0.91	0.91
The proposed method	0.86	0.60	0.22	0.39
	Sample 9	Sample 10	Sample 11	Sample 12
SVR-PHM	0.68	0.91	0.83	0.66
The proposed method	0.61	0	0.11	0.23
	Sample 13	Sample 14	Sample 15	Sample 16
SVR-PHM	0.78	0.73	0.64	0.73
The proposed method	0.74	0.42	0.33	0.69
	Sample 17	Sample 18	Sample 19	
SVR-PHM	0.92	0.51	0.82	
The proposed method	0.28	0.55	0.67	

Table 4. Comparison of prediction errors

	The proposed method	SVR-PHM
MSE	16.21	125.32
RMSE	4.03	11.19
MAE	3.47	8.58

Figure 12 gives a comparison between the survival function of the selected 19 samples predicted by the proposed right-time prediction method and traditional SVR-PHM. The blue line is the survival function obtained by SVR-PHM, the red line is obtained by the proposed method, and the black dotted line corresponds to the real failure time of the sample. For most of the 19 prediction samples, the predicted failure probability of the proposed method is close to or significantly lower than that of SVR-PHM at the failure time, which indicates the superiority of the proposed method.

For a more accurate comparison, Table 3 lists the survival probability of the 19 samples predicted by the two methods at real failure time. Among the 19 samples, only one sample has a significantly higher survival probability with the proposed method than with SVR-PHM, which marked red in table. For four samples in table, the predicted survival probability by the two methods is similar with difference less than 0.05, and these four samples are marked orange in table. For the left 14 samples, the survival probability predicted by the proposed method is significantly less than that of SVR-PHM.

Table 4 gives the MSE, RMSE and MAE comparison of RUL prediction results of the two methods. For SVR-PHM, the RUL is determined by survival probability with a threshold of 0.8, and the minimum error can be achieved with this threshold. For the proposed right-time prediction method, the threshold is chosen as 0.6.

In this demonstration, the proposed method achieves better prediction results than the traditional SVR-PHM, the reasons for the superiority of the proposed method are analyzed:

## References

- Bai J M, Zhao G S, Rong H J. Novel direct remaining useful life estimation of aero-engines with randomly assigned hidden nodes. *Neural Computing and Applications* 2020; 32(18): 14347-14358, <https://doi.org/10.1007/s00521-019-04478-1>.
- Baraldi P, Mangili F, Zio E. A belief function theory based approach to combining different representation of uncertainty in prognostics. *Information Sciences* 2015; 303: 134-149, <https://doi.org/10.1016/j.ins.2014.12.051>.
- Cai B P, Shao X Y, Liu Y H, Kong X D, Wang H F, Xu H Q, Ge W F. Remaining useful life estimation of structure systems under the influence of multiple causes: subsea pipelines as a case study. *IEEE Transactions on Industrial Electronics* 2020; 67(7): 5737-5747, <https://doi.org/10.1109/TIE.2020.3000000>.

The proposed method estimating the degradation state of the device firstly, and the RUL is predicted after the critical state reached. Therefore, only the degradation process of critical state need to be modeled by PHM. Compared with the traditional method which modeling the whole degradation process, the proposed method reduces the difficulty of modelling and makes the established model more consistent with the real degradation process.

The input variable of PHM in the proposed method is the characteristics extracted from the real long run degradation data, while the traditional method takes indicator value as the input variable. For the traditional method, the indicator value during the remaining lifetime must be predicted before RUL prediction. The proposed method avoids the uncertainty in the process of indicator value prediction.

The traditional method takes indicator value as the input variable of PHM. This makes the model highly affected by the indicator fluctuation. The indicator volatility increased the difficulty of model fitting, makes the model difficult to accurately express the real degradation process. The proposed method takes the characteristics extracted from the whole degradation process as the input variable of PHM, and can make the RUL prediction with a global perspective.

## 4. Conclusion

This paper presents a right-time RUL prediction method for reducing the prognostics uncertainty of mechanical systems under unobservable degradation based on HMM and PHM. In this method, HMM is used for state estimating, and PHM is used for obtaining the survival function. With the concept of right-time prediction, RUL is only predicted after the last degradation state reached. The prediction uncertainty can be greatly reduced by maximizing the available degradation data and meanwhile leaving sufficient time for decision-making by right-time prediction. By demonstrating on a turbofan engine degradation simulation data set and comparing with a traditional SVR-PHM, the effectiveness of the proposed method is proved.

The main contributions of this paper are as follows:

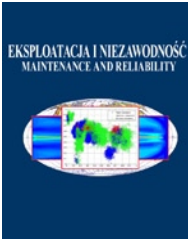
According to the authors know, the concept of right-time prediction is first put forward in this paper. Correspondingly, the offline modelling and online predicting parts of traditional RUL prediction are divided into more accurate offline modelling, online state estimating and online RUL predicting three parts. When the RUL is predicted, the whole degradation trend is more clear for better modelling, and the previous degradation data can be used for uncertainty reduction.

Taking the degradation characteristics extracted from the long run data as the input of PHM. It can not only avoid the uncertainty from indicator predicting, but also avoid the influence of the RUL prediction results from indicator fluctuations. With the data set, it is verified that the degradation characteristics extracted by partitioning the degradation states are effective.

While good prediction results have been obtained by the proposed method, the method still has considerable room for improvements. At present, the input variables of PHM are degradation characteristics extracted only from the previous data. How to update the survival function with the indicator value of the last degradation state, and that will be focused on in further research.

- doi.org/10.1109/TIE.2019.2931491.
4. Chen Z, Li Y, Xia T, Pan E. Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy. *Reliability Engineering & System Safety* 2019; 184: 123-136, <https://doi.org/10.1016/j.res.2017.09.002>.
  5. Cox D R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 1972; 34(2): 187-220, <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>.
  6. Deng Y, Bucchianico A D, Pechenizkiy M. Controlling the accuracy and uncertainty trade-off in RUL prediction with a surrogate Wiener propagation model. *Reliability Engineering & System Safety* 2020; 196: 1-10, <https://doi.org/10.1016/j.res.2019.106727>.
  7. Djeziri M A, Benmoussa S, Sanchez R. Hybrid method for remaining useful life prediction in wind turbine systems. *Renewable Energy* 2018; 116: 173-187, <https://doi.org/10.1016/j.renene.2017.05.020>.
  8. Du Y, Wu T, Zhou S, Makis V. Remaining useful life prediction of lubricating oil with dynamic principal component analysis and proportional hazards model. *Proceedings of the Institution of Mechanical Engineers Part J-Journal of Engineering Tribology* 2020; 234(6): 964-971, <https://doi.org/10.1177/1350650119874560>.
  9. Engel S J, Gilmartin B J, Bongort K, Hess A. Prognostics, the real issues involved with predicting life remaining. *IEEE Aerospace Conference*. 2000: 457-469.
  10. Equeter L, Ducobu F, Rivière-Lorphèvre E, Serra R, Dehombreux P. An analytic approach to the Cox proportional hazards model for estimating the lifespan of cutting tools. *Journal of Manufacturing and Materials Processing* 2020; 4(27): 1-13, <https://doi.org/10.3390/jmmp4010027>.
  11. Fink O, Wang Q, Svensén M, Dersin P, Ducoffe M. Potential, Challenges and Future Directions for Deep Learning in Prognostics and Health Management Applications. *Engineering Applications of Artificial Intelligence* 2020; 92: 1-15, <https://doi.org/10.1016/j.engappai.2020.103678>.
  12. Hu J W, Chen P. Predictive maintenance of systems subject to hard failure based on proportional hazards model. *Reliability Engineering & System Safety* 2020; 196: 1-9, <https://doi.org/10.1016/j.res.2019.106707>.
  13. Hu Y, Li H, Shi P, Chai Z, Wang K, Xie X, Chen Z. A prediction method for the real-time remaining useful life of wind turbine bearings based on the Wiener process. *Renewable Energy* 2018; 127: 452-460, <https://doi.org/10.1016/j.renene.2018.04.033>.
  14. Jacyna M, Semenov I. Models of vehicle service system supply under information uncertainty. *Eksplotacja i Niezawodność - Maintenance and Reliability* 2020; 22(4): 694-704, <https://doi.org/10.17531/ein.2020.4.13>.
  15. Lee D, Choi D. Analysis of the reliability of a starter-generator using a dynamic Bayesian network. *Reliability Engineering & System Safety* 2020; 195: 1-11, <https://doi.org/10.1016/j.res.2019.106628>.
  16. Li H, Wang W, Li Z W, Dong L Y, Li Q Z. A novel approach for predicting tool remaining useful life using limited data. *Mechanical Systems and Signal Processing* 2020; 143: 1-22, <https://doi.org/10.1016/j.ymsp.2020.106832>.
  17. Li X P, Huang H Z, Li F Q, Ren L M. Remaining useful life prediction model of the space station. *Eksplotacja i Niezawodność - Maintenance and Reliability* 2019; 21(3): 501-510, <https://doi.org/10.17531/ein.2019.3.17>.
  18. Liang Z, Parlikad A K. Predictive group maintenance for multi-system multi-component networks. *Reliability Engineering & System Safety* 2020; 195: 1-18, <https://doi.org/10.1016/j.res.2019.106704>.
  19. Lin L, Sun Z, Xu X, Zhang K. Multi-zone proportional hazard model for a multi-stage degradation process. *8th International Manufacturing Science and Engineering Conference*. 2013: 1-8, <https://doi.org/10.1115/MSEC2013-1113>.
  20. Liu D, Wang S P, Tomovic M M. Degradation modeling method for rotary lip seal based on failure mechanism analysis and stochastic process. *Eksplotacja i Niezawodność - Maintenance and Reliability* 2020; 22(3): 381-390, <https://doi.org/10.17531/ein.2020.3.1>.
  21. Ma B, Yan S, Wang X, Chen J, Zheng C. Similarity-based failure threshold determination for system residual life prediction. *Eksplotacja i Niezawodność - Maintenance and Reliability* 2020; 22(3): 520-529, <https://doi.org/10.17531/ein.2020.3.15>.
  22. Maior C B S, Moura M D C, Lins I D. Particle swarm-optimized support vector machines and pre-processing techniques for remaining useful life estimation of bearings. *Eksplotacja i Niezawodność - Maintenance and Reliability* 2019; 21(4): 610-619, <https://doi.org/10.17531/ein.2019.4.10>.
  23. Man J, Zhou Q. Prediction of hard failures with stochastic degradation signals using Wiener process and proportional hazards model. *Computers & Industrial Engineering* 2018; 125: 480-489, <https://doi.org/10.1016/j.cie.2018.09.015>.
  24. Ossai C I. Prognostics health estimation of lithium-ion batteries in charge-decay estimation uncertainties - A comparative analysis. *International Journal of Prognostics and Health Management* 2018; 9(2): 1-8.
  25. Qiu G, Gu Y, Chen J. Selective health indicator for bearings ensemble remaining useful life prediction with genetic algorithm and Weibull proportional hazards model. *Measurement* 2020; 150: 1-14, <https://doi.org/10.1016/j.measurement.2019.107097>.
  26. Rabiner L R. A tutorial on hidden Markov-models and selected applications in speech recognition. *Proceedings of the IEEE* 1989; 77(2): 257-286, <https://doi.org/10.1109/5.18626>.
  27. Ramezani S, Moini A, Riahi M. Prognostics and health management in machinery: A review of methodologies for RUL prediction and roadmap. *International Journal of Industrial Engineering and Management Science* 2019; 6(1): 38-61.
  28. Razavi-Far R, Chakrabarti S, Saif M, Zio E. An integrated imputation-prediction scheme for prognostics of battery data with missing observations. *Expert Systems with Applications* 2019; 115: 709-723, <https://doi.org/10.1016/j.eswa.2018.08.033>.
  29. Sankararaman, Shankar. Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction. *Mechanical Systems & Signal Processing* 2015; 52: 228-247, <https://doi.org/10.1016/j.ymsp.2014.05.029>.
  30. Sankararaman S, Goebel K. Uncertainty in prognostics and systems health management. *International Journal of Prognostics & Health Management* 2015; 6: 1-11.
  31. Saxena A, Goebel K, Simon D, Eklund N. Damage propagation modeling for aircraft engine run-to-failure simulation. *2008 international conference on prognostics and health management*. 2008: 1-9, <https://doi.org/10.1109/PHM.2008.4711414>.
  32. Sikorska J, Hodkiewicz M, Ma L. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing* 2011; 25(5): 1803-1836, <https://doi.org/10.1016/j.ymsp.2010.11.018>.
  33. Sun J, Zuo H, Wang W, Pecht M G. Prognostics uncertainty reduction by fusing on-line monitoring data based on a state-space-based degradation model. *Mechanical Systems and Signal Processing* 2014; 45(2): 396-407, <https://doi.org/10.1016/j.ymsp.2013.08.022>.
  34. Tayade A, Patil S, Phalle V, Kazi F, Powar S. Remaining useful life (RUL) prediction of bearing by using regression model and principal component analysis (PCA) technique. *Vibroengineering Procedia* 2019; 23: 30-36, <https://doi.org/10.21595/vp.2019.20617>.

35. Tran V T, Hong T P, Yang B S, Tan T N. Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine. *Mechanical Systems & Signal Processing* 2012; 32: 320-330, <https://doi.org/10.1016/j.ymssp.2012.02.015>.
36. Wang H Y, Song W Q, Zio E, Kudreyko A, Zhang Y J. Remaining useful life prediction for Lithium-ion batteries using fractional Brownian motion and Fruit-fly Optimization Algorithm. *Measurement* 2020; 161: 1-9, <https://doi.org/10.1016/j.measurement.2020.107904>.
37. Wang T. Trajectory similarity based prediction for remaining useful life estimation. University of Cincinnati; 2010.
38. Xiuli, Wang, Bin, Jiang, Ningyun, Lu. Adaptive relevant vector machine based RUL prediction under uncertain conditions. *ISA Transactions* 2019; 87: 217-224, <https://doi.org/10.1016/j.isatra.2018.11.024>.
39. Yang C, Zhang Y, Xu X, Li W. Molecular subtypes based on DNA methylation predict prognosis in colon adenocarcinoma patients. *Aging* 2019; 11(24): 11880-11892, <https://doi.org/10.18632/aging.102492>.
40. You M Y, Li L, Meng G, Ni J. Two-zone proportional hazard model for equipment remaining useful life prediction. *Journal of Manufacturing Science and Engineering-Transactions of the Asme* 2010; 132(4): 1-6, <https://doi.org/10.1115/1.4001580>.
41. Yu W N, Kim I Y, Mechefske C. An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme. *Reliability Engineering & System Safety* 2020; 199: 1-12, <https://doi.org/10.1016/j.ress.2020.106926>.
42. Zhang W, Zhang Y. Integrated survival analysis of mRNA and microRNA signature of patients with breast cancer based on Cox model. *Journal of Computational Biology: a Journal of Computational Molecular Cell Biology* 2020, <https://doi.org/10.1089/cmb.2019.0495>.
43. Zhou Q, Son J, Zhou S, Mao X, Salman M A. Remaining useful life prediction of individual units subject to hard failure. *IIE Transactions* 2014; 46(10): 1017-1030, <https://doi.org/10.1080/0740817X.2013.876126>.



Article citation info:

Rośkowicz M, Godzimirski J, Jaształ M, Gąsior J. Improvement of fatigue life of riveted joints in helicopter airframes. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 165–175, <http://dx.doi.org/10.17531/ein.2021.1.17>.

## Improvement of fatigue life of riveted joints in helicopter airframes

Marek Rośkowicz<sup>a</sup>, Jan Godzimirski<sup>a</sup>, Michał Jaształ<sup>a\*</sup>, Jarosław Gąsior<sup>b</sup>

Indexed by:



<sup>a</sup>Department of Mechatronics, Armament and Aviation, Military University of Technology, ul. gen. Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland

<sup>b</sup>Military Centre for Standardization, Quality and Codification, ul. Nowowiejska 28a, 00-909 Warsaw, Poland

### Highlights

- Titanium driven blind bolts improve fatigue life of repaired airframes of helicopters.
- Hybrid (rivet & adhesive) joints also increase fatigue life of repaired airframe.
- Better fatigue life for less rigid adhesive compound with thin layer of glass fabric.
- Proposed joints reduce the problem of secondary fatigue damage to the repaired airframe.

### Abstract

Using original cold-formed rivets in repairs of airframes of helicopters is difficult due to no access to inside parts of the airframe. Thus, the main aim of the study was to investigate the possibility to use the blind rivets or hybrid joints by verification the fatigue performance of such joints that must be better than with original rivets. Riveted and hybrid joints have been experimentally tested under static and fatigue loads. Furthermore, numerical calculations of stress distribution for strapped joint have been conducted. The test results covered fatigue life of lap joints and models of repaired airframe sheets using ordinary mushroom head rivets ref. 3558A-4-10, titanium driven blind bolts with pin, ref. MBF2110AB-05-150 and modified hybrid joints. Using titanium driven blind bolts with pin instead of ordinary hammer-bucked rivets, can improve the fatigue life of element made of aluminum alloy AW 2024T3. There are advantages of replacing riveted joints with modified hybrid (rivet & adhesive) joints in threefold increase in fatigue life of repaired airframe structures.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

riveted joints; hybrid joint; fatigue life, titanium driven blind bolts with pin, ordinary mushroom head rivets.

## 1. Introduction

Currently, there are many aircraft in operation around the world with a service life of more than 30 years. As an example, in the Polish Air Forces there are many helicopters in operation for over 35 years (e.g. Mi-24 helicopters). Many of them have fatigue problems in the riveted joints of the structural components. The damage is in the form of loose rivets, fatigue cracks of the rivets and fatigue cracks in the joined material in the vicinity of the mounting holes [21]. Maintenance is part of operating costs which include repairs, overhauls and replacement of damaged parts. Hence, repair technologies play an important role in maintenance costs and readiness [25, 38].

Airframes of helicopters are joined together mainly with cold-formed rivets due to the materials of which they are made, i.e., aluminum alloys of the AW 2xxx group [43]. In manufacturing processes, ordinary rivets with heads of various types are usually used, made of aluminum alloys with shearing strength  $R_t = 245$  MPa [40]. To a lesser extent, blind rivets are used in inaccessible sites where no bucking bar necessary in the process can be used [10,11].

Repair of damaged joints using the prescribed restorative technology solves the problem to a limited extent. The need for bilateral access to the repair node is a major constraint in the repair of semi-monocoque helicopter structures. This is particularly the case with the need for rapid repair of small-scale combat damage (bullet shot hole) which should be made in a short period of time. In-service ex-

perience indicates that there is a phenomenon of secondary fatigue damage which may occur in the repair nodes within a few months of the repair [42].

Hence, there is a need for more efficient solutions to improve the fatigue life of repair nodes. The solution may be to “shift” the sites of fatigue crack initiation beyond the critical cross-sections of the mounting holes or reduce the stress concentration factors in the vicinity of the holes. Moving the crack initiators beyond the critical zones of the rivet holes can be achieved by generating high pressure forces in the vicinity of the mounting holes. For this purpose, instead of riveted joints, bolted joints should be used, which, however, still does not solve the problem of limited two-sided access to mechanical fasteners in the assembly procedure. Thus, authors proposed that, the target can be achieved by replacing conventional rivets with a blind bolt fastener which have much higher clamping force or by using hybrid joints i.e. mechanical-adhesive joints. It is worth mentioning, that in the publications on hybrid joint solutions known to authors, nobody take into account the problem of the destruction of the adhesive joint by the pressures generated by mechanical joints during their assembly, therefore authors proposed original modification of the adhesive joints with layers of glass fabric. Nevertheless, it should be proved that the fatigue performance of such joints is at least as good with original rivets.

The problems of durability and fatigue strength of riveted joints are widely described in the literature and authors provides a brief review

(\*) Corresponding author.

E-mail addresses: M. Rośkowicz - [marek.roskowicz@wat.edu.pl](mailto:marek.roskowicz@wat.edu.pl), J. Godzimirski - [jan.godzimirski@wat.edu.pl](mailto:jan.godzimirski@wat.edu.pl), M. Jaształ - [michal.jaształ@wat.edu.pl](mailto:michal.jaształ@wat.edu.pl), J. Gąsior - [j.gasior@ron.mil.pl](mailto:j.gasior@ron.mil.pl)

of a literature search on this topic in next paragraph. Furthermore, authors reviewed the papers focused on the bonded repair of aluminium structures [26] and composite structures [19, 20], adhesive materials [3], fatigue strength [1,6], static strength after impact [22], environmental durability [7, 17, 23], analytical methods [8, 9] and finite element method [16, 27, 28]. In the 2017 Budhe et al. [4] issued valuable analysis covers articles published from 2009 to 2016 presenting an updated review of adhesively bonded joints. The papers described current trends regarding i.e. improving fatigue life of joints. Budhe et al. analysed main aspects over conventional joining method including mechanical fasteners and adhesive. His conclusions pointed out stress concentration as major cause of joint damage. The nowadays solution to prevent this phenomenon is to construct hybrid joints, fix with appropriate torque and filling adhesive which gradually take place in aviation [2, 5].

World literature on the repair of riveted joints is very specific. Namely, there are publications which provide dedicated solutions for specific materials of airframe and fasteners, repairs technology and geometry of structural elements of fuselage. The scientific approach to designing and assessing repairs before implementing probably started in the early 1970's [41]. Unfortunately, a large number of literature items on this subject are in the form of organizational reports or presented in conference proceedings, both having limited accessibility. Adopting repair guidelines given by the manufacturer for typical minor damages is a routine activity with many users. Procedures not listed in such repair manuals, naturally call for intervention of the scientists for detailed research on this case [18]. Thus, presented work is a response to the urgent need to develop a modern repair procedure for the aging fleet of the aircraft including Mi-24 helicopter fleet.

## 2. The main problems of durability and fatigue strength of riveted joints

The topic of fatigue of riveted joints is complex due to a multitude of variables related to the structure of the joint, the manufacturing process, and loading conditions during operation. The durability and fatigue strength of riveted joints is determined by the following factors, among others: quality of the surface of the rivet hole, rivet type and material of which it is made, how the rivet fits with the installation hole, potential burrs on the surface of the riveted joint, and the force with which the rivet clamps the elements together. To characterize and describe the impact of those factors, fatigue tests are conducted, initiated by Segerfröjd et al. [34].

Depending on the rivet type and rivet clamping force, fatigue cracks in the sheets to be joined can occur in various locations. There is particularly extensive literature on how rivet clamping force affects the site where cracks first occur, e.g., Hartman [14], Schijve [30], Müller

[24], Harish [13], Skorupa [36]. The sites of fatigue damage of joints for rivets with countersunk and protruding heads are the following:

- fatigue cracks are initiated and propagate in the net section (in the cross-section of the rivet center line) on the edge of the hole, where two overlapping sheets join together, Fig. 1.a;
- fatigue cracks initiate away from the rivet hole, but propagate through the hole, and they are usually slightly offset relative to the rivet center line, Fig. 1.b;
- fatigue cracks initiate above the rivet hole on the edge of the clamped area delineated by the rivet head on the sheet material, and propagate away from rivet holes, Fig. 1.c.

The initiation of fatigue cracks in the net section of the sheet on the edge of rivet hole or in the dimple under the rivet head is shown in Fig. 1a. It occurs both for rivets with countersunk and protruding heads, if the clamping force is limited and a small part of the load is transferred through friction between the sheets.

The nucleation of the crack shown in Fig. 1b. results from a modified distribution of residual stress with increased rivet head squeezing force resulting in better clamping. A significant effect of the riveting force on the moving the location where the crack is initiated away from the net section in the rivet center line is very well demonstrated in research conducted by Müller [24] and presented in Fig. 2. Notably, this type of crack initiation is largely produced by fretting (friction and corrosion wear). In assessing the effect of fretting on the initiation of cracks, we should also mention that the accumulation of the products of wear between the lapped surfaces significantly increases the friction coefficient. This has, obviously, large effect on the loads transferred by friction. For instance, in studies on fretting conducted by Szolwinski [39] for 2024-T351 alloy, the initial value of  $\mu$  friction coefficient with no fretting was about 0.15. However, already after several thousands of cycles, a steep increase in that factor was observed, which was related to the formation of fretting products, until friction coefficient of 0.65 was reached. Interestingly, the growth of friction coefficient observed in the studies (Fig. 3) was consistent, both qualitatively and quantitatively, with increased friction force during fatigue studies conducted by Harman [15]. Friction between the lapped surfaces of the parts to be joined significantly contributes to the transmission of the load occurring at the joint. Experimental studies demonstrate that increasing friction force with growing number of load cycles stabilizes on a certain level and is then responsible for the transfer of 80%-90% of load by friction.

For correctly made riveting, the clamping force is very large, and fatigue cracks will initiate and further propagate above the hole (Fig. 1c). This case is also illustrated in Fig. 2 by the study results presented as points in the "above hole" area in the chart above. In this case, the load in the rivet shank - hole system is replaced by the transmission by friction between the lapped sheets. The situation described above applies to correctly make riveting with ordinary (solid) rivets, and, in particular, when rivet nuts are used with large clamping force exerted of the lapped surfaces.

In such cases, cracks occur on the sheets rubbing against each other (Fig. 1c), usually in several adjacent areas and do not propagate through the rivet hole [29, 30]. Maximum stresses accumulate in areas removed from the rivet holes, where secondary bending and fretting, acting jointly, initiate and grow the cracks.

Figure 2 also shows that the maximum fatigue stresses applied during the testing affect the crack initiation sites. For larger maximum stresses, the maximum secondary-bending torque is closer to the net section in the outlying rivet row due to limited effect of the rivet head (which restrains secondary bending), and, in addition, there are larger pressures of the rivet shank on the hole (overcoming friction forces between the joined parts), which, overall, leads to the initiation of cracks closer to the rivet hole than in the event of smaller maximum fatigue stresses.

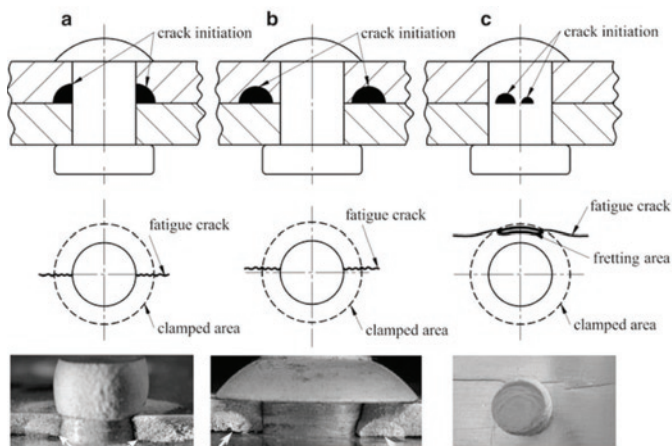


Fig. 1. Most common locations where cracks initiate in lap joints: a) on the edge of the rivet hole, b) away from the hole, c) above the rivet hole [35]



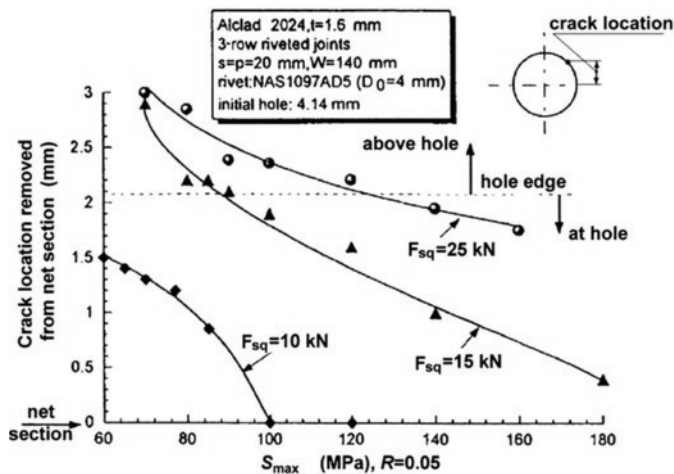


Fig. 2. Relationship between the location of cracks in lap joints and stresses for various rivet clamping forces [35]

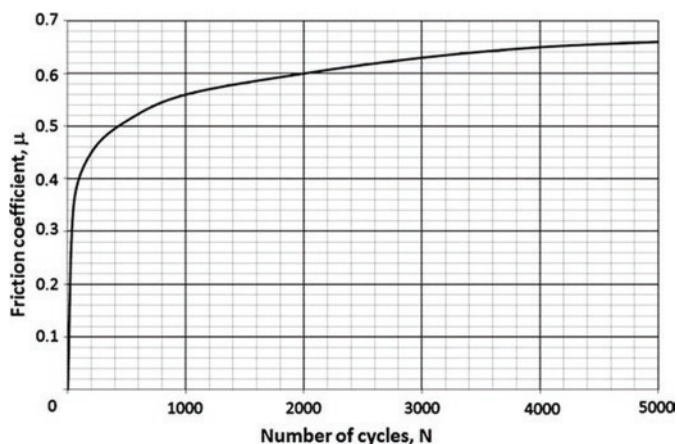


Fig. 3. Change of friction coefficient observed during fretting testing on 2024-T351 alloy [39]

Fatigue life of riveted joints can also be affected by various adhesive materials placed between the riveted parts during installation (sealers or adhesives). In such hybrid joints, cracks can initiate on sites different from those in the case of riveted joints without adhesive, and such cracks can propagate slower (as observed by Schütz [32] in testing conducted for lap joints with protruding-head rivets with and without sealers). Cracks in specimens without sealer were initiated on the edge of the rivet hole, whereas when a sealer is applied, the propagating cracks were away from the net section of the rivet row. In both cases, the external sheet of the joint was damaged, and fatigue life of the sealed joint was larger than in joints with no seal.

In analyzing fatigue life of riveted joints it is also worthwhile to note a major difference between isolated fatigue cracks and multiple cracks in riveted joints of aircraft skin. Fatigue cracks occurring in riveted joints of aircraft skin can be multiple-site damage (MSD) whose growth is determined by mutual action of the adjacent rivet holes, or single-site damage (SSD) which occurs as isolated defects mainly due to manufacturing and material defects. In terms of durability of riveted joints, multiple-site cracks are particularly hazardous for two reasons:

- Growth of MSDs from short to long is much quicker than in the case of single-site damage of the same size;
- Critical length of MSDs with the required residual strength is much less than in the case of single-site damage.

These conclusions have been reaffirmed by a number of experimental studies, for instance conducted by Schra [31] and many others [12, 37].

The purpose of the tests the results of which are presented in this paper was to find effective ways to improve fatigue life of riveted joints made in construction and repair shops in helicopters operated in the Polish Air Force for several dozen years. An improvement in fatigue life of riveted joints was achieved in two ways: one was to replace ordinary mushroom head rivets ref. 3558A-4-10 made of W65 alloy (Russian marking of aluminum alloy), 4 mm in diameter, with titanium driven blind bolts with pin, ref. MBF2110AB-05-150, with round head, diameter 4.2 mm), and the other was to use hybrid (adhesive & riveted) joints instead of riveted joints.

### 3. Fasteners and specimens used in experimental studies

The studies used two types of fasteners: ordinary mushroom head rivets 3558A-4-10 and driven blind bolts with pin MBF2110AB-05-150 with round head. Both rivet types have mushroom heads and require the same diameter of the hole. Ordinary rivets during the clamping expand and fill out the installation hole, but following the swaging of such rivet types, the clamp-up pressure applied to the joined elements is small. The situation is different for titanium rivets MBF2110AB-05-150 which, when clamped in place, still leave about 50 μm clearance between the shank and the hole edge and fill the hole incompletely following installation, but provide for a very large clamp-up pressure acting on the surfaces to be joined. The structure and principle of operation of such rivet types is similar to Hi-lok and Lockbolt rivets, which also ensure a significant clamp-up pressure applied to the joined surfaces. The magnitude of that pressure depends obviously on the diameter of the rivet. For instance, based on experimental studies presented in [33], the tension in the rivet shank with which the joint is clamped down has been found to be 50%-80% of the (static) tensile strength of the rivet along its shank. Hence, the Hi-Lok rivets tested in study [33] had their clamp-up pressure applied to the joint at about 5 kN, and Lockbolt rivets – about 6 kN. Unlike such fasteners, ordinary rivets are characterized by very small clamp-up pressure of the joint, i.e., within the range of 0-0.5 kN. In addition, the tested joints with rivet nuts contained significant amounts of oxides, being the products of wear (micro-friction) at crack initiation sites, which means that the crack was initiated by fretting.

The results of fatigue life testing of joints made with titanium Hi-Lok rivets have been very interesting, as their fatigue life was almost 2.5 times as much as of joints with steel Hi-Lok fasteners [33] (they differed by clamp-up pressure of the joint). In turn, joints with titanium Lockbolt rivets featured much poorer fatigue life than titanium Hi-Lok rivets despite the same configuration of the joint and rivet installation parameters. This was due to a difference in diameter between Hi-Lok rivet heads, which are about 10% larger than the diameter of Lockbolt rivet heads, which in turn affects the intensity of secondary bending in e.g. lap joints [33].

Experimental studies in this paper have been conducted on three types of specimens which differ by geometry. In addition, for each type of specimen, joints were made using ordinary mushroom head rivets 3558A-4-10 (Variant I) and driven blind bolts with pin MBF2110AB-05-150 (Variant II). The following types of specimens were used in the tests:

- single-lap joint made of two plated duralumin sheets AW 2024T3 2 mm thick and sized 100x25 mm (long x wide) fastened with either two ordinary rivets 3558A-4-10 (Variant I) or two titanium blind rivets MBF2110AB-05-150 with round head (Variant II) – Fig. 4;
- single-lap joint made of two non-plated duralumin sheets AW 2024T3 2 mm thick and sized 110x50 mm (long x wide) fastened with either six ordinary rivets 3558A-4-10 (Variant I) or six titanium rivets MBF2110AB-05-150 with round head (Variant II) – Fig. 5;
- strapped joint – repair of damaged part made of AW 2024T3 alloy 2 mm thick and sized 210x80 mm (long x wide) made by a strap

with an insert, damage in the form of a hole with diameter of 20 mm (Fig. 7), repair in the form of a riveted strip 70 mm in diameter with insert 20 mm in diameter (a round insert connected with a rivet with a patch is used in repairing airframe covers in order to reduce hole deformation; the patch and insert made of sheet AW2024T3); part were fastened with ordinary rivets 3558A-4-10 (Variant I) or titanium rivets MBF2110AB-05-150 (Variant II) – Fig. 6.

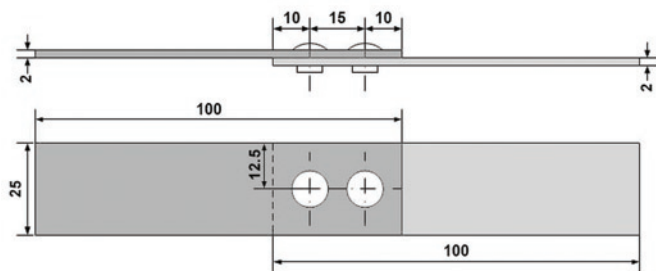


Fig. 4. Diagram of a single-lap specimen with two fasteners

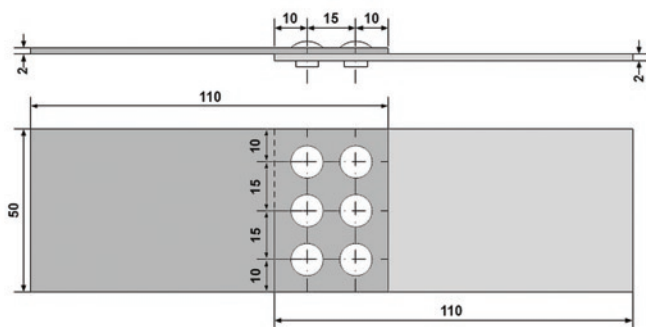


Fig. 5. Diagram of a single-lap specimen with six fasteners

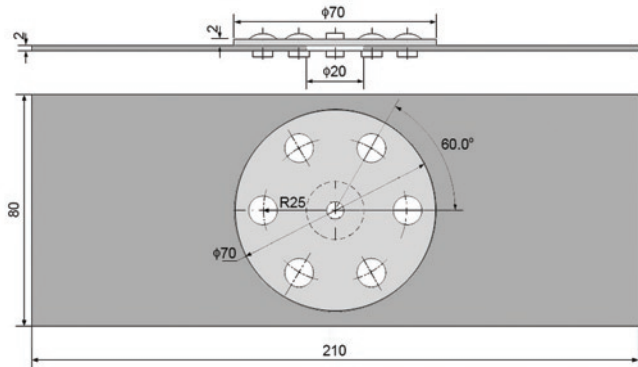


Fig. 6. Diagram of strapped-joint specimen – repair of damage

Hybrid joints were prepared based on strapped-joint specimens, using two types of adhesive material: Epidian 57/Z1 Sarzyna by Ciech S.A. (Poland) and Raychem S1125 by Cheney Manor Industrial Estate (UK). The materials used here differed significantly in terms of flexibility. Epidian 57/Z1 is a rigid material, while Raychem S1125 is a flexible material with high viscoelasticity. Stress-strain curves for these materials are presented in Fig. 7.

The surfaces to be adhesively bonded were cleaned with a 3M 3809 Fine one-sided abrasive sponge, washed with extraction gasoline and dried before applying the adhesive. Hybrid joints were made by installing fasteners prior to curing. The researchers expected that the adhesive bond with titanium rivets MBF2110AB-05-150 might be

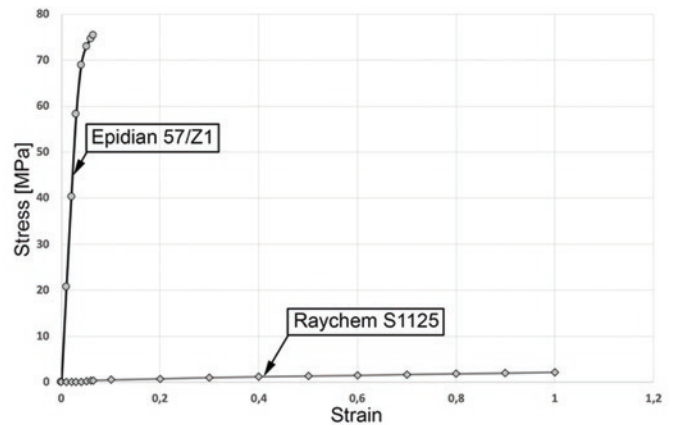


Fig. 7. Stress-strain curves for adhesive materials (speed 2mm/min)

weaker due to its high clamping force which would squeeze out the adhesive from the space between two parts to be joined. Therefore, specimens were also prepared in which a single layer of glass fabric Synglass E81 with basis weight of 101 g/m<sup>2</sup> was inserted between the parts to be joined. The bonds made with Epidian 57/Z1 were cured in two steps, i.e., for 12 hours at room temperature (about 20°C) and for 6 hours at 80°C, and Raychem bonds were cured for seven days at room temperature. The strength and durability testing was conducted on Instron 8802 machine.

#### 4. Lap joints – results of experimental studies

Lap joints have been tested under static loads and fatigue life ranges. The following results were obtained for static tests:

- load capacity of single lap specimens fastened with two ordinary rivets 3558A-4-10 (Variant I) was 7500±200 N, and with two rivets MBF2110AB-05-150 (Variant II) it was 10,500±150 N;
- load capacity of single lap specimens fastened with six ordinary rivets 3558A-4-10 was 23,000±180 N, and with six rivets MBF2110AB-05-150 it was 31,200±150 N.

The load capacity was proportional to the quantity of fasteners used in the joint. Hence, for joints made with titanium rivets MBF2110AB-05-150 it was about 40% larger than for joints with 3558A-4-10 rivets, which was also due to the materials of which the fasteners are made (rivets 3558A-4-10 – aluminum alloy, rivets MBF2110AB-05-150 – titanium alloy and steel pin). Joints made with 3558A-4-10 rivets were damaged by shearing, and the joints with MBF2110AB-05-150 fasteners were damaged by breaking fastener heads. As a result of surface stresses occurring in joints with titanium bolts, significant ovalization of installation holes was noted.

In preparation for fatigue life testing of the above-mentioned specimens, also the boundary load values of the fatigue cycles needed to be determined. Maximum loads were selected by applying the rule that nominal stresses in sections with or without holes should be about 230 MPa, which is the upper yield strength of duralumin AW 2024T3 divided by safety factor used in aviation, i.e., n=1.5. The value of the safety factor comes from the provisions of JAR 25.303. Guided by such rule, single-lap specimens joined together with two fasteners were fatigue-loaded between the minimum force  $F_{min} = 3500$  N and the maximum force  $F_{max} = 5000$  N in sinusoidal cycles with 20 Hz frequency. The results of fatigue tests for five specimens with rivets

Table 1. Number of cycles to damage of a single-lap joint fastened with two ordinary rivets 3558A-4-10

	Specimen 1	Specimen 2	Specimen 3	Specimen 4	Specimen 5
Number of cycles to joint damage	551,746	714,441	464,999	602,610	626,371
Average number of cycles to joint damage	592,033				

Table 2. Number of cycles to damage of a single-lap joint fastened with two titanium rivets MBF2110AB-05-150

	Specimen 1	Specimen 2	Specimen 3	Specimen 4	Specimen 5
Number of cycles to joint damage	1,010,888	1,063,639	1,100,000 test broken	1,005,000 test broken	1,005,000 test broken
Average number of cycles to joint damage	1,036,905				

3558A-4-10 are specified in Table 1, and with MBF2110AB-05-150 fasteners in Table 2.

For three specimens of joints with titanium fasteners MBF2110AB-05-150, testing was interrupted after about 1,000,000 cycles, because qualitative confirmation has been obtained that confirms higher fatigue life of joints with titanium rivets. A view of the joints damaged during the fatigue test for two types of fasteners is presented in Fig. 8.

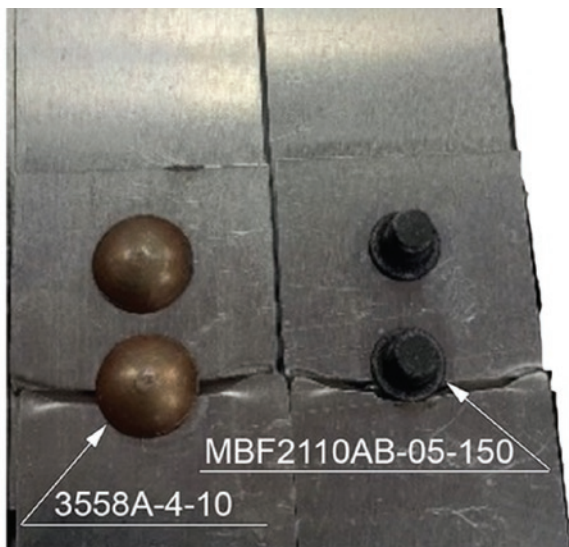


Fig. 8. View of lap joint damage for two types of fasteners (3558A-4-10 and MBF2110AB-05-150)

In addition, one of the specimens of the lap joint with two titanium fasteners which was not damaged, was studied with computer-assisted tomograph v/tome/x m, by GE, equipped with panel detector, and projection system using a cone-shaped x-ray beam (Fig. 9). An image of the joint shows that the hole is not filled out by the shank completely, which is characteristic of such fasteners.

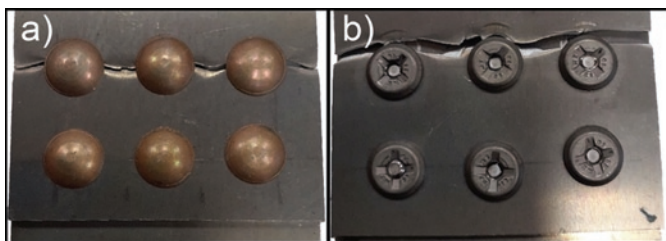


Fig. 9. CT image of a joint made with titanium blind bolts following a load of 1,100,000 cycles

On the next stage of durability studies, single-lap specimens joined together with six ordinary rivets 3558A-4-10 or titanium rivets MBF2110AB-05-150 were tested. The joints were loaded with sinusoidal fatigue cycle in the range between  $F_{max}=15,000$  N and  $F_{min}=10,500$  N with frequency of 20 Hz, until the sheets were damaged. The specimen with fasteners 3558A-4-10 was damaged following 207,952

cycles, whereas the specimen with fasteners MBF2110AB-05-150 was damaged following 1,140,700 cycles. The nature of the joint damage also depended on the type of fastener, i.e., for 3558A-4-10 fasteners the fracture was along the line of holes under rivet heads (Fig. 10a), and for MBF2110AB-05-150 fasteners the fracture was along the edge of heads, away from the critical section (Fig. 10b).

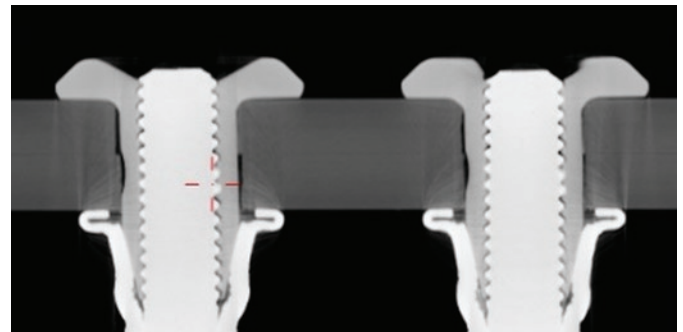


Fig. 10. Fatigue crack of sheets fastened with six rivets: (a) 3558A-4-10, (b) MBF2110AB-05-150

The durability tests demonstrated better fatigue life of joints made with titanium rivets MBF2110AB-05-150 and a different nature of damage – the crack did not follow through the weakest section but began on the edge of a head. Consistently with the current state of knowledge and the results of studies presented in global literature [13,14,24,29,30,35], cracks in joints with MBF2110AB-05-150 rivets initiated away from the critical section (as well as away from the clamped area) due to high clamp-up pressures applied to the sheets by the fasteners. For 3558A-4-10 rivets, the rivet clamping force is relatively small, due to which the load is transferred mainly by rivet shanks pressing against the surface of installation holes. In such case, cracks are initiated usually on hole edges.

## 5. Strapped joints (repair of damage) – numerical calculations

In strapped specimens, a problem might occur of multiple cracks due to the interference between the installation holes and the adjacent central hole having 20 mm in diameter. That is why numerical calculations were conducted on the initial stage of the analysis to assess this condition.

The calculations were conducted using the Static Structural module of ANSYS v.19.2. software. A damaged specimen was modelled (a hole 20 mm in diameter) and a specimen repaired by riveting down a metal strap (hole 20 mm in diameter and installation holes). The following boundary conditions were adopted: force of  $F=30,000$  N is applied to the lateral surface as presented in Fig. 11 (the value of the force results from the adopted criterion of nominal stresses in a section with a hole at approx. 230 MPa), and the opposite lateral surface is fixed, also as shown in Fig. 11. The FE model was built based on hexagon elements (using a “Hex Dominant Method”) using tools for grid thickening in the region of expected large stress gradients. Due to the fact that the results of numerical analyses were meant to be qualitative rather than quantitative ones, the FE model was built using automatic algorithms offered by the ANSYS software.

The material of the AW 2024T3 sheet was modelled as elastic and plastic material with strain hardening (yield strength  $R_e=330$  MPa, strain hardening modulus of 1000 MPa), because in calculations for the linear model of the material, stresses significantly exceeding the strength of tested material were found. Distribution of Huber-Mises reduced stresses and maximum principal stress, which characterize better the strain of the material in fatigue tests. First, calculations

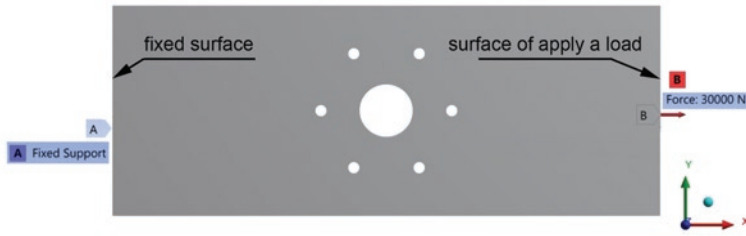


Fig. 11. Boundary conditions of the model under analysis

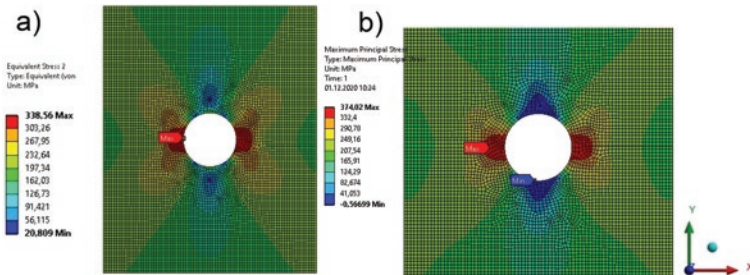


Fig. 12. Distribution of stress in a part with 20 mm hole loaded with 30,000 N: (a) Huber-Mises stresses, (b) maximum principal stress

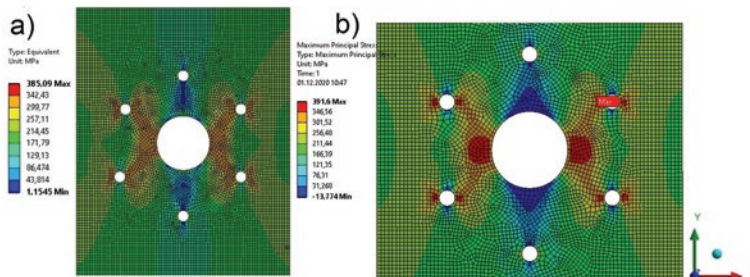


Fig. 13. Distribution of stress in a piece with 20 mm hole and installation holes 4.2 mm in diameter loaded with 30,000 N: (a) Huber-Mises stresses, (b) maximum principal stress

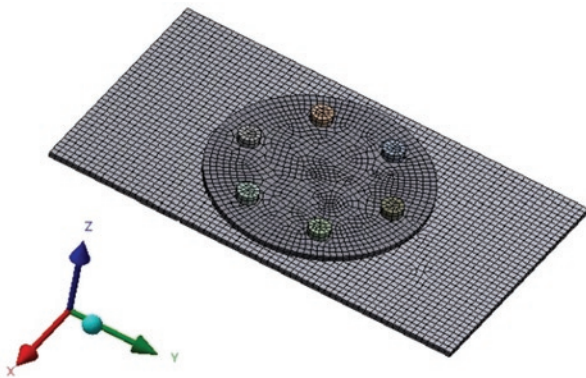


Fig. 14. Numerical model of strapped joint of a specimen with riveted patch

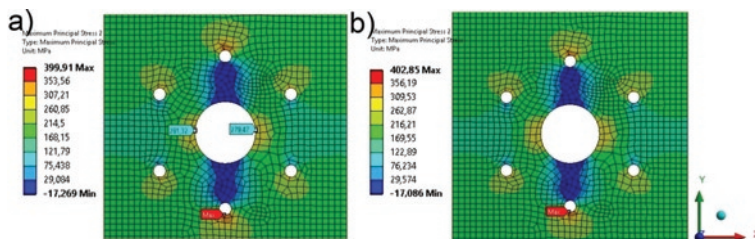


Fig. 15. Distribution of maximum principal stresses in a piece repaired with: (a) titanium rivets, (b) duralumin rivets

were conducted for the distribution of reduced stresses by Huber/Mises hypothesis (Fig. 12a), maximum principal stress (Fig. 12b), in a piece with a single hole 20 mm in diameter, loaded with 30,000 N. For the part with hole  $d = 20$  mm with the load of 30 kN it was found that the yield strength of the material AW 2024T3 in its critical section was exceeded.

In the next step, calculations were conducted for the distribution of reduced stresses by Huber/Mises hypothesis (Fig. 13a), maximum principal stress (Fig. 13b), in a piece with a central hole 20 mm in diameter and six riveting holes, loaded with 30,000 N.

In the piece with installation holes  $d=4.2$  mm, maximum stresses at load of 30,000 N are by about 50 MPa more than in the piece with a single central hole only. Consequently, the specimen with the 20 mm hole should feature, among other things, better fatigue life than the specimen with rivet holes. Upon obtaining the results for the model of a plate with a central hole and rivet holes, model specimens were developed with a round strap of 70 mm in diameter mounted on the damaged part using duralumin rivets and titanium rivets (Fig. 14). Models of joints with titanium and duralumin rivets were developed with the assumption that the damaged part and the strap are made of aluminum alloy modelled as bilinear material with yield strength of  $\sigma_y=330$  MPa and strain hardening modulus  $D=1000$  MPa; linear properties of the rivet material (aluminum and titanium, respectively) were adopted. Boundary conditions for load and strain hardening were adopted identical to the previous model. In addition, friction factor at  $\mu=0.1$  was assumed between rivet heads and joined elements and between the strap and the damaged part. The rivet shanks were also assumed to fill in the rivet holes without friction and fit the sheet material subject to deformation ("no separation" function built into Ansys was used). The initial analyses of rivet joints covered cases meant to be only a point of reference for further analysis, and namely, they assumed that both duralumin and titanium rivets do not clamp together the sheets to be joined, and the entire load is transferred by friction force and by shearing of rivet shanks. Distribution of stresses in the damaged part using various fasteners are presented in Fig. 15.

It was found that if clamping forces in the parts to be joined produced by the installed rivets and the bucking of the shanks of duralumin rivets are not taken into account, the most loaded during the transfer of joint loads are the external rivet holes (Fig. 15), which is not reaffirmed by experimental studies where duralumin and titanium rivets are clamped together with an appropriate force.

Finally, specimen model was developed with a strap installed on the damaged part using duralumin rivets and titanium rivets, taking into account clamp-up pressures in the parts to be joined, generated by the installation of rivets as well as stresses occurring in the hole due to the bucking of duralumin rivets. Models of specimens with titanium and duralumin rivets were developed with the assumption that the damaged part and the strap are made of aluminum alloy modelled as bilinear material with yield strength of  $\sigma_y=330$  MPa, with linear properties of the rivet material (aluminum and titanium, respectively) defined. In addition, friction factor at  $\mu=0.1$  was assumed between rivet heads and the element and between the strap and the damaged part. Rivet shanks were also assumed to fill in the rivet holes without friction and fit the sheet material subject to deformation ("adjust to touch" function built into Ansys was used). Duralumin rivets were also assumed to clamp down the element with 500 N, and expanding rivet shanks press down on the internal surface of the holes, defined as pressure of 150 MPa. Titanium rivets clamp down the sheets with 5 kN and fill out the rivet holes without pressing down on the surface (due to no expansion inside the hole). Boundary conditions for load and strain hardening were adopted identical to the previous model. The clamping down of

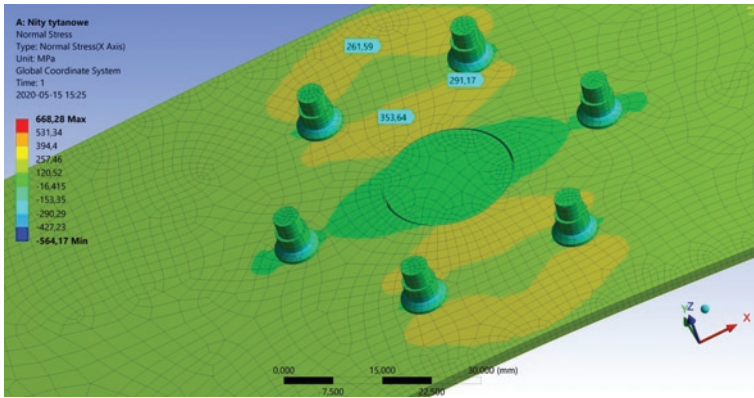


Fig. 16. Model of a specimen with titanium rivets and distribution of normal stresses along a piece straining load

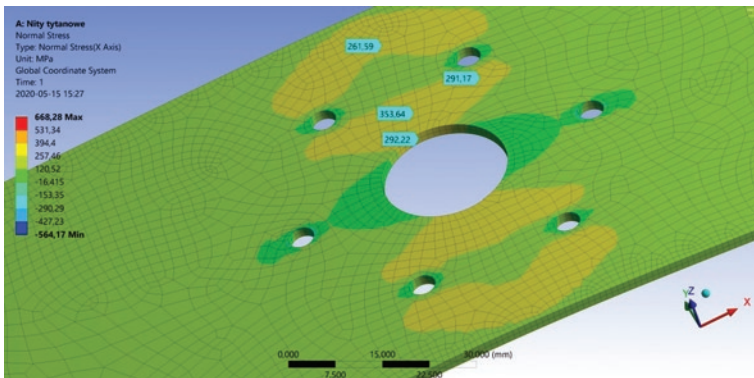


Fig. 17. Model of a specimen with hidden titanium rivets to facilitate the imaging of the distribution of normal stresses in rivet hole areas

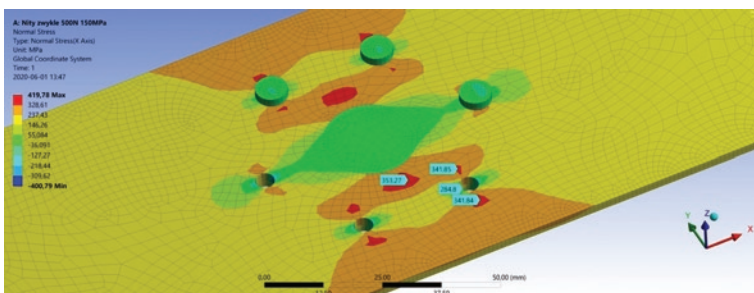


Fig. 18. Distribution of stresses in an element being repaired using expanded duralumin rivets

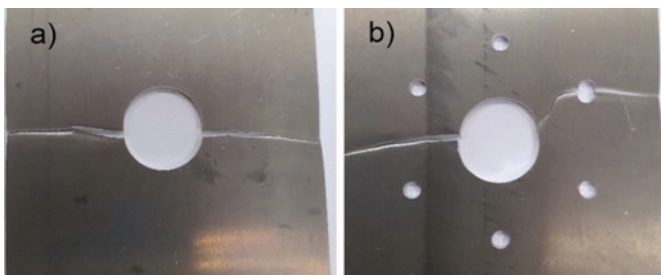


Fig. 19. Fatigue damage of a WA 2024T3 alloy part 2 mm thick and with the size of 210x80 mm: (a) with a hole 20 mm in diameter; (b) with a hole 20 mm in diameter and installation holes with the diameter of 4.2 mm

the sheets with rivets was modelled in ANSYS using “Bolt pretension” tool which enables one to set the force with which the factory and shop heads act on the surfaces to be joined.

In analyzing the results of calculations for titanium rivets (Fig. 16 and 17), it was found that the use of a strap locked in place with titanium rivets, each of which presses the patch down with 5kN, removes

maximum stresses away from the edge of the  $d = 20$  mm hole, which should make it harder to initiate a crack. Above all, the application of such fasteners reduces the stresses on rivet holes, which should improve fatigue life of the joint considered here.

As mentioned when modelling a specimen with duralumin rivets, rivet clamp-up pressure of 500 N was assumed, and pressure inside of the rivet holes with 150 MPa was modelled, representing pressures from the expanding rivet shank. For such case, removal of maximum stresses away from the edge of the central hole was achieved, whereas significant concentration of stresses in the area of installation holes was found which can result in nucleation of fatigue cracks (Fig. 18). Based on this, it can be concluded that the relatively small force with which duralumin rivets clamp down the sheets is insufficient to remove the area of stress concentration away from the rivet holes (as is the case for titanium rivets), which may result in lower fatigue life of that joint.

## 6. Experimental tests of fatigue life of rivet and hybrid joints

### 5.1. Rivet joints

Fatigue tests of strapped joints (Fig. 6) were conducted with sinusoidal loads of  $F_{\min} = 20,000$  N,  $F_{\max} = 30,000$  N and frequency of 8 Hz. At maximum load, nominal stresses in a section which is weakened by the 20 mm hole were 250 MPa, which is below the yield strength of AW 2024T3 duralumin (which is about 330 MPa).

Fatigue life of the piece made of WA 2024T3 alloy 2 mm thick and size of 210x80 mm with a hole of 20 mm was 407,800 cycles, whereas of a part with additional rivet holes of  $d = 4.2$  mm was 211,200. Fatigue cracks propagated through the central hole of  $d = 20$  mm and through one of the rivet holes  $d = 4.2$  mm. Obviously, the development of the crack results from the concentration of loads around the holes and coincides with the critical section of the specimen. The absence of crack symmetry can be explained by early initiation of the crack in the rivet hole and its subsequent expansion due to non-symmetric load. The appearance of the damage to the parts being tested is shown in Fig. 19.

Fatigue life of the specimen with rivet holes proved to be smaller than for the specimen with hole of  $d = 20$  mm alone, which – as demonstrated by numerical calculations – results from a larger concentration of stresses at a hole with  $d = 4.2$  mm.

Then, fatigue life of strapped joints was tested (repair of a damaged part made of AW 2024T3 alloy). The specimen of a joint made with ordinary rivets 3558A-4-10 was fatigue-damaged following 301,600 cycles (Fig. 20 a), and the damage differed from the damage made with static strain test (Fig. 20 b), and was similar to the damage of the part with central hole and installation holes (Fig. 19 b). The damage occurring in the static strain test runs through the critical section of the specimen, i.e., the central hole and two rivet holes. For the fatigue-stressed specimen ( $F_{\max} = 30,000$  N), the mechanism of damage changes from local disruption of the specimen at the stress concentration site into the initiation of crack in one of the holes, development of the crack and complete breakup with non-symmetric load.

Also the other specimen repaired with ordinary rivets 3558A-4-10 was tested, and fatigue life of that specimen was 567,200 cycles, and the cracks propagated from two rivet holes. In this case, the cracks initiated symmetrically in two rivet holes.

Also strapped joints were tested, in which titanium rivets MB-F2110AB-05-150 were used for installation. The specimen with titanium rivets was fatigue-damaged following 957,200 cycles – Fig. 21 a. In this case, the crack did not run through the hole with  $d = 4.2$  mm. Clamp-up pressures which are many times larger for titanium rivets result in a larger part of the load being transferred by sheet friction

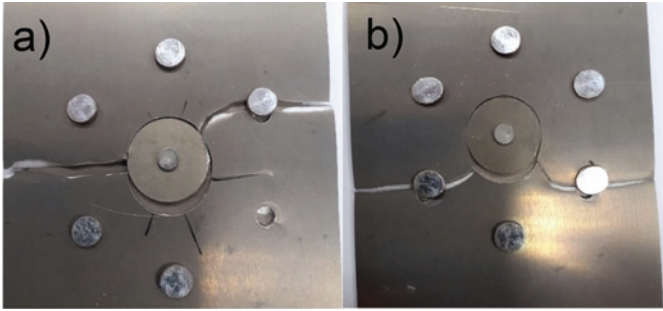


Fig. 20. Destruction of the strapped specimen fastened with ordinary rivets 3558A-4-10: (a) following fatigue test, (b) following static strain test

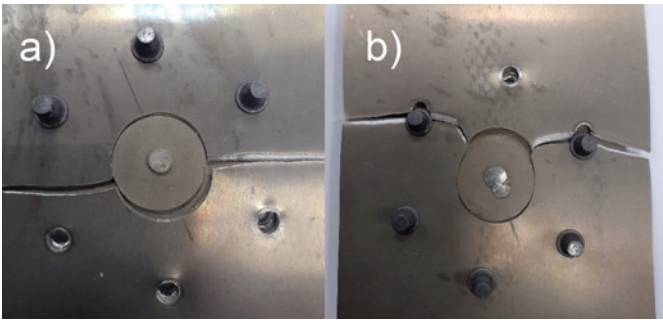


Fig. 21. Fracture of the strapped specimen riveted with fasteners: (a) following fatigue test, (b) following static strain test

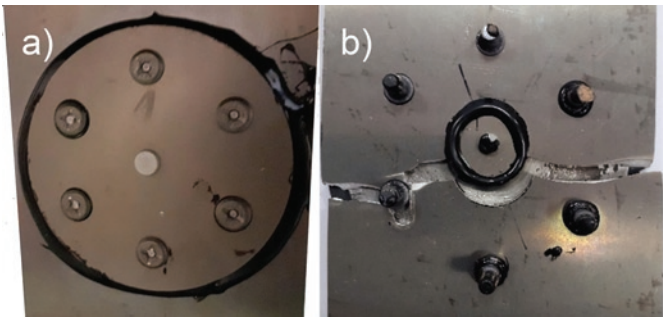


Fig. 22. View of a hybrid strapped joint (MBF2110AB-05-150 rivets and Raychem S1125 compound): (a) after preparation of specimen, (b) damaged in durability tests

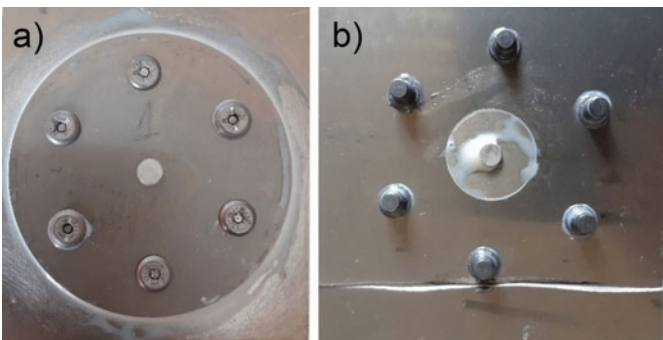


Fig. 23. View of a hybrid strapped joint (MBF2110AB-05-150 rivets and Epidian 57/Z1 compound): (a) after preparation of specimen, (b) damaged in durability tests

forces at the joint, and crack initiation sites are removed away from rivet hole critical sections. This makes the section along the center line of the 20 mm hole becoming the critical section, including due to maximum stresses occurring at this site, resulting from secondary bending (due to geometry of the strapped joint). Like for repairs made with ordinary rivets, the damage occurring in the static strain test (Fig.

21 b) runs through the critical section of the specimen, i.e., the central hole and two rivet holes.

## 5.2. Hybrid joints

In seeking ways to enhance fatigue life of strapped joints, fatigue life tests were conducted also for a hybrid joint which uses, in addition to titanium rivets MBF2110AB-05-150, adhesive material Raychem S1125 to install the strap. Such prepared specimen (Fig. 22 a) was also loaded with the following force range:  $F_{max} = 30,000$  N,  $F_{min} = 20,000$  N in a sinusoidal cycle with 8 Hz frequency. Following 3,059,100 load cycles and no specimen damage, the load was increased up to the range of  $F_{max} = 35,000$  N,  $F_{min} = 25,000$  N. Following further 328,200 cycles, the joint was damaged (Fig. 22 b), and the cracking was initiated on the hole with  $d = 20$  mm (the fracture at the rivet occurred in the last phase of destruction). The improved fatigue life is due to additional adhesive forces bonding the surfaces, which exceed friction forces usually occurring in joints of such type between the parts joined together. The initiation of the crack on the central hole of the specimen was probably due to maximum stresses occurring at the site from secondary bending (due to the geometry of the strapped joint).

Hybrid joints were also tested by replacing the Raychem S1125 adhesive compound (with strong viscoelastic properties) with Epidian 57/Z1 structural adhesive which is much stronger than the Raychem compound (Fig. 23 a). The hybrid joint specimen was loaded with forces within the following range:  $F_{max} = 30,000$  N,  $F_{min} = 20,000$  N in a sinusoidal cycle with 8 Hz frequency. The joint was damaged following 1,506,300 load cycles. The crack initiated on the edge of the shop head along the specimen loading axis, and propagated first towards one edge of the specimen, and then in both directions (Fig. 23 b).

Inspection of the damaged specimen found that the pressures occurring during the installation of the rivets squeeze the adhesive compound out of the space between the joined parts. Any imperfection in the adhesive bond, especially around the installation holes, may affect fatigue life of the joint. Hence, more specimens of hybrid (bonded and riveted) joints were prepared, this time using ordinary rivets 3558A-4-10. Two variants of the strapped joint with Epidian 57 adhesive were prepared. They differed by the presence of a single layer of Synglass E81 glass fabric with basis weight  $101 \text{ g/m}^2$  in the adhesive layer. Such prepared specimens were loaded with forces within the following range:  $F_{max} = 30,000$  N,  $F_{min} = 20,000$  N in a sinusoidal cycle with 8 Hz frequency. The hybrid joint without glass fabric was damaged following 751,100 load cycles, and with the glass fabric following 1,017,400 load cycles. In the joint with no fabric, the crack propagated from the hole with  $d = 20$  mm (Fig. 24 a), and in the joint with fabric – it propagated from the outlying fastener (Fig. 24 b). In addition, durability was determined for a specimen joined with ordinary rivets 3558A-4-10 and bonded with Raychem S1125 adhesive. The joint was damaged following 950,597 load cycles.

Similar testing was conducted for a hybrid joint using titanium rivets and Raychem S1125 compound, but with modifying the bond by adding Synglass E81 glass fabric. In expectation for improved fatigue performance, such prepared specimen was loaded with increased forces within the range of:  $F_{max} = 35,000$  N,  $F_{min} = 25,000$  N in a sinusoidal cycle with 8 Hz frequency. The sample was damaged following 3,114,700 load cycles. As shown in Fig. 25, the crack propagated from hole  $d = 20$  mm.

Fatigue life of the strapped joints tested is compared in Table 3.

Testing results summarized in Table 3 and numerical calculations show that the smaller rivet holes result in more concentration of stresses than the large central hole with the diameter of  $d = 20$  mm. Hence, it turns out that, in terms of fatigue life, repair of pieces with a central hole and a strap fastened with ordinary rivets 3558A-4-10 (strapped joints) is less effective than no repair at all (a piece with central hole with the diameter of  $d = 20$  mm had larger fatigue life than a piece repaired with ordinary rivets). Blind rivets MBF2110AB-05-150 were an effective

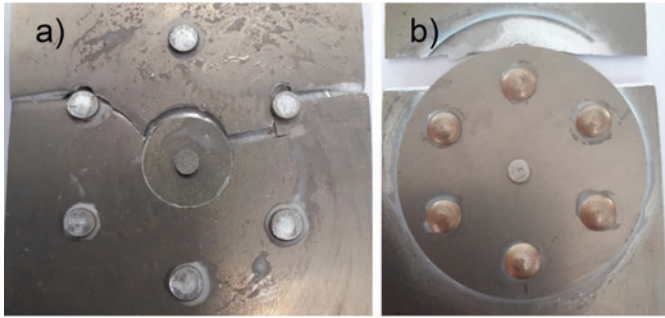


Fig. 24. Fatigue damage of a hybrid joint (3558A-4-10 rivets plus Epidian 57/Z1 compound): (a) with no glass fabric, (b) with glass fabric



Fig. 25. Fatigue damage of a hybrid strapped joint (MBF2110AB-05-150 rivets and Raychem S1125 compound with a layer of Synglass E81 fabric)

solution. Fatigue life of the repair joint with titanium rivets was twice as much as for a piece with a central hole.

Experimental tests have clearly demonstrated that titanium blind rivets perform better than ordinary rivets in terms of ensuring fatigue life of joints of pieces made of AW 2024T3 aluminum alloy.

Also, the use of hybrid (adhesive & riveted) joints has been found to be viable in strapped joints, e.g., in repairs of damaged skin of aircraft. Better fatigue life was achieved by using a less rigid adhesive compound (lower Young's modulus), i.e., Raychem S1125 adhesive. The rigid adhesive bond achieved with Epidian 57/Z1 resulted in poorer fatigue life.

A major problem with making hybrid joints before the adhesive compound is cured is the quality of the adhesive bonds so made. Pressures exerted by fasteners during installation squeeze out the adhesive compound out in the area of the joint, and the effective area of the adhesive joint is less than the geometrical size of a bonded strap (Fig. 26a). Not insignificant in terms of fatigue life of hybrid joints is the absence of the adhesive material around installation holes where stresses concentrate. The presence of adhesive in such sites would significantly reduce stress concentration factors. As demonstrated by the outcomes of experimental tests, a solution to this problem may be to add a filler layer to the adhesive material to reduce the adverse effect of the adhesive being squeezed out during the installation of fasteners. These researchers have successfully applied a layer of glass fabric whose presence in the joint does not adversely affect the adhesive strength of the bond (Fig. 26 b,c).

## 7. Conclusions

Based on experimental studies and numerical calculations we have found that fatigue life of joints in workpieces made of aluminum alloy AW 2024T3 can be improved by using titanium driven blind bolts with pin MBF2110AB-05-150 instead of ordinary, hammer-bucked rivets 3558A-4-10. Another beneficial solution to improve the durability of riveted joints is to replace them with hybrid (rivet & adhesive) joints.

The use of MBF2110AB-05-150 rivets instead of 3558A-4-10 in helicopter repairs should significantly reduce the problem of secondary fatigue damage to the repaired airframe structures.

Table 3. Comparison of fatigue life of workpieces with holes and strapped joints

No.	Specimen type	Cycles to damage
1	Piece with hole $d = 20$ mm	407,800
2	Piece with hole $d = 20$ mm and six installation holes $d_1=4.2$ mm	211,200
3	Strapped joint – 3558A-4-10 rivets	301,600
4	Hybrid strapped joint – 3558A-4-10 rivets and Epidian 57/Z1 adhesive compound	751,100
5	Modified hybrid strapped joint – 3558A-4-10 rivets, Epidian 57/Z1 with glass fabric	1,017,400
6	Strapped joint – 3558A-4-10 rivets, Raychem S1125 adhesive compound	950,597
7	Strapped joint – MBF2110AB-05-150 titanium rivets	957,200
8	Hybrid strapped joint – MBF2110AB-05-150 titanium rivets and Epidian 57/Z1 adhesive compound	1,506,300
9	Modified hybrid strapped joint – MBF2110AB-05-150 rivets, Epidian 57/Z1 with glass fabric	> 3,015,300 (3,015,300 cycles with load of 30 – 20 kN and additional 1,947,000 cycles with load of 35 - 25 kN)
10	Hybrid strapped joint – MBF2110AB-05-150 rivets, Raychem S1125 compound	> 3,059,100 (3,059,100 cycles with load of 30 – 20 kN and additional 328,200 cycles with load of 35 - 25 kN)
11	Modified hybrid strapped joint – MBF2110AB-05-150 rivets, Raychem S1125 with glass fabric	3,114,700 (with load increased to 35 – 25 kN)

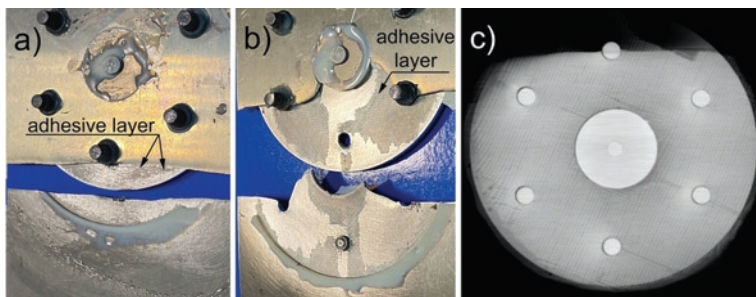


Fig. 26. View of adhesive bond: (a) in hybrid strapped joint without glass fabric, (b) in hybrid strapped joint with a layer of glass fabric, (c) following the analysis with computer-assisted tomograph in the hybrid strapped joint with a layer of glass fabric

Our studies have confirmed that the fasteners compared during the testing transfer loads differently: by friction of strongly pressed down factory and shop heads of MBF2110AB-05-150 and shearing of bucked rivets 3558A-4-10. Numerical calculations have found that both the pressing down joined elements by heads with appropriate force and pressure of shanks of deformed rivets on hole walls remove

the largest stresses away from hole edges. This effect is much stronger for titanium rivets than for duralumin rivets, which results in longer delay of crack initiation in joints with titanium rivets, thus improving fatigue life of such joints relative to joints using ordinary duralumin rivets.

Experimental tests have also demonstrated the viability of replacing riveted joints with hybrid (riveted & adhesive) joints in repairs of airframe structures. In addition to other useful properties of such joints, e.g., their airtightness, fatigue life of the repaired site can be significantly improved. The magnitude of increased fatigue life depends on the type of the adhesive material used in the joint. Adhesive compounds with moderate rigidity provide for better effects, because then fasteners are loaded more evenly and are better strained. Hence damage of the adhesive bond (which is damaged first) of a hybrid joint will occur later and its fatigue life will be better compared to “pure” riveted joints. At the same time, one should remember that as a result of pressures occurring during the installation of fasteners, the adhesive joint is degraded. This adverse outcome can be significantly reduced by physically modifying the adhesive bond with fillers, e.g., thin fabrics made of glass fibers.

## References

1. Abdel Wahab M M. Fatigue in adhesively bonded joints: a review. *ISRN Materials Science* 2012; 2012: 1-25, <https://doi.org/10.5402/2012/746308>.
2. Alderliesten R C. Introduction to aerospace structure and materials, Delft University of Technology, 2018.
3. Banea M D, da Silva L F M, Campilho R D S G, Sat, C. Smart adhesive joints: an overview of recent developments. *The Journal of Adhesion* 2014; 90: 16-40, <https://doi.org/10.1080/00218464.2013.785916>.
4. Budhe S, Banea M, D, de Barros S, da Silva L. An updated review of adhesively bonded joints in composite materials, *International Journal of Adhesion & Adhesives* 2017; 72: 30-42, <https://doi.org/10.1016/j.ijadhadh.2016.10.010>.
5. Chang P, Wang J, Wing Kong Chiu, Nabil M. Chowdhury. Experimental and finite elements studies of bolted, bonded and hybrid step lap joints of thick carbon fibre/epoxy panels used in aircraft structures, *Composites Part B* 2016; 100: 68-77, <https://doi.org/10.1016/j.compositesb.2016.06.061>.
6. Chaves F J P, da Silva L F M, de Moura M F S F, Dillard D A, Esteves V H C. Fracture mechanics tests in adhesively bonded joints: a literature review. *The Journal of Adhesion* 2014; 90: 955-92, <https://doi.org/10.1080/00218464.2013.859075>.
7. Costa M, Viana G, da Silva L F M, Campilho R D S G. Environmental effect on the fatigue degradation of adhesive joints: a review. *The Journal of Adhesion* 2016; 93: 1-2, 127-146, <https://doi.org/10.1080/00218464.2016.1179117>.
8. da Silva L F M, das Neves P J C, Adams R D, Wang A, Spelt J K. Analytical models of adhesively bonded joints-Part II: comparative study *International Journal of Adhesion and Adhesives* 2009; 29: 331-41, <https://doi.org/10.1016/j.ijadhadh.2008.06.007>.
9. da Silva, L F M, das Neves P J C, Adams R D, Spelt J K. Analytical models of adhesively bonded joints-Part I: literature survey. *International Journal of Adhesion and Adhesives* 2009; 29: 319-30, <https://doi.org/10.1016/j.ijadhadh.2008.06.005>.
10. Gąsior J, Komorek A, Rośkiewicz M, Tkaczuk S. Ocena możliwości zastąpienia nitów typu solid w połączeniach konstrukcji lotniczych (Assessment of the potential to replace solid rivets in aircraft structural joints). *Technologia i Automatykacja Montażu* 2018; 2/2018: 53-56.
11. Godzimirski J, Rośkiewicz M. Selection of joints for testing fatigue life of aviation rivets. *Technologia i Automatykacja Montażu* 2020; 2/2020: 17-20.
12. Gruber M L, Wilkins K E, Worden R E. Investigation of fuselage structure subject to widespread fatigue damage. In: Bigelow, C.A. (ed.) *Proceedings of FAA/NASA Symposium on the Continued Airworthiness of Aircraft Structures*, Atlanta, GA, 28-30 Aug 1996; DOT/FAA/AR-97/2: 439-459.
13. Harish G, Farris T N, Wang H L, Grandt A F. Nucleation and growth of cracks in lap joints. In: 1999 USAF Aircraft Structural Integrity Program Conference, 30 Nov-2 Dec 1999; San Antonio; TX: 1-14.
14. Hartman A. Fatigue tests on single lap joints in clad 2024-T3 aluminum alloy manufactured by a combination of riveting and adhesive bonding. Report NLR M.2170. NLR, Amsterdam, 1966.
15. Hartman A. Some tests on the effect of fatigue loading on the friction in riveted light alloy specimens. Report NLR M. 2008. NLR, Amsterdam, 1961.
16. He, X., A review of finite element analysis of adhesively bonded joints. *International Journal of Adhesion and Adhesives* 2011; 31: 248-64, <https://doi.org/10.1016/j.ijadhadh.2011.01.006>.
17. Heshmati M, Haghani R, Al-Emrani M. Environmental durability of adhesively bonded FRP/steel joints in civil engineering applications: state of the art. *Composites Part B: Engineering* 2015; 81: 259-75, <https://doi.org/10.1016/j.compositesb.2015.07.014>.
18. Jones R, Baker A, Matthews N, Champagne V. *Aircraft Sustainment and Repair*. Butterworth-Heinemann, 2018; <https://doi.org/10.1016/C2014-0-03919-6>.
19. Katnam K B, Comer A J, Roy D, da Silva L F M, Young T M. Composite repair in wind turbine blades: an overview. *The Journal of Adhesion* 2015; 91: 113-39, <https://doi.org/10.1080/00218464.2014.900449>.
20. Katnam K B, da Silva L F M, Young T M. Bonded repair of composite aircraft structures: a review of scientific challenges and opportunities. *Progress in Aerospace Sciences* 2013; 61: 26-42, <https://doi.org/10.1016/j.paerosci.2013.03.003>.
21. Klimaszewski S, Leski A, Dragan K, Kurdelski M, Wrona M. Helicopter Structural Integrity Program Of Polish Mi-24 Hind Helicopters.



- In: Bos M.J. (eds) Proceedings of the 25th Symposium of the International Committee on Aeronautical Fatigue, Bridging the Gap between Theory and Operational Practice. Springer, Dordrecht 2009, [https://doi.org/10.1007/978-90-481-2746-7\\_16](https://doi.org/10.1007/978-90-481-2746-7_16).
22. Komorek A, Przybyłek P. Examination of the influence of cross-impact load on bend strength properties of composite materials, used in aviation. *Eksplatacja i Niezawodność - Maintenance and Reliability* 2012; 14(4): 265-269.
  23. Marques E A S, da Silva L F M, Banea M D, Carbas R J C. Adhesive joints for low- and high-temperature use: an overview. *The Journal of Adhesion* 2015; 91: 556-85, <https://doi.org/10.1080/00218464.2014.943395>.
  24. Müller R P G. An experimental and analytical investigation on the fatigue behavior of fuselage riveted lap joints. The significance of the rivet squeeze force, and a comparison of 2024-T3 and Glare 3. Ph.D. thesis, TU Delft, Delft, 1995.
  25. Pitta S, de la Mora Carles V, Roure Fernández F, Crespo Artiaga D, Rojas Gregorio J I. On the static strength of aluminium and carbon fibre aircraft lap joint repairs, *Composite Structures* Volume 201, 1 October 2018: 276-290, <https://doi.org/10.1016/j.compstruct.2018.06.002>.
  26. Rośkowicz M, Smal T. Research on durability of composite materials used in repairing aircraft components. *Eksplatacja i Niezawodność - Maintenance and Reliability* 2013; 15 (4): 349-355.
  27. Rudawska A, Dębski H. Experimental and numerical analysis of adhesively bonded aluminium alloy sheets joints. *Eksplatacja i Niezawodność - Maintenance and Reliability* 2011; 1(49): 4-10.
  28. Sauer R A. A survey of computational models for adhesion. *The Journal of Adhesion* 2016;92:81-120, <https://doi.org/10.1080/00218464.2014.1003210>.
  29. Schijve J. Fatigue life until small cracks in aircraft structures. Durability and damage tolerance. In: Harris, Ch.E. (ed.) Proceedings of the FAA/NASA International Symposium on Advanced Structural Integrity Methods for Airframe Durability and Damage Tolerance, Hampton, VA, 4-6 May 1994, NASA CP 3274: 665-680.
  30. Schijve, J. Multiple-site-damage of riveted joints. In: Atluri S.N., Harris C.E., Hoggart A., Miller N., Sampath, S.N.: International Workshop on Structural Integrity of Ageing Airplanes, Durability of Metal Aircraft Structures, Atlanta, GA, 31 Mar-2 Apr 1992. Atlanta Technical Publication, Atlanta; 1992: 2-27.
  31. Schra L, Ottens H H, Vlieger H. Fatigue crack growth in simulated Fokker 100 lap joints under MSD and SSD conditions. Report NLR CR 95729 C. NLR, Amsterdam; 1995.
  32. Schütz W. Zeitfestigkeit einschichtiger Leichtmetall-Nietverbindungen. Bericht Nr. F-47. Laboratorium für Betriebsfestigkeit, Darmstadt, 1963.
  33. Segerfröjd G, Wang G S, Palmberg B, Blom A F. Fatigue Behavior of Mechanical Joints: Critical Experiments and Statistical Analyses, ICAF 97: Fatigue in New and Ageing Aircraft: Proceedings of the 19th Symposium of the International Committee on Aeronautical Fatigue, Engineering Materials Advisory Services, Clifton-upon-Teme, England, U.K., 18-20 June 1997: 575-598.
  34. Segerfröjd G, Zuccherini S, Giovannelli G, Magnusson L. Fatigue behavior of Mechanical Joints - An experimental evaluation of ten different fastener systems and their influence on fatigue life. Combined Report, The Aeronautical Research Institute of Sweden, Report No. FKH R-4105, Sweden, January, 1997.
  35. Skorupa A, Skorupa M. Riveted Lap Joints in Aircraft Fuselage - Design, Analysis and Properties. Springer Dordrecht Heidelberg New York London, 2012, <https://doi.org/10.1007/978-94-007-4282-6>.
  36. Skorupa M, Skorupa A, Machniewicz T, Korbel A. An experimental investigation on the fatigue performance of riveted lap joint. In: Bos, M.J. (ed.) Proceedings of the 25th Symposium of the International Committee on Aeronautical Fatigue, Bridging the Gap between Theory and Practice, Springer, Rotterdam, 27-29 May 2009: 449-473, [https://doi.org/10.1007/978-90-481-2746-7\\_26](https://doi.org/10.1007/978-90-481-2746-7_26).
  37. Steadman D, Carter A, Ramakrishnan R. Characterisation on MSD in an in-service fuselage lap joint. In: 3rd Joint FAA/DoD/NASA Conference on Aging Aircraft, Albuquerque, N.M., Sept 1999 (Proceedings on CD-Rom); 1999.
  38. Stępień S, Szajnar S, Jaształ M. Problems of military aircraft crew's safety in condition of enemy counteraction. *Eksplatacja i Niezawodność - Maintenance and Reliability* 2017; 19 (3): 441-446, <https://doi.org/10.17531/ein.2017.3.15>.
  39. Szolwinski M P. The mechanics and tribology of fretting fatigue with application to riveted lap joints, PhD dissertation West Lafayette, IN, USA Purdue University, USA, 1998.
  40. Szymczyk E. Numeryczna analiza zjawisk lokalnych w połączeniach nitowych konstrukcji lotniczych (Numerical analysis of local phenomena in riveted joints of aircraft structures). WAT, Warszawa, 2016.
  41. Turaga V R S, Umamaheswar, Ripudaman Singh, Modelling of a patch repair to a thin cracked sheet, *Engineering Fracture Mechanics*, Volume 62, Issues 2-3, January-February 1999: 267-289, [https://doi.org/10.1016/S0013-7944\(98\)00088-5](https://doi.org/10.1016/S0013-7944(98)00088-5).
  42. Underhill P R, Juurlink J, DuQuesnay D L. The use of safety cuts in fatigue damaged fastener hole repair, *International Journal of Fatigue* 91, Part 1, October 2016: 242-247, <https://doi.org/10.1016/j.ijfatigue.2016.06.014>.
  43. Urban M R. Analysis of the fatigue life of riveted sheet metal helicopter airframe joints, *International Journal of Fatigue*; 2003; 25: 1013-1026, <https://doi.org/10.1016/j.ijfatigue.2003.08.003>.
  44. Witkowski R. Wprowadzenie do wiedzy o śmigłowcach (Introduction to helicopter topics). Biblioteka Naukowa Instytutu Lotnictwa, Warsaw, Poland, 1998.

## Prediction of remaining useful life for lithium-ion battery with multiple health indicators

Chun Su<sup>a, b, \*</sup>, Hongjing Chen<sup>a</sup>, Zejun Wen<sup>b</sup>

<sup>a</sup>School of Mechanical Engineering, Southeast University, Nanjing 211189, China

<sup>b</sup>Hunan Provincial Key Lab of Health Maintenance for Mechanical Equipment, Hunan University of Science and Technology, Xiangtan 411201, Hunan, China

Indexed by:



### Highlights

- Four types of health indicators (HIs) are built with the battery operating data.
- GRNN is applied to estimate the battery's remaining capacity with the HIs.
- Based on the predicted capacity value, the battery's RUL is estimated with NAR.

### Abstract

Lithium-ion (Li-ion) battery has become a primary energy form for a variety of engineering equipments. To ensure the equipments' reliability, it is crucial to accurately predict Li-ion battery's remaining capacity as well as its remaining useful life (RUL). In this study, we propose a novel method for Li-ion battery's online RUL prediction, which is based on multiple health indicators (HIs) and can be derived from the battery's historical operation data. Firstly, four types of indirect HIs are built according to the battery's operation current, voltage and temperature data respectively. On this basis, a generalized regression neural network (GRNN) is presented to estimate the battery's remaining capacity, and the non-linear autoregressive approach (NAR) is applied to predict the battery's RUL based on the estimated capacity value. Furthermore, to reduce the interference, twice wavelet denoising are performed with different thresholds. A case study is conducted with a NASA battery dataset to demonstrate the effectiveness of the method. The result shows that the proposed method can obtain Li-ion batteries' RUL effectively.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

lithium-ion (Li-ion) battery, remaining useful life (RUL), health indicator (HI), generalized regression neural network (GRNN), non-linear autoregressive (NAR).

## 1. Introduction

Lithium-ion (Li-ion) battery is the key power unit for a variety of engineering equipments, including portable computers, electric automobiles, satellites, and spacecrafts, etc. With the increased cycles of charge-discharge, the battery's capacity will deteriorate gradually, and when the capacity is inferior to a given threshold, failure may occur. This phenomenon will reduce the equipment's reliability and even lead to catastrophic failures [31]. Therefore, it is crucial to timely forecast Li-ion battery's capacity as well as its remaining useful life (RUL).

For Li-ion battery, its RUL can be defined as the number of remaining charge-discharge cycles before the battery's capacity deteriorates to a predetermined failure threshold [1]. The prediction methods for battery RUL can be categorized into three types, i.e. model-based approach, data-driven approach and hybrid approach [23]. Model-based approaches adopt mathematical representation or failure physics model to describe the degradation process of battery capacity, including electrochemical model [14, 16, 25], equivalent circuit model [6] and empirical model [33], etc. Considering the interactions among the various factors and the calculation of the parameters, it is not easy to establish a reliable and accurate prediction model for the electrochemical model and equivalent circuit model. Conversely, by fitting

a large amount of degradation data, empirical model is easier to be established [32].

Based on the battery's historical operational data, data-driven approach can extract the feature information and obtain the inherent degradation tendency. As a kind of data-driven approach, time series analysis has been widely used in RUL prediction, including autoregressive (AR) [12], autoregressive and moving average (ARMA) [18], autoregressive integrated moving average (ARIMA) [39] and the improved models [13, 35]. In recent years, artificial intelligence (AI) algorithm becomes a hot point in data-driven approaches due to its powerful ability in self-learning and data mining. The frequently-used AI algorithms include artificial neural network (ANN) [11, 21, 30], support vector machine (SVM) [15, 26], relevance vector machine (RVM) [36], etc.

By combining two or more model-based or data-driven approaches, hybrid approach can overcome the limitations of a single method and thus improve the accuracy and efficiency of the prediction [5, 37]. Hybrid approach can be a combination of model-based approach and data-driven approach, and it can also be a combination of two or more types of data-driven approaches [1, 22].

In most existing studies, the Li-ion battery's RUL is predicted with static capacity data. In fact, the battery's remaining capacity can only be measured when it is out of service, meanwhile it also requires strict

(\* ) Corresponding author.

E-mail addresses: C. Su - [suchun@seu.edu.cn](mailto:suchun@seu.edu.cn), H. Chen - [s820086099@163.com](mailto:s820086099@163.com), Z. Wen - [zjwen732@163.com](mailto:zjwen732@163.com)

testing conditions and environment. Therefore, it is difficult to be applied in engineering practice. To solve this problem, some other discriminant criteria can be used to estimate the battery's remaining capacity as well as its RUL, e.g. internal resistance [9, 10], incremental capacity analysis (ICA) [27, 28], open circuit voltage (OCV), entropy [2, 7, 29], etc [34]. Here, we call them the health indicators (HIs).

In this study, on the basis of battery's operation data, a novel hybrid approach is proposed for estimating Li-ion battery RUL, and multiple HIs are adopted concurrently. Generally, there are three steps in it. Fig. 1 is the flowchart of the proposed method.

**Step 1:** Calculating HIs. Based on the data, including battery's current, voltage and working temperature, four types of HIs are built accordingly. They are calculated resistance (CR), rate of temperature change (TR), duration time of equal discharging voltage difference (DTEDVD) and sample entropy of discharge voltage (SampEn), respectively.

**Step 2:** Estimating the remaining capacity. A generalized regression neural network (GRNN) with cross-validation is applied to estimate the battery's remaining capacity. To reduce the interference and improve the accuracy of the estimation, the wavelet denoising is performed for two times and with different thresholds.

**Step 3:** Predicting RUL. With the estimated remaining capacity, the battery's RUL is predicted via the non-linear autoregressive (NAR) method.

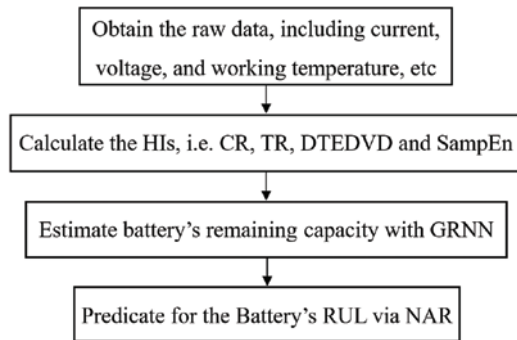


Fig. 1. Process of the proposed approach

The remainder of this article is organized as follows. Section 2 introduces the four HIs and their calculation methods respectively. Section 3 illustrates the basic theory of GRNN, NAR and wavelet denoising. Based on the Li-ion battery data from NASA, Section 4 conducts a case study of Li-ion battery RUL prediction, and the effectiveness of the proposed approach is illustrated. Conclusions and future works are given in Section 5.

## 2. Health indicators

To illustrate the discharging characteristics of the battery changing with the number of the charge cycles, in this study four HIs are extracted, they are CR, TR, DTEDVD and SampEn respectively. The value of the above four HIs will change with the increase of cycles, and the corresponding calculation methods are described as the follows.

### 2.1. Calculation of CR

For Li-ion battery, its internal resistance is closely related to the battery's remaining capacity. Saha et al [19] found that the remaining capacity of Li-ion batteries is linearly related to the sum of their electrolyte impedance and the charge transfer impedance.

CR is the measured internal resistance, which can be obtained with battery's terminal voltage and current during its operating state. Fig. 2 depicts the voltage and current's change at the instant of battery discharge. When a constant DC current passes through the battery, the terminal voltage changes  $\Delta U_n$  at the time  $t_1$ . At that time, the polariza-

tion effect has not occurred in the battery, therefore the voltage change is caused by ohm resistance. In this paper, the CR refers to the ohmic resistance  $R_n$ . According to Ohm's Law, it can be obtained as [18]:

$$R_n = \frac{\Delta U_n}{I_n} \quad n = 1, 2, \dots, N \quad (1)$$

where  $\Delta U_n$  is the voltage change caused by ohmic resistance of cycle  $n$ ;  $I_n$  is the load current in cycle  $n$ ; and  $N$  is the total cycle life of the battery.

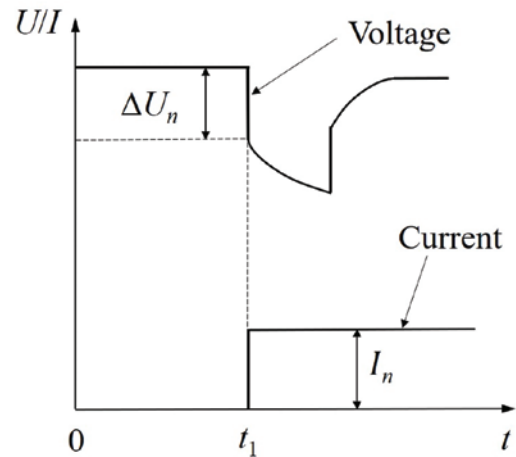


Fig. 2. Pulse discharge response of the battery

### 2.2. Calculation of TR

During the discharging process, the battery's internal temperature will change. Here we define TR as the rate that the temperature changes in the discharge process. TR will increase with the increase of cycles. In cycle  $n$ , TR can be calculated as follows:

$$TR_n = (TP_e^n - TP_s^n) / t_n \quad n = 1, 2, \dots, N \quad (2)$$

where  $TR_n$  denotes the rate of temperature change in cycle  $n$ ;  $TP_e^n$  is the end temperature of cycle  $n$ ;  $TP_s^n$  is the initial temperature of cycle  $n$ ; and  $t_n$  is the duration time of cycle  $n$ 's discharge process.

### 2.3. Calculation of DTEDVD

For Li-ion battery, the time for full discharge will be shortened gradually with the increase of the cycles. Therefore, the duration time of equal discharging voltage difference (DTEDVD) will also be shortened with the increase of the cycles' number. For a constant voltage interval  $[U_L, U_H]$ . We define  $t_D^n$  as the duration time in cycle  $n$  when the voltage is reduced from  $U_H$  to  $U_L$ , as shown in Fig. 3.

$$t_D^n = |t_H^n - t_L^n| \quad n = 1, 2, \dots, N \quad (3)$$

where  $t_D^n$  denotes the duration time of equal discharging voltage difference in cycle  $n$ ;  $t_H^n$  denotes the moment that the discharge voltage is  $U_H$  in cycle  $n$ ; and  $t_L^n$  denotes the moment that the discharge voltage is  $U_L$  in cycle  $n$ .

### 2.4. Calculation of SampEn

Sample entropy (SampEn) was firstly proposed by Richman and Moorman to measure the complexity of time series [17]. In this study, we regard the discharging process as a time series, and the SampEn of discharge voltage is used as an HI for the Li-ion battery.

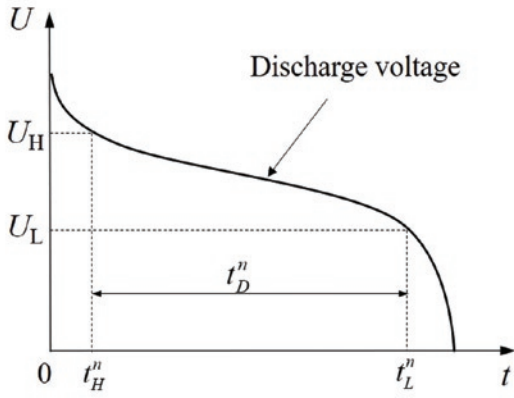


Fig. 3. Duration time of equal discharging voltage difference

For a given embedding dimension  $m$ , tolerance  $r$  as well as the number of data point  $T$ , we can obtain a time-series data set  $X = \{x_1, x_2, x_3, \dots, x_T\}$  with a constant time interval  $\tau$ . Here we define a template vector of length  $m$ ,  $X_m(\alpha) = [x(\alpha), x(\alpha+1), \dots, x(\alpha+m-1)]$   $\alpha = 1, 2, \dots, T-m+1$ . The distance between two template vectors of  $X_m(\alpha)$  and  $X_m(\beta)$  is define as  $d_m[X_m(\alpha), X_m(\beta)]$  ( $\alpha \neq \beta$ ). It can be any types of distance functions, including Euclidean distance, or Chebyshev distance.

$V^m(\alpha)$  is defined as the number of  $d_m[X_m(\alpha), X_m(\beta)] \leq r$  ( $\alpha \neq \beta$ ), and the definition of the function is as the follows [28]:

$$B_\alpha^m(r) = \frac{1}{T-m+1} V^m(\alpha) \quad \alpha = 1, 2, \dots, T-m+1 \quad (4)$$

$$B^m(r) = \frac{1}{T-m} \sum_{\alpha=1}^{T-m} B_\alpha^m(r) \quad (5)$$

where  $B^m(r)$  represents the probability that two sequences will match for  $m$  points. Similarly, we can obtain  $B^{m+1}(r)$ . Hence, the sample entropy can be defined as:

$$SampEn(r, m, T) = -\ln \left[ \frac{B^{m+1}(r)}{B^m(r)} \right] \quad (6)$$

### 3. Methodology

In this study, the battery's remaining capacity is estimated with the proposed HIs. On this basis, by analyzing the degradation trend of the remaining capacity, the battery's RUL is predicted.

GRNN is used to establish the mapping relationship between the HIs and the battery's capacity, and hereafter to estimate the remaining capacity. Here, NAR model is used to demonstrate the degeneration of battery capacity. Moreover, wavelet denoising is applied to reduce the interference of the estimated capacity. Here, we illustrate the basic theories for GRNN, NAR and wavelet denoising as the follows.

#### 3.1. Theory of GRNN

With a flexible network structure, GRNN model has strong ability for non-linear mapping. Therefore, it is suitable for solving non-linear problems [38].

Generally, there are four layers in GRNN, i.e. input layer, pattern layer, summation layer, and output layer, as shown in Fig. 4. The input of the network is  $\mathbf{X} = [x_1, x_2, \dots, x_e]^T$ , and its output is

$\mathbf{Y} = [y_1, y_2, \dots, y_k]^T$ .  $d$  is the dimension of the input vector,  $k$  is the dimension of the output sample, and  $m$  is the number of the learning sample.

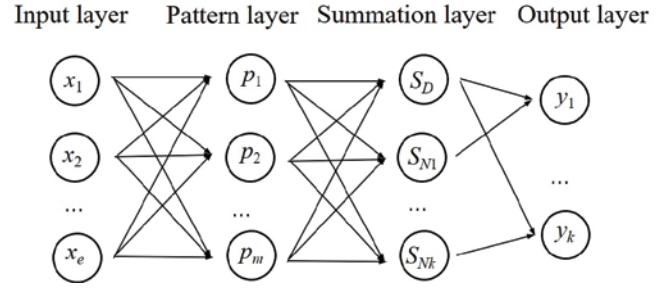


Fig. 4. Basic structure of the GRNN

- (1) Input layer: Among it, each neuron is a simple distribution unit, which will transfer the input variable directly to the pattern layer. In addition, the number of the input neurons in a GRNN equals to the dimension of the input vector  $e$ .
- (2) Pattern layer: The number of neurons equals to the number of learning sample  $m$ . Each neuron corresponds to a different sample, and the neurons transfer function can be calculated as [36]:

$$p_i = \exp \left[ -\frac{(\mathbf{X} - \mathbf{X}_i)^T (\mathbf{X} - \mathbf{X}_i)}{2\sigma^2} \right] \quad i = 1, 2, \dots, m \quad (7)$$

where  $\sigma$  is the smoothing parameter;  $\mathbf{X}$  is the input variable; and  $\mathbf{X}_i$  is the corresponding learning sample of neuron  $i$ .

- (3) Summation layer: It sums the output of the pattern layer. In this layer, two types of neurons are used for summation. The simple summation of the pattern outputs is determined with:

$$S_D = \sum_{i=1}^m p_i \quad (8)$$

And, the weighted summation of the pattern output can be determined by:

$$S_{Nj} = \sum_{i=1}^m y_{ij} p_i \quad j = 1, 2, \dots, k \quad (9)$$

where  $y_{ij}$  is the weight of the  $i$ -th neuron in the pattern layer, which is connected to the summation layer.

- (4) Output layer: The results calculated in the summation layer will be sent to the output layer. The number of neurons is equal to the dimension  $k$  of the output vector in the sample. The output of the neurons is calculated as follows:

$$y_j = \frac{S_{Nj}}{S_D} \quad j = 1, 2, \dots, k \quad (10)$$

where  $y_j$  is the output of the  $j$ -th node in the output layer.

In this study, the cross-validation is used to gain the optimized parameters.

#### 3.2. Theory of NAR

Here, by combining the variable's information during the early time, NAR model is used to describe the variable information at a particular time. It is a form of time series [8]:

$$y(t) = f[y(t-1), y(t-2), \dots, y(t-p)] \quad (11)$$

where  $p$  is the delay time, namely the order of the neural network (NN); and  $f(\cdot)$  is the model of NAR.

The structure of NAR-NN is shown as in Fig. 5. It is consisted of four layers, i.e. input layer, hidden layer, output layer and the delay function.

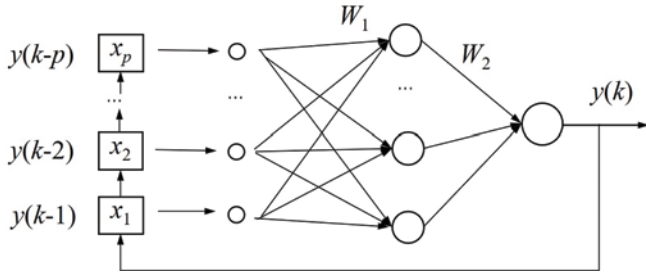


Fig. 5. Structure of NAR-NN

In Fig.5,  $x_u$  ( $1 \leq u \leq p$ ) is the input;  $y(k)$  is the output;  $W_1$  is the connection weight between the input layer and the hidden layer;  $W_2$  is the connection weight between the hidden layer and the output layer; and  $p$  is the delay time of the output.

The output of the node  $v$  in the hidden layer is:

$$h_v = g_1 \left( \sum_{u=1}^p w_{uv} x_u + c_v \right) \quad v = 1, 2, \dots, V \quad (12)$$

where  $g_1(\cdot)$  is the activation function of the hidden nodes;  $V$  is the number of the hidden nodes;  $x_u$  is the  $u$ -th delay output of the output signal  $y$ ;  $w_{uv}$  is the connection weight between the hidden node  $u$  and the delay node  $v$ ; and  $c_v$  is the  $v$ -th neural threshold of the hidden layer.

The output of NAR is:

$$y = g_2 \left( \sum_{v=1}^V w_v h_v + d \right) \quad (13)$$

where  $w_v$  is the connection weight between the hidden node  $v$  and the output node; and  $d$  is the neural threshold of the output layer.

### 3.3. Wavelet denoising

Wavelet analysis is a kind of method for signal analysis, which can separate the noise signals and extract effective data from the noise-mixed data. In wavelet analysis, the data can be decomposed into two parts with the expansion and translation of the mother wavelet, i.e. approximate part and detail part [24].

Let  $H(x)$  denote the measured signal mixed with noise, the wavelet transform with the mother wavelet  $\psi(x)$  can be expressed as:

$$c(a, b) = \langle H(x), \psi_{a,b}(x) \rangle = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} H(x) \psi \left( \frac{x-b}{a} \right) \quad (14)$$

where  $a$  and  $b$  are the expansion and translation parameters of the mother wavelet respectively; and  $c(a, b)$  is the wavelet coefficient of  $H(x)$ .

During the process of wavelet denoising, the selection of wavelet denoising threshold is closely related to the consequence of noise reduction. In this study, a twice denoising method with different thresh-

olds is applied [3, 4]. The first denoising threshold is made by the Sqtwolog method:

$$\lambda_{sqt} = \sqrt{2 \log N} \quad (15)$$

And, the second denoising threshold is finished by the minimax method:

$$\lambda_{\min} = 0.3936 + 0.1829 \log_2^N \quad (16)$$

where  $N$  is the number of signal  $H(x)$ . In this paper, it equals to the cycle life of the battery.

The thresholds of the above two methods are usually different, and the threshold of the Sqtwolog method is larger than that of the minimax method. Therefore, the first wavelet denoising can remove the larger noise signal, while the second wavelet denoising can remove the smaller noise signal. By using the twice wavelet denoising with different thresholds, the interference among the measured data can be reduced effectively.

### 3.4. Implementation of the proposed approach

Here, the training batteries are used to establish the mapping relationship between the HIs and the battery capacity. The basic steps for estimating battery RUL are described as follows:

- (1) Based on the historical operation data, the HIs of the training batteries and testing batteries are calculated respectively.
- (2) By using the training batteries, train a GRNN model to establish the mapping relationship between the HIs and the battery's capacity.
- (3) With the HIs data of testing batteries in the previous cycles, the capacity of the testing batteries is estimated on the basis of the well-trained GRNN model.
- (4) The wavelet denoising is applied to deduce the noising of the estimated capacity.
- (5) With the established capacity sequence and by using the NAR model, the remaining capacity and RUL of the testing battery can be predicted.

## 4. Prognostics Experiment

### 4.1. Experiment Data and the Procedure

In this section, a case study is conducted to demonstrate the effectiveness and efficiency of the proposed approach, where the Li-ion battery data come from the data repository of the NASA Ames Prognostics Center of Excellence (PCoE) [20]. Here, three groups of batteries are taken as examples, i.e. B05, B06 and B07. Among them, B05 and B06 are used as the training batteries, and B07 is used as the target battery. The experimental parameters of the discharge process are listed as in Table 1.

Table 1. Discharging parameters of the NASA batteries

Battery No.	$T/^\circ\text{C}$	$I_{dis}/\text{A}$	$V_{up}/\text{V}$	$V_{low}/\text{V}$	$C_{new}/\text{Ah}$
B05	24	2	4.2	2.7	1.86
B06	24	2	4.2	2.5	2.04
B07	24	2	4.2	2.2	1.89

The capacity's degradation trends with the cycles are shown as in Fig. 6. Obviously, the battery's capacity shows a non-monotonic decline with the increase of the cycles. Meanwhile, the degradation is accompanied with the phenomena of capacity self-regeneration as

well as local fluctuations. Therefore, it is essential to develop an effective prognostic approach to accurately estimate the capacity and predict its RUL by overcoming the uncertainty fluctuation.

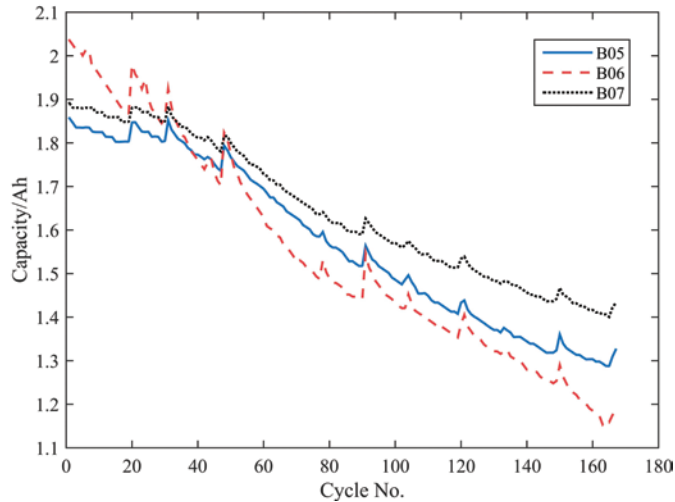
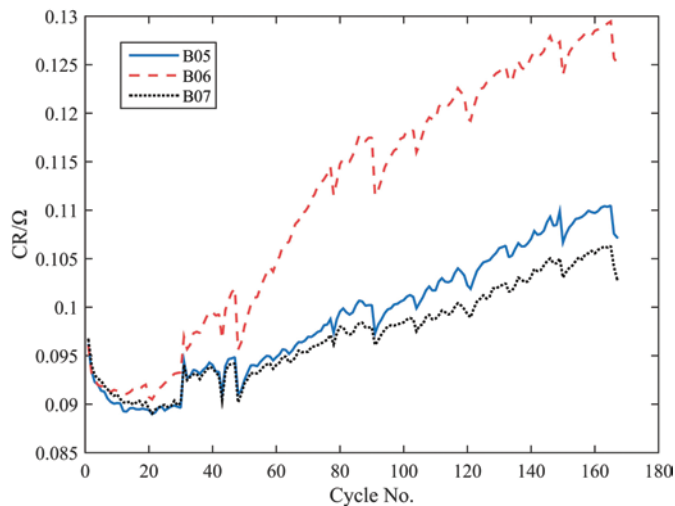
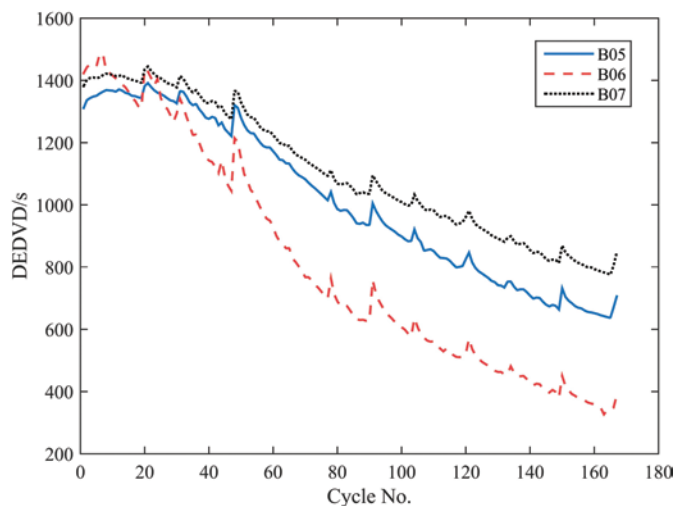


Fig. 6. Capacity degradation of tested batteries



(a) Variations of CR



(c) Variations of DTEDVD

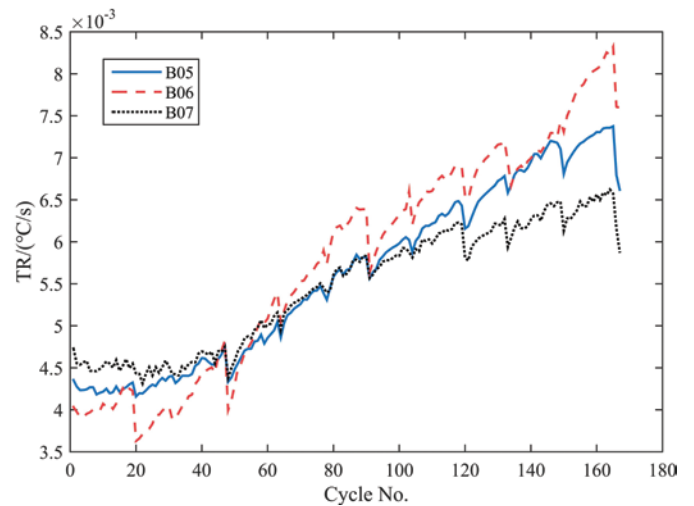
process, we calculate the HIs for batteries of B05, B06 and B07 respectively. Fig. 7 shows the variation of the four HIs with the increase of cycles respectively.

Next, we calculate the Pearson correlation coefficient between the four HIs and the original capacity of Li-ion batteries, as shown in Table 2. Obviously, the HIs have a close correlation with the capacity. Furthermore, the DTEDVD has a negative correlation with the capacity, while the other three HIs are in positive correlation.

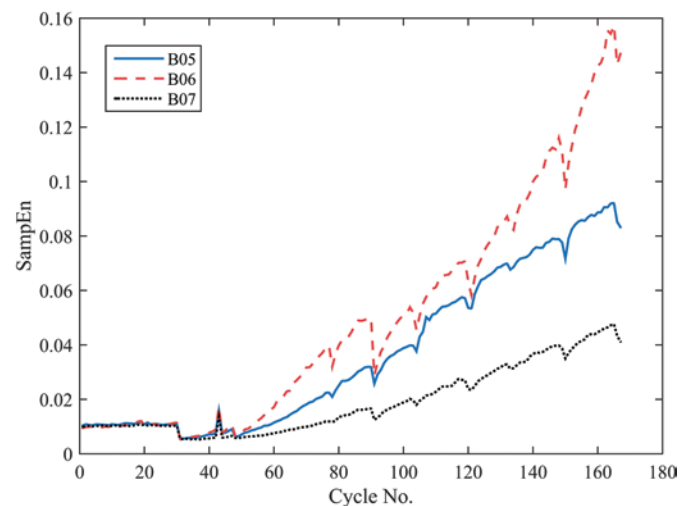
### 4.3. Establishing GRNN model

With the obtained HIs data, GRNN model is used to train and establish the mapping relationship between the HIs (input) and the capacity (output). In this study, the data of B05 and B06 batteries are used for training. On that basis, the well-trained GRNN model is used to estimate B07's capacity with B07's HIs. The predicted results are shown in Fig. 8.

As shown in Fig. 8, the proposed method has a relatively high prediction accuracy concerned with the capacity in the long-term trend. However, the predicted result is not so good for the first 40 points. The reason is that the HIs are not so stable during the early stage. Therefore, the first 40 points can be removed to obtain a more accurate prediction result.



(b) Variations of TR



(d) Variations of SampEn

Fig. 7. Variations of the four HIs changing with the charging cycles

### 4.2. Calculating HIs

In this section, by using the battery's historical operation data (including current, voltage and temperature) during the discharging

Moreover, the indicators, including root means square error (RMSE), the maximum estimation error and the maximum covariance (*cov*) value, are utilized to evaluate the accuracy of the estima-

Table 2. Correlation coefficient between HIs and capacity

Battery No.	B05	B06	B07
CR	-0.9791	-0.9893	-0.9658
TR	-0.9947	-0.9819	-0.9901
DTEDVD	0.9988	0.9957	0.9990
SampEn	-0.9522	-0.8936	-0.8876

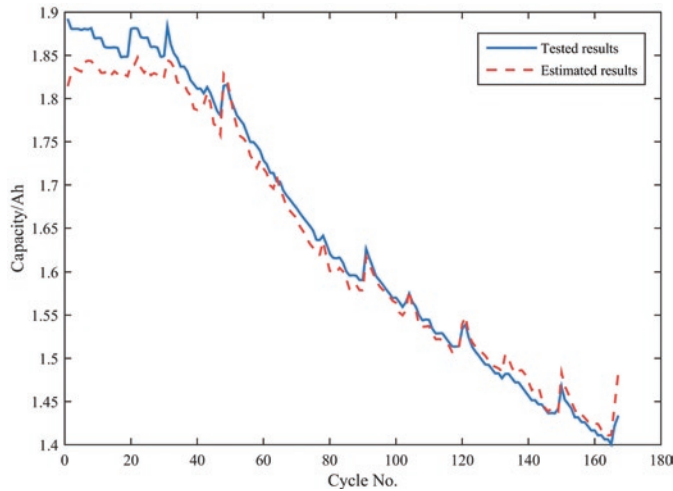


Fig. 8. The estimated and tested capacity of B07

Table 3. Accuracy analysis of the capacity estimation

Indicators	All points	Remove the first 40 points
RMSE	0.2002	0.0140
Max error	0.0823	0.0267
Max cov	0.0259	0.0110

tion results, as shown in Table 3. Obviously, the capacity estimation is with high accuracy.

#### 4.4. RUL prediction

To remove the influence of self-regeneration phenomena and local fluctuations, a twice wavelet denoising is implemented with different thresholds. The result is shown as in Fig. 9.

In this case, the threshold of Sqtwolog method is larger than that of minimax method. As shown in Fig. 9, the first denoising can only eliminate the large noise signal, and the second denoising can elimi-

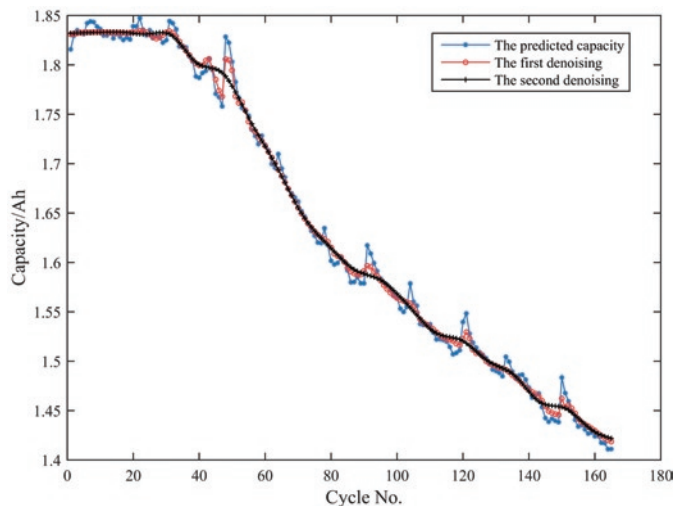


Fig. 9. The predicted capacity of B07

nate the smaller noise signal. Thus, with the proposed twice wavelet denoising, the noise can be eliminated more thoroughly and a more accurate prediction can be obtained.

Based on the historical operation data, we can estimate the remaining capacity for each past cycle with the above steps. In addition, by selecting the obtained data points as the input set, the future degraded capacity can be predicted with the NAR model. On that basis, we can obtain the corresponding value of the RUL. Here, the number of the hidden layer is set as 1, the delay time is 20. It is supposed that the

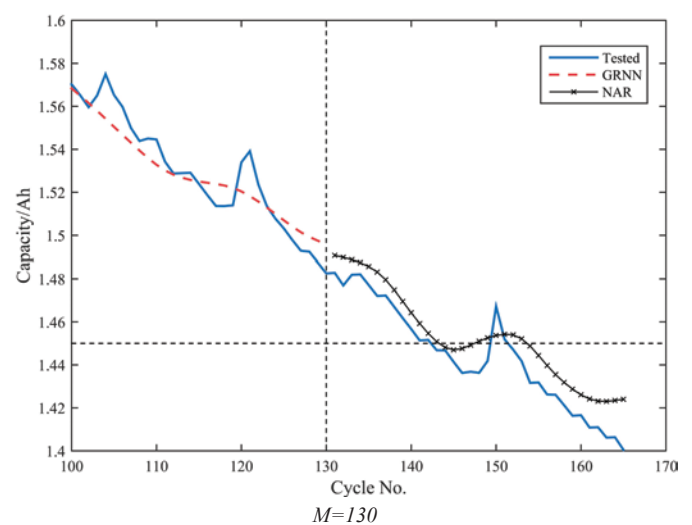
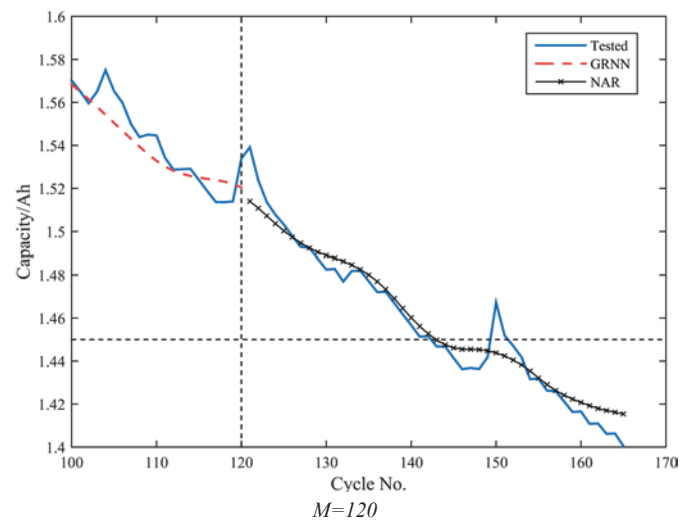
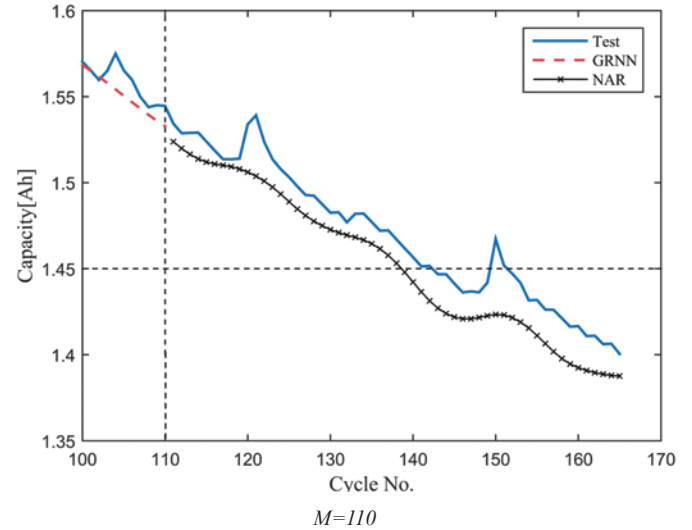


Fig. 10. Predictions for B07's RUL under different M

Table 4. The predicted results of B07's RUL

M	Actual EOL	Actual RUL	Predicted RUL	Percentage Error/%
110	142	32	28	14.3
120	142	22	23	4.3
130	142	12	14	14.3

failure threshold of the battery capacity is 1.45Ah, and the total cycle life of B07 is 142. Here, we predict the B07's RUL when the operated cycles ( $M$ ) are 110, 120, 130 respectively. In addition, considering that there is randomness in the training model, the predicted RUL will change with the experiments, thus we repeat the experiments for 100 times. Fig. 10 and Table 4 show some of the predicted results.

Set the range of the 100 predicted results as the RUL uncertainty bound (RUB). For the battery of B07, the RUB is shown as in Table 5.

Table 5. The RUB of B07 battery

M	Actual EOL	Actual RUL	RUB	Range
110	142	32	[28, 55]	27
120	142	22	[23, 45]	22
130	142	12	[14, 35]	21

With the above predicted results, we can conclude that the proposed methods can obtain the decline tendency of the Li-ion battery capacity with a high accuracy. However, the RUL's predicted result is not so stable. The possible reason is that the tested life cycle of the battery is not long enough. Therefore, the training data are not sufficient. Furthermore, considering that the capacity change during each cycle is too small, and a small error in capacity prediction may lead to a larger error in the RUL's prediction. It can also be found that the

Table 6. Predicted RUL results based on ARIMA model

M	Actual RUL	Predicted RUL	Error	Percentage Error/%
110	32	38	6	18.9
120	22	28	6	27.3
130	12	18	6	50.0

range of RUB will be narrowed down with the increase of the input cycles. It shows that the RUL prediction is sensitive to the number of input data. Therefore, by increasing the number of the training samples, the stability and robustness of the predictions can be improved.

## References

- Chang Y, Fang H, Zhang Y. A new hybrid method for the prediction of the remaining useful life of a lithium-ion battery. *Applied Energy* 2017; 206: 1564–1578, <https://doi.org/10.1016/j.apenergy.2017.09.106>.
- Chen L, Xu L, Zhou Y. Novel approach for lithium-ion battery on-Line remaining useful life prediction based on permutation entropy. *Energies* 2018; 11(4): 820. <https://doi.org/10.3390/en11040820>.
- Donoho D L, Johnstone I M. Threshold selection for wavelet shrinkage of noisy data. *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 1994; A24–A25, <https://doi.org/10.1109/iembs.1994.412133>.
- Donoho D L, Johnstone I M. Adapting to unknown smoothness via wavelet shrinkage. *Fundamental Papers in Wavelet Theory*. Princeton University Press 2014: 833–857, <https://doi.org/10.1515/9781400827268.833>.
- Feng J, Kvam P, Tang Y. Remaining useful lifetime prediction based on the damage-marker bivariate degradation model: A case study on lithium-ion batteries used in electric vehicles, *Engineering Failure Analysis* 2016; 70: 323–342, <http://dx.doi.org/10.1016/j.engfailanal.2016.04.014>.
- Hu X S, Li S, Peng H. A comparative study of equivalent circuit models for Li-ion batteries. *Journal of Power Sources* 2012; 198:359-367, <https://doi.org/10.1016/j.jpowsour.2011.10.013>.
- Hu X, Li S E, Jia Z, et al. Enhanced sample entropy-based health management of Li-ion battery for electrified vehicles. *Energy* 2014; 64: 953–960. <https://doi.org/10.1016/j.energy.2013.11.061>.
- Ibrahim M, Jemei S, Wimmer G, et al. Nonlinear autoregressive neural network in an energy management strategy for battery/ultra-capacitor

ARIMA model is also widely used in the battery's RUL prediction [21]. To compare the characteristics of different prediction models, here we make a comparison for RUL prediction with the ARIMA model. For the same input, the ARIMA method has a unique prediction result. Thus, the prediction results do not have a RUB. The RUL predicted results with ARIMA model are shown as in Table 6.

By comparing the three predicted results of different  $M$ , it can be found that for different  $M$ , the errors keep the same. It means that the prediction accuracy is not sensitive to the number of input cycles.

Comparing Table 4 and Table 6, it can be found that the predicted results of NAR model are more accurate than the results of ARIMA model. However, the predicted results are not as stable as the results of ARIMA model. Furthermore, since NAR model is sensitive to the number of the input data, the proposed NAR method is more suitable for large data set.

## 5. Conclusions

In this study, an innovative hybrid data-driven method is proposed to predict Li-ion battery's RUL, which is based on multiple HIs, including CR, TR, DTEDVD and SampEn. Furthermore, the GRNN model, NAR model and twice wavelet denoising are integrated. Case study shows that the proposed method can achieve Li-ion battery's RUL prediction with a high accuracy. Compared with ARIMA model, NAR model is more sensitive to the size of training samples. Meanwhile, the proposed method is established on the basis of battery historical data, it can overcome the limitation of online capacity measurement approaches. Moreover, the proposed RUL prediction method does not concern the physical and chemical reactions in the battery, thus it is well suitable for other kinds of batteries.

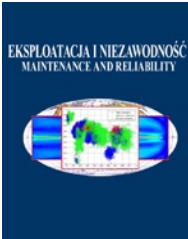
The future works can be focused on the following aspects: (1) to improve the robustness of the RUL prediction results, the NAR model can be improved further. (2) new HIs can be searched for more stable prediction, especially for the initial stage. (3) new RUL prediction methods can be developed by considering various operating environments of the Li-ion batteries.

### Acknowledgement

This work is supported by the National Natural Science Foundation of China under No.71671035; the Open Fund of Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment, China under No.201901; Scientific Research Fund of Hunan Provincial Education Department, China under Grant No. 17A069, and the Open Fund of Jiangsu Wind Power Engineering Technology Center of China under ZK19-03-03.



- hybrid electrical vehicles. *Electric Power System Research* 2016; 136: 262–269, <https://doi.org/10.1016/j.epsr.2016.03.005>.
9. Kim I S. A technique for estimating the state of health of lithium batteries through a dual-sliding-mode observer. *IEEE Transactions on Power Electronics* 2009; 25(4): 1013–1022, <https://doi.org/10.1109/tpel.2009.2034966>.
  10. Kim J, Lee S, Cho B H. Complementary cooperation algorithm based on DEKF combined with pattern recognition for SOC/capacity estimation and SOH prediction. *IEEE Transactions on Power Electronics* 2011; 27(1):436–451, <https://doi.org/10.1109/tpel.2011.2158554>.
  11. Li Y, Wang K. Modified convolutional neural network with global average pooling for intelligent fault diagnosis of industrial gearbox. *Eksploracja i Niezawodność – Maintenance and Reliability* 2020; 22 (1): 63–72, <http://dx.doi.org/10.17531/ein.2020.1.8>.
  12. Liu D, Luo Y, Liu J, et al. Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation AR model and RPF algorithm. *Neural Computing and Applications* 2014; 25: 557–572, <https://doi.org/10.1007/s00521-013-1520-x>.
  13. Long B, Xian W, Jiang L, Liu Z. An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries. *Microelectronics Reliability* 2013; 53(6): 821–831, <https://doi.org/10.1016/j.microrel.2013.01.006>.
  14. Lyu C, Lai Q, Ge T, et al. A lead-acid battery's remaining useful life prediction by using electrochemical model in the particle filtering framework. *Energy* 2017; 120: 975–984, <https://doi.org/10.1016/j.energy.2016.12.004>.
  15. Patil M A, Tagade P, Hariharan K S, et al. A novel multistage support vector machine based approach for Li ion battery remaining useful life estimation. *Applied Energy* 2015; 159: 285–297, <https://doi.org/10.1016/j.apenergy.2015.08.119>.
  16. Prasad G K, Rahn C D. Model based identification of aging parameters in lithium ion batteries. *Journal of Power Sources* 2013; 232: 79–85, <https://doi.org/10.1016/j.jpowsour.2013.01.041>.
  17. Richman J S, Moorman J R. Physiological time-series analysis, using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology* 2000; 278(6): H2039–H2049, <https://doi.org/10.1152/ajpheart.2000.278.6.h2039>.
  18. Saha B, Goebel K, Christophersen J. Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement and Control* 2009; 31: 293–308, <https://doi.org/10.1177/0142331208092030>.
  19. Saha B, Poll S, Goebel K, et al. An integrated approach to battery health monitoring using Bayesian regression and state estimation. *Proceedings of IEEE Autotestcon, 2007*; 646–653, <https://doi.org/10.1109/autest.2007.4374280>.
  20. Saha and B, Goebel K. Battery Data Set, NASA ames prognostics data repository. NASA Ames Research Center, 2007.
  21. Sbarufatti C, Corbetta M, Giglio M, et al. Adaptive prognosis of lithium-ion batteries based on the combination of particle filters and radial basis function neural networks. *Journal of Power Sources* 2017; 344: 128–140, <https://doi.org/10.1016/j.jpowsour.2017.01.105>.
  22. Song Y, Liu D, Yang C, et al. Data-driven hybrid remaining useful life estimation approach for spacecraft lithium-ion battery. *Microelectronics Reliability* 2017; 75: 142–153, <https://doi.org/10.1016/j.microrel.2017.06.045>.
  23. Su C, Chen H J. A review on prognostics approaches for remaining useful life of lithium-ion battery. *IOP Conference Series: Earth and Environmental Science* 2017; 93(1): 012040, <https://doi.org/10.1088/1755-1315/93/1/012040>.
  24. Valencia D, Orejuela D, Salazar J, et al. Comparison analysis between rigsure, sqtwolog, heursure and minimaxi techniques using hard and soft thresholding methods. *Proceedings of XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA) 2016*; 1–5, <https://doi.org/10.1109/STSIVA.2016.7743309>.
  25. Waag W, Kbitz S, Sauer D U. Experimental investigation of the lithium-ion battery impedance characteristic at various conditions and aging states and its influence on the application. *Applied Energy* 2013; 102: 885–897. <https://doi.org/10.1016/j.apenergy.2012.09.030>.
  26. Wei J, Dong G, Chen Z. Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression. *IEEE Transactions on Industrial Electronics* 2017; 65(7): 5634–5643, <https://doi.org/10.1109/tie.2017.2782224>.
  27. Weng C, Cui Y, Sun J, et al. On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression. *Journal of Power Sources* 2013; 235(4): 36–44, <https://doi.org/10.1016/j.jpowsour.2013.02.012>.
  28. Weng C, Sun J, Peng H. Model parametrization and adaptation based on the invariance of support vectors with applications to battery state-of-health monitoring. *IEEE Transactions on Vehicular Technology* 2015; 64(9): 3908–3917, <https://doi.org/10.1109/tvt.2014.2364554>.
  29. Widodo A, Shimb M C, Caesarendra W, et al. Intelligent prognostics for battery health monitoring based on sample entropy. *Expert Systems with Applications* 2011; 38(9): 11763–11769, <https://doi.org/10.1016/j.eswa.2011.03.063>.
  30. Wu J, Zhang C, Chen Z. An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks. *Applied Energy* 2016; 173: 134–140, <https://doi.org/10.1016/j.apenergy.2016.04.057>.
  31. Xia Q, Wang Z, Ren Y, Sun B, et al. A reliability design method for a lithium-ion battery pack considering the thermal disequilibrium in electric vehicles. *Journal of Power Sources* 2018; 386:10–20, <https://doi.org/10.1016/j.jpowsour.2018.03.036>.
  32. Xia Q, Yang D, Wang Z, Ren Y, et al. Multiphysical modeling for life analysis of lithium-ion battery pack in electric vehicles. *Renewable and Sustainable Energy Reviews* 2020; 131: 109993, <https://doi.org/10.1016/j.rser.2020.109993>.
  33. Xing Y, Ma E W M, Tsui K L, et al. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. *Microelectronics Reliability* 2013; 53(6): 811–820, <https://doi.org/10.1016/j.microrel.2012.12.003>.
  34. Yan S, Ma B, Zheng C. Health index extracting methodology for degradation modelling and prognosis of mechanical transmissions. *Eksploracja i Niezawodność – Maintenance and Reliability* 2019; 21 (1): 137–144, <http://dx.doi.org/10.17531/ein.2019.1.15>.
  35. Yu J, Yang J, Tang D, Dai J. Early prediction of remaining discharge time for lithium-ion batteries considering parameter correlation between discharge stages. *Eksploracja i Niezawodność – Maintenance and Reliability* 2019; 21 (1): 81–89, <http://dx.doi.org/10.17531/ein.2019.1.10>.
  36. Zhang C L, He Y G, Yuan L F, et al. Capacity prognostics of lithium-ion batteries using EMD denoising and multiple kernel RVM. *IEEE Access* 2017; 5: 12061–12070, <https://doi.org/10.1109/access.2017.2716353>.
  37. Zheng X, Fang H. An integrated unscented Kalman filter and relevance vector regression approach for lithium-ion battery remaining useful life and short-term capacity prediction. *Reliability Engineering & System Safety* 2015; 144: 74–82, <https://doi.org/10.1016/j.res.2015.07.013>.
  38. Zhou J, He Z, Gao M, et al. Battery state of health estimation using the generalized regression neural network. *Proceedings of 8th International Congress on Image and Signal Processing (CISP), 2015*; 1396–1400. <https://doi.org/10.1109/cisp.2015.7408101>.
  39. Zhou Y, Huang M. Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model. *Microelectronics Reliability* 2016; 65: 265–273, <https://doi.org/10.1016/j.microrel.2016.07.151>.



Article citation info:

Dziubak T, Wysocki T, Dziubak S. Selection of vehicles for fleet of transport company on the basis of observation of their operational reliability. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 184–194, <http://dx.doi.org/10.17531/ein.2021.1.19>.

## Selection of vehicles for fleet of transport company on the basis of observation of their operational reliability

Indexed by:



Tadeusz Dziubak<sup>a</sup>, Tadeusz Wysocki<sup>a</sup>, Sebastian Dziubak<sup>a</sup>

<sup>a</sup>Military University of Technology, Faculty of Mechanical Engineering, ul. Gen. Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland

### Highlights

- A research method was proposed which aims to study the operational reliability of the vehicles.
- The operational reliability indices describing the failure frequency of the vehicles were used.
- An evaluation and comparison of the operational reliability of four makes of the vehicles operated was carried out.
- The reasonability to unify the fleet of the light commercial vehicles was confirmed.

### Abstract

The exploitation reliability of the contemporary light commercial vehicles of four makes is subject to a quantitative analysis in the article. The issue concerning the quantitative evaluation of the operational reliability of the fleet of the commercial vehicles in the transport company and their selection for this fleet is not very numerously described in the literature. For the purposes of evaluating the operational reliability, results of monitoring failures of 115 vehicles during their several years of exploitation in the similar conditions in the urban area and its surroundings were used. A proposed test method uses the relative quantitative measures of the operational reliability indices targeted at the mileage of vehicles and frequency of the failures. The aforementioned enabled to identify the make of the vehicles, which during exploitation was characterized by the lowest failure frequency and thus reliability. This, in turn, allowed for a favorable selection of the vehicles within the next years of the operation of the company. The results of the study confirmed the possibility of using the proposed method to evaluate the operational reliability of the commercial vehicles in any transport company.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

exploitation, light commercial vehicles, exploitation reliability, indices of operational reliability evaluation.

## 1. Introduction

The selection of the vehicles for a fleet of the company should depend primarily on a profile of its economic activity and current transport needs. The transport fleet is of necessity in the economic activity where the sales and distribution of goods to the customer or the carriage of passengers is carried out. However, while selecting the vehicles, special attention should be paid not only to their technical parameters, including cargo parameters, but also to the efficiency and economy of the drive unit (engine power, fuel consumption, fulfilment of the Euro standards). An important performance parameter is also their failure frequency. The lower the failure frequency, the lower the costs of maintaining the fleet and the risk of failure to complete the task. Certainly, other criteria should be adopted in case of selecting, for example, passenger cars for the department of goods distribution in the undertaking, and others in the event of the commercial vehicles for the purposes of the goods distribution.

The planning and implementation of the exploitation process of the motor vehicles in a complex transport system requires the efficient use of the means of transport and constitutes one of the main manners of achieving a competitive advantage in the present-day market.

When a utility vehicle (commercial vehicle or light commercial vehicle) is selected by the entrepreneur, exploitation costs have a major impact, apart from the vehicle model, make and purchase price, as well as the exploitation decline in value.

Physical aging of the vehicle components resulting from the processes in the individual systems and matching elements, momentary overloads of these components and random events (e.g., collisions and road accidents) affect both the durability and reliability of the entire vehicle [37]. The evaluation of costs should take into account not only the predictable (planned) operating costs, but also the “incidental costs”, which are strictly related to the operational reliability of the vehicles. Their importance is demonstrated by their significant impact on the guarantee and a completion of the transport service in a fashion manner. A failure frequency of the vehicles and the losses incurred due to it are in many cases the main factor for which the user loses confidence in a given make. Thus, the procedure for selecting vehicles by the user should consider an aspect of their technical failure frequency (operational reliability), including the type of failure, its frequency and repair costs.

There are researches in the source literature which present a reliability approach to the issues of the optimal selection of the means of transport to be used by a company for its transport operations [25], but

E-mail addresses: T. Dziubak - [tadeusz.dziubak@wat.edu.pl](mailto:tadeusz.dziubak@wat.edu.pl), T. Wysocki - [tadeusz.wysocki@wat.edu.pl](mailto:tadeusz.wysocki@wat.edu.pl), S. Dziubak - [sebastian.dziubak@wat.edu.pl](mailto:sebastian.dziubak@wat.edu.pl)

they are not numerous. Furthermore, the network analysis methods (CPM, PERT, GERT) often presented in the publications do not take this approach, as they are limited to the determination of the critical path for the planned transport of goods. In particular, as emphasized in [41], in an era of the strong market competition, the optimal selection of the means of transport for the undertaken transport task reduces the own costs of the company. This allows for a significant reduction in the price of the service and makes it easier to adjust the time and frequency of goods delivery to the customer expectations. However, in the procedure of optimal selection of the means of transport for the purposes of carrying out the undertaken transport task, it is proposed to take into account, *inter alia*, the criteria of availability, utilization of loading capacity and applicable legal restrictions, e.g., the EURO standards concerning the selected means of transport. The basis for such considerations can be found in the papers [42].

## 2. State of the art related to the research subject matter

In a system of exploitation, many different reliability indices are used in order to evaluate the efficiency of the use and handling of the vehicles, for example in respect of to the selected vehicle system [28, 32] or system component [33] in terms of the total reliability.

The authors of the paper [32], while using the reliability indices, concluded that the reliability of the individual structural systems of the two researched bus makes is different, in particular in terms of their mileage to first failure, whereby the system of the public transport buses that is characterized by the highest failure frequency turned out to be the engine itself. Whereas, the lowest risk of failure was determined in the power and braking systems. These results are inconsistent with the data obtained by the authors of the paper [28], who had evaluated a quantity of failures per public transport bus and per 10,000 km and the mileage between the failures and had stated that the braking system and external lighting, alarm system, tires and steering system had made the greatest impact (32-50%). The failure data of the individual systems were in consonance with the exponential distribution. The study is highly reliable as it had been conducted over a three-year period on a population of 160 public transport buses. The reliability curve shows that as an age of the vehicle increases, so does the failure frequency, and an established optimum useful life of the vehicle is 72 months. The aforementioned data is similar to the results of the comparative reliability studies of two selected makes of the commercial vehicles under the long-term rental conditions presented in the paper [36]. A description of reliability of the subassemblies and components was drawn up taking into account the mileage to first failure and the mean mileage between the repairs.

In these publications it was stated that the systems that feature the highest failure frequency of the studied vehicles in terms of the mileage to first failure were engine and powertrain, electrical, brake and exhaust systems. The aforementioned partially confirms the results of the analysis presented in [28]. Furthermore, it proves that one group of the vehicles is characterized by much higher reliability and lower exploitation costs than the vehicles of the second make, which coincides with a similar ascertainment presented in [32]. The similar findings were made by the authors of the paper [31], proving that the greatest repair costs are generated by failures to the suspension system and wheels and steering system, engine and powertrain and braking system. The analyzed systems of the vehicle were also mentioned by other authors [28, 32, 36] as the systems generating the highest quantity of failures and the highest repair costs. The above analyses fail to cover the repair time, which is also an important cost-generating factor, as apparently the minor failures may be the cause of the long-standing taking out of service of the vehicles. The problem of time in the reliability evaluation was presented by the authors of the paper [13], who used three reliability indices based on the averaged time in order to evaluate the reliability of the main components of the electric vehicles: MTTF – mean time to failure, MTTR - mean time to repair

(downtime), MTBF - mean time between failures (cycle time that is a sum of operating time and downtime). Similarly defined reliability indices were presented by the authors of the paper [43] to evaluate the vehicle failure frequencies (Load-Haul-Dump) in an underground rock mine in Canada. They used the total time between failures, the time to repair (TTR), the time between failures (TBF) and, moreover, the machine downtime during a shift to evaluate the systems and components to which the generated breakdowns and failures were ascribed.

The author of the papers [39, 40] asserts that the aforementioned methods of the reliability description applied to the objects working to first failure, e.g., using only the reliability function  $R(t)$  or failure intensity  $\lambda(t)$  are insufficient. Reliability was presented by the author as a comprehensive property including: reliability, availability and maintainability (RAM). The RAM method is used in the varied industrial sectors, including aerospace, defense, energy, processing and transport [18, 34, 40]. For example, in the paper [18] the authors describe the possibilities of using the RAM model in the industrial practice to identify a critical equipment in terms of the frequent failures or high maintenance requirements. The paper [34] presents the use of the RAM analysis as a helpful tool in the design of systems, in the implementation of structural changes to achieve a minimum quantity of failures and increase the mean time between failures. The research [40] studies a relationship between the readiness, maintainability of the means of rail transport and the costs of unscheduled downtimes and vehicle maintenance. A model of the optimal implementation of the periodic inspections and repairs was conceptualized from a point of view of the costs, taking into account the current data on the vehicle failures.

The reliability indices are also used to analyze the exploitation costs of vehicles, while selecting a utility vehicle (commercial vehicle or light commercial vehicle) by the entrepreneur, have a major impact, apart from the model, vehicle make, purchase price and exploitation decline in value. In the vehicle exploitation, the planned costs of: repairs and technical maintenance, employees of maintenance and repair station personnel, parts and assemblies, fuel and oil, as well as other fluids and consumables [7, 10, 11] and the so-called “incidental costs”: unplanned downtimes resulting from a potential loss of trust of the ordering party, the need to provide replacement vehicles [29], changes in the allocation of the vehicles to the specific operations [26].

Another condition, brought up in the source literature, related to the selection of the vehicles in a company that has been gaining in importance and popularity within the recent years, is the environmental safety of the transport fleet, which is determined by the various factors, such as, for example: vehicles in a bad technical condition and emitting an increased amount of toxic compounds into the atmosphere [3, 9], an implementation of technologies (catalytic reactor, DPF and electronic injection, EGR system) reducing an impact of toxic exhaust gases on the environment [27], properly selected engine oil of low viscosity, which increases engine efficiency and simultaneously reduces a fuel consumption [14] and an oil contamination with fuel which may condense on the walls of the cylinder liner and flow into the oil sump reducing oil viscosity and carbon black resulting from an operation of the EGR system [1, 22]. As a result of this phenomenon, the chemical and physical properties of the engine oil deteriorate, which may result in the accelerated wear of the engine parts and its higher failure frequency. The elongated maintenance intervals in respect of the internal combustion engines extended by the manufacturers, aimed at reducing the vehicle exploitation costs and the environmental impact of the used oil are conducive to this process. At the present moment, a low fuel consumption and a compliance with the exhaust emission standards is a key exploitation factor in the vehicle selection for the company fleet.

Many national and international publications describe a problem of optimizing an allocation of the vehicles to their tasks. This solution should take into account the time of the individual tasks as this time

determines the risk of the vehicle failure. A typical multi-vehicle task allocation scenario consists in a deployment of the fleet of vehicles to visit a group of the target locations while minimizing the certain objectives such as: the total cost of traveling of the vehicles with a limited loading capacity to deliver the order to the several customers and return to the depot [5], the fuel consumption costs, cargo storage costs, cargo purchase costs and lead times depending on a cargo weight and a vehicle loading capacity [20], the maintenance costs including preventive maintenance of the heavy vehicles travelling long distances in the rather difficult road conditions [23], the costs and lead times of the cargo carriage including technical and economic factors [41], the total cost per unit of a fuel consumption fulfilling the requirements of three interdependent factors: transport, maintenance and environment [44], the total travel time of a fleet of the distributed vehicles that has to visit multiple destinations while encountering obstacles [3], the time of assigning planes to apron parking lots upon their landing [15], the total distance travelled by the robots [50], the maximum time for the robots to perform their tasks [12] or finding a maximum quantity of persons assigned to the vehicles in the flooded areas [51], the costs of a non-compliance with the service schedule [45], the time of waste collection, the duration of route allocation and the use of the resources [21]. The solution to the transport task may be to calculate the optimal route taking into account the exploitation parameters [47].

However, most of these researches focus primarily on how to effectively assign tasks to the vehicles, assuming that these vehicles can move freely between any two chosen locations where there is no external interference.

According to the author of the paper [35], the modern-day vehicles, which are equipped with on-board telematic systems connecting electronic devices between them, also have a specific exploitation potential, which can be used to improve a value of the specific reliability and exploitation indices of the vehicles. Transport telematics is the application of information, telecommunication, automatic, measurement and transport management technologies to achieve the best possible results in the trafficability performance. Telematics systems include the Internet, mobile networks, radio communication systems, geographical and road databases, satellite navigation systems, weather and traffic monitoring devices, radars and street cameras.

An improvement in the vehicle reliability and exploitation indices can be achieved by the use of the wireless technologies used in the vehicle-to-vehicle or vehicle-to-infrastructure communication. For example, in [2] a model was developed for the purpose of evaluating an impact of the "vehicle-to-vehicle" (V2V) and the "vehicle-to-infrastructure" (V2I) technology for the travel time and its reliability and for the scenario of a motorway work zone, which causes an increase in the travel time, queuing length and fuel consumption. The authors of the paper [19] presented the thesis that "vehicle-to-vehicle" communication (V2V) while using the short-range communication technology (DSRC) can help dramatically reduce a quantity of the vehicle collisions. Whereas, the paper [49] discusses the VANETs system, which is part of the intelligent transport system (ITS), using the "vehicle-to-vehicle" (V2V) and "vehicle-to-infrastructure" (V2I) communication to obtain information on the traffic and vehicle condition so that the road accidents can be prevented. The basic problems of the cooperating vehicle-to-infrastructure system (CVIS) were analyzed in the paper [52]. It discusses how the vehicle and network communication influences the benefits received from the common calculations in the CVIS and what should be an optimal cooperation in terms of reliability. Thus, the vehicles equipped with such technologies are better perceived by the purchaser and will be more willingly proposed to equip the fleet of the company.

The studies on evaluating the exploitation reliability of both inter-city and urban rail transport contained in the available literature are aimed at: a reduction of the exploitation costs and an improvement of the maintenance efficiency, a flexible management of exploitation and ensuring safety of the urban railway system [6], a development of a model of the railway vehicle maintenance strategy based on the

reliability indices - mean time to first failure, mean time between failures, mean time to recovery [16], an elaboration of a model for the short-term prediction of train failures taking into account the intensity of failures and a configuration of the spare parts [24].

It should be highlighted that the source literature does not often include the comprehensive studies describing the methods of the optimal selection of rolling stock (means of transport) in the transport company considering their operational reliability.

Therefore, there is a need to apply the comprehensive methods that enable to select the means of transport in the transport company in the optimal manner, taking into account their operational reliability. In the available literature, such researches are not easily available, hence there is a need to supplement this knowledge in terms of the reliability-based approach to solving the problem of selecting and optimizing the structure of the means of road transport in the company.

The commercial vehicles are essential for an existence of the transport company on the market. Therefore, the fleet should consist of the vehicles with the lowest failure frequency, which has a positive impact on their operational reliability.

The commercial vehicle is an intricate technical system which exploitation reliability is defined as a characteristic of its ability to be used or handled under the certain operating conditions. The evaluation of the operational reliability of the commercial vehicle is enabled by the various measures, including quantitative measures in the form of the varied numerical indices of the statistical nature. The availability of information on a course of the performance of the commercial vehicle in a concerned transport company is a key condition for the daily use of such indices [6].

The analysis carried out on the problem of the exploitation reliability of the vehicles shows that they usually involve a minimization of the costs linked to their use and operation, and ensuring the maximum reliability of the transport system. The impact of the vehicle use on the natural environment and the resulting safety and reliability aspect is also subject to the analyses. From the standpoint of authors of many studies an improvement of the reliability and exploitation indices of the vehicles can be achieved using the wireless technologies used in the vehicle communication and the telematic equipment of the on-board systems. The evaluation of the exploitation intensity of the vehicle is conducted while considering, inter alia, the mileage value, engine capacity, vehicle age, repair costs, revenues, technical readiness and intensity of the vehicle operation. It is not very often that the studies analyze and provide models for evaluating the operational reliability of the commercial vehicles in the transport company. This evaluation is of particular importance when selecting vehicles for the company.

This publication is an attempt to present the model that enables a manufacturer of the means of road transport, in particular light commercial vehicles, to select a vehicle manufacturer for the transport company while regarding its operational reliability.

### 3. General systematics of the exploitation reliability of the vehicle

The exploitation of a vehicle comprises a set of the intentional, interconnected organizational and technical operations from the moment the vehicle is taken into operation in order to use it optimally and according to its intended use until its totaling.

The optimal use of the vehicle is conditioned by many of its properties, among which the exploitation reliability expressed as a set of properties that describe the readiness for use and correctness of operation, durability, repairability and maintainability of the vehicle is of great importance. It is a measure of an ability of the vehicle to perform its functions efficiently, if needed and for a duration of the required time or mileage and under the specific environmental conditions.

In the exploitation practice, the readiness for use of the vehicle is treated as a fraction of a period of its usability, during which the vehicle performs or is able to perform its utility functions.

The correctness of operation is characterized by an ability of the vehicle to perform its work without any failure, and its durability - to maintain its usability until the vehicle reaches the limit state which means that it is technically impossible or economically inexpedient to continue its exploitation. The reparability, on the other hand, should be understood as a characteristic of the vehicle, which means its adaptation to the recovery of its condition by removing malfunctions or failures through maintenance and repair. The maintainability is a property of the vehicle which consists in maintaining the determined values of its exploitation parameters during its non-use [17].

The exploitation reliability of each vehicle is subject to the continuous and unfavorable changes due to an occurrence of operational wear and tear phenomenon. Having a knowledge of this reliability largely enables efficient and effective management in the company.

For the purposes of evaluating the exploitation reliability of the vehicles in the company the specifications, known as the reliability indices, which enable to conduct a quantitative evaluation of the impact of the varied factors occurring within the process of exploitation of the individual groups of the vehicles on their reliability and the obtained operational effects [46, 48]. Furthermore, the reliability indices facilitate a correct organization of the use and maintenance processes, including planning of the spare parts procurement within the company.

It can be assumed that the specific exploitation performance of the vehicle that determine its exploitation reliability are, first of all, as follows:

- correctness of operation, characterized by its ability to perform work failure-free,
- reparability, characterized by its ability to recover its usability.

The aforementioned properties of the vehicle are verifiable and measurable during its exploitation. While determining their characteristics, the conditions under which the vehicle was used or operated are taken into account. Thus, reliability in the broader sense, i.e., the exploitation reliability or, in a narrower sense, the operational reliability of the selected vehicle or a specific group of the vehicles in the company can be determined.

The operational reliability of the commercial vehicle is nothing else than its property - its ability to be used under the certain operating conditions and for a certain period of time or a certain service live. The operational reliability of the vehicle expresses a confidence of the decision-maker, dispatcher or driver that the vehicle will be fit to commence a transport task and will maintain this usability to complete that task(s). The level of this confidence may depend on a type of the task, requirements of the decision maker, conditions of completion and other factors [17].

The reliability theory makes extensive use of statistical calculation methods, and a global or a targeted operational reliability of the concerned make of the commercial vehicle can be determined as a value of a specific quantitative measure of a statistical nature or a group of these measures.

In addition to the operational reliability of a technical system, other types of the technical system, such as maintenance or economic reliability, are distinguished in the source literature.

The maintenance reliability expresses a confidence of the decision-maker and/or operator that an object (component, device, system) can be operated (recovered) under the certain conditions and over a certain time. On the other hand, the economic reliability - that an object will perform a specific exploitation (operational) task at an economic profit.

This trust may be based on the experience gained from observing the functioning of a certain population of the objects of the same type, carried out under the same or similar operating conditions, frequently also on a theoretical analysis and a simulation of the exploitation processes [17].

#### 4. Evaluation criteria of the operational reliability of the vehicle

The studying and evaluation of the operational reliability of the commercial vehicles plays an important role in conveying information to the entity that operates them. Each entity that commissions in-service testing expects different information on the condition of its own fleet, so for each entity the reliability testing is usually targeted on the specific characteristics of its vehicles and is usually conducted under a different focus using the different reliability measures and indices or characteristics.

In this paper, the multi-criteria in-service reliability study of four groups (makes) of the commercial vehicles in the transport company existing on the modern market used three basic measures (quantitative indices), which are used in the literature to evaluate motor vehicles, rail vehicles, electric vehicles [17].

- mean mileage do to first failure  $L_{u,k}$  of the vehicles in the studied "k" group,
- mean failure intensity  $\lambda_{u,k}(l)$  of the vehicles in the studied "k" group,
- mean mileage between failures (breakdowns)  $L_{m,k}$  in the studied "k" group.

The studies concerned a determination of the operational reliability of these groups (makes) of the vehicles targeted on the frequency of an occurrence of failures (breakdowns) preventing periodically (temporarily) the use of these vehicles by the concerned company.

The indices of the operational reliability of the concerned transport means used in the course of the study are presented below:

- 1) Mean mileage  $L_{u,k}$  to first failure "u" of the vehicles in the studied „k" vehicle group/make is defined by the relation:

$$L_{u,k} = \frac{1}{N(0)} (L_1 + L_2 + \dots + L_{N(0)}) = \frac{1}{N(0)} \sum_{i=1}^{N(0)} L_i \quad [\text{km}], \quad (1)$$

where:  $N(0)$  – quantity of the vehicles studied in the concerned vehicle group/make,  $L_i$  – mileage to first failure of  $i$  – th vehicle in the concerned group (make).

- 2) Mean failure intensity  $\lambda_{u,k}(l)$  of the vehicles in the studied "k" vehicle group/make for their assumed mileage/service life  $l$  ( $l = 75,000$  km) is defined by the relation:

$$\lambda_{u,k}(l) = \frac{n(l)}{N(0) * L} \left[ \frac{\text{failures}}{\text{km}} \right] = \frac{l * n(l)}{N(0) * L} \left[ \frac{\text{failures}}{75,000 \text{ km}} \right], \quad (2)$$

where:  $n(l)$  – quantity of failures (breakdowns) of the individual vehicles for the assumed mileage/service life  $l$ ,  $L$  – total mileage of the studied "k" vehicle group (make) to their first and subsequent failures for the assumed service life/mileage  $l = 75,000$  km.

- 3) Mean mileage  $L_{m,k}$  between failures "m" (breakdowns) in the studied „k" vehicle group (make) is defined by the equation:

$$L_{m,k} = \frac{1}{\sum_{j=1}^{N(0)} \left( \sum_{i=1}^{n_j} n_{ij} - 1 \right)} \sum_{j=1}^{N(0)} \sum_{i=2}^{n_j} l_{ij} = \frac{1}{\sum_{j=1}^{N(0)} (n_j - 1)} \sum_{j=1}^{N(0)} \sum_{i=2}^{n_j} l_{ij} \quad [\text{km}], \quad (3)$$

where:  $n_{ij}$  –  $i$ -th failure in the  $j$  - th vehicle,  $l_{ij}$  – mileage between  $i$  - th, and preceding ( $i-1$ ) failure of  $j$  - th vehicle,  $n_j$  –  $j$  - th vehicle in the studied group of the vehicles,  $N(0)$  – a quantity of the vehicles studied in the concerned "k" vehicle group/make.

It should be emphasized the above listed utility reliability indices have different physical quantities, which would significantly impede their use and comparison during a targeted analysis of this reliability for the studied groups (makes) of the light commercial vehicles.

In order to enable a mutual comparison, the indices (1, 2, 3) adopted for the analysis were subjected to a standard normalization procedure, depriving them, inter alia, of a physical quantity and standardizing their order of values to the range [0; 1].

The standardized index of mean mileage to first failure  $L_{u,knorm}$  of the concerned "k" vehicle group/make was calculated according to the following relationship:

$$L_{u,knorm} = \frac{L_{u,k} - L_{u,min}}{L_{u,max} - L_{u,min}}, \quad (4)$$

where:  $L_{u,k}$  – mean mileage to first failure of the concerned "k" vehicle group (make),  $L_{u,min}$  – minimum mileage to the first failure of the vehicle required in the company was of 50,000 km,  $L_{u,max}$  – maximum value of mean mileage to first failure among the A, B, C and D vehicle makes considered within the study.

The quantitative condition for the data adopted for the calculations:

$$L_{u,min} \leq L_{u,k},$$

what means that a minimum mileage to the first failure of the vehicles required in the company with regard to the possessed fleet may not exceed a monitored/disclosed mean value for this mileage.

The standardized index of mean failure intensity  $\lambda_{k,norm}(75,000)$  at the beginning of exploitation for the mileage adopted in the company  $l = 75,000$  km was calculated with the use of the following relationship:

$$\lambda_{k,norm}(75,000) = 1 - \frac{\lambda_{u,k}(75,000) - \lambda_{u,n}}{\lambda_{u,max} - \lambda_{u,min}}, \quad (5)$$

where:  $\lambda_{u,k}(75,000)$  – mean failure intensity of the vehicles of the concerned make,  $\lambda_{u,min}$  – minimum value of mean failure intensity for the studied vehicle makes,  $\lambda_{u,max}$  – mean failure intensity of the vehicles required in the company should not exceed 0.5 failure/75,000km.

The quantitative condition for the data adopted for the calculations:

$$\lambda_{u,max} > \lambda_{u,min},$$

what means that a permissible, required by the company, mean failure intensity of the vehicles with regard to the possessed fleet cannot be less than monitored/disclosed value of the intensity.

The standardized index of mean mileage between failures  $L_{m,knorm}$  for the vehicles of the studied make was calculated from the following relationship:

$$L_{m,knorm} = \frac{L_{m,k} - L_{m,min}}{L_{m,max} - L_{m,min}}, \quad (6)$$

where:  $L_{m,k}$  – mean mileage between failures/breakdowns of the studied vehicle make,  $L_{m,min}$  – minimum mileage between failures required in the company was of 10,000 km,  $L_{m,max}$  – maximum value of mean mileage between failures among the considered vehicle makes.

The quantitative condition for the data adopted for the calculations:

$$L_{m,min} \leq L_{m,k},$$

what means that a minimum mileage between failures required in the company with regard to the possessed fleet cannot exceed a monitored/disclosed mean value for this mileage.

## 5. Algorithm for the evaluation of the operational reliability of the vehicles in the transport company

Unquestionably, an effective methodology for studying the reliability of the commercial vehicles in the transport company is the key to their utility evaluation and, following an appropriate decision on the selection (choice) of the most reliable vehicle makes for the fleet already in possessed.

For the need of the undertaken evaluation, an algorithm for evaluating the operational reliability of the commercial vehicles had been developed in the form of a diagram, which was then used for such an evaluation of the commercial vehicles in the transport company existing on the Polish market (Fig. 1). The study covered four makes (groups) of the light commercial vehicles – delivery vans. In each group there were vehicles of the same homologation category N1 used since their manufacture for several years under the similar conditions in the concerned transport company. The quantity of the vehicles in the individual groups (makes) varied. The monitoring in this study covered the years (2010-2015) of the use of these vehicles

The versatility of the above diagram consists, inter alia, in the fact that the mathematical models (indices) of the operational reliability of the individual means of transport used in any study, according to the range of information (data) available, and depending on a course of their exploitation – use, can be always freely changed. According to the following algorithm, the operational reliability indices were calculated for the studied vehicle makes in the concerned transport company.

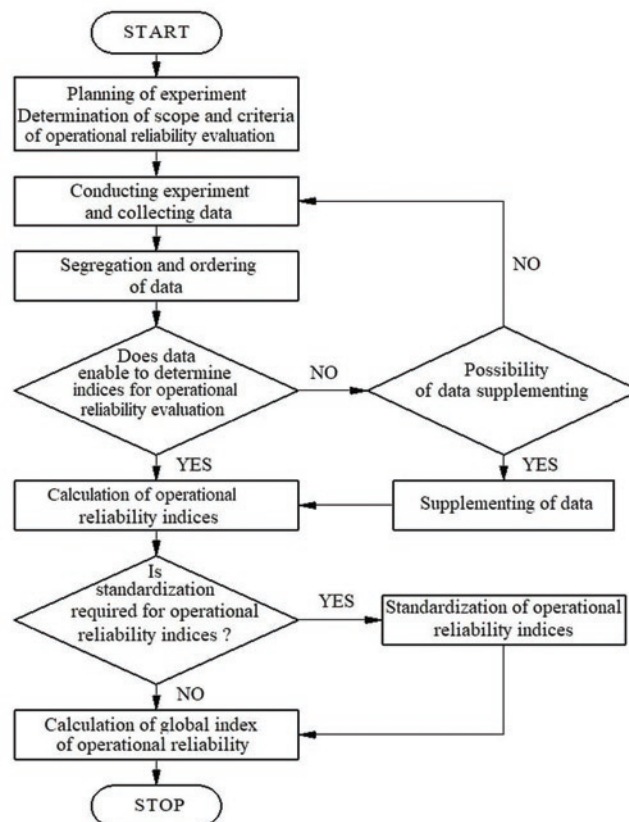


Fig. 1. Algorithm for evaluation of operational reliability of vehicles

## 6. Selection of the vehicle makes in the transport company on the basis of the evaluation of their operational reliability

It is common knowledge that the benefits of unifying the model/make of the manufacturer of the vehicle fleet in the specialized transport company can be great. They result, in particular, from an actual possibility of a significant reduction of the overall costs of servicing and repairing a uniform fleet compared to another company with the similar fleet, but including the varied models/makes. These benefits are without any doubts even greater if the company unifies its vehicle fleet to the model/make of the manufacturer that currently has the least failure frequency in the market in relation to other vehicle makes, which may result, inter alia, from a reduction in the overall costs of repairing such unified fleet and its greater operational reliability.

It was proposed to the specialized road transport company, currently operating on the national market, to unify the make of light commercial vehicles (category N1) from the four vehicle makes already used on the market (A, B, C and D) on the basis of the results of ob-

servicing their operational reliability, in particular the failure frequency or breakdown frequency that prevent them on a periodical basis from being used. As known, the benefits of such a procedure for the company are multiple, especially it enables, inter alia, to reasonably limit a quantity of the vehicle makes possessed to the one make, characterized by the vehicles with a relatively low failure frequency, which significantly reduces the costs of the conducted economic activity.

On the request of the company, the study was undertaken, during which data was collected containing targeted and detailed information on the exploitation of those vehicles, concerning their quantity in terms of featuring the concerned make of the manufacturer and a mileage to the first and subsequent failures (breakdown). A duration of the vehicle downtime in the service station for at least one day - 8 working hours was assumed to be the failure (breakdown). In the studied company, A make consisted of 19 vehicles, B make - 20, C make - 36, and D make - 40 vehicles.

Due to the large number of data collected during the concerned research, this study presents the selected monitoring data concerning only the failure frequency of A-make vehicles (Table 1 and 2).

Table 1. List of the A make vehicles along with their mileage to first failure

A make		Failure [km]						
Vehicle Number		1	2	3	4	5	6	7
A1	Mileage [km]	35 031	43 312	46 485	48 651			
A2		25 962						
A3		77 859						
A4		43 535						
A5		55 503						
A6		70 638						
A7		91 465						
A8		79 339	80 601	88 760				
A9		57 051	81 500					
A10		77 000	112 967					
A11		50 307	100 702	125 472	133 117	135 525	165 556	
A12		56 012	67 509	71 586	79 542			
A13		78 199						
A14		87 655						
A15		24 786	106 163	106 775				
A16		103 363						
A17		90 263						
A18		39 720	57 500	65 642	96 292	120 021	130 417	147 162
A19		45 068	63 996	71 063	72 939	103 074	129 011	

Table 2. Mileages between subsequent failures of the vehicles of A make [km]

A make	Failure number					
Vehicle number	2-1	3-2	4-3	5-4	6-5	7-6
A1	8 281	3 173	2 166			
A8	1 262	8 159				
A9	24 449					
A10	35 967					
A11	50 395	24 770	7 645	2 408	30 031	
A12	11 497	4 077	7 956			
A15	81 377	612				
A18	17 780	8 142	30 650	23 729	10 396	16 745
A19	18 928	7 067	1 876	30 135	25 937	

On the basis of the data presented in Table 1, the value of the index relating to the mean mileage to the first failure and the mean failure intensity to the mileage of 75,000 km of the vehicles of A make was determined. It can be easily perceived that only a few vehicles of this make features a great failure frequency. The data in this table were used to determine the mileage between the subsequent failures of the vehicles of this make (Table 2).

On the basis of the data presented in Table 2, a large dispersion of mileage between the subsequent failures of the vehicles of the A make is visible. The value of this dispersion results from an impact of the varied factors during the exploitation of these vehicles, in that also random factors.

The indices of the mean failure intensity of the vehicles featuring the individual makes were determined for the set-point value of 75,000 km of the service life/mileage. The results of the calculations are presented in Table 3.

Mean mileage of the vehicles of A make to first failure:

$$L_{u,A} = \frac{1}{N(0)} (L_1 + L_2 + \dots + L_{N(0)}) = \frac{1}{N(0)} \sum_{i=1}^{N(0)} L_i$$

$$L_{u,A} \cong 62,566 \text{ km.}$$

Maximum value of the mean mileage to the first failure had the vehicles of C make –  $L_{u,max} = 71,835$  km (Table 3).

Standardized index of the mean mileage to first failure of the vehicles of A make:

$$L_{u,Anorm} = \frac{L_{u,A} - L_{u,min}}{L_{u,max} - L_{u,min}}$$

$$L_{u,Anorm} \cong 0.576.$$

Mean failure intensity  $\lambda_{u,A}(75,000)$  of the vehicles of A make to the mileage of 75 000 km:

$$\lambda_{u,A}(75,000) = \frac{l * n(l)}{N(0) * L}$$

$$\lambda_{u,A}(75,000) \cong 0.141 \left[ \frac{\text{failures}}{75,000 \text{ km}} \right].$$

Standardized index of the mean failure intensity of the vehicles of A make:

$$\lambda_{u,Anorm}(75,000) = 1 - \frac{\lambda_{u,A}(75,000) - \lambda_{u,n}}{\lambda_{u,max} - \lambda_{u,min}}$$

$$\lambda_{u,Anorm}(75,000) \cong 0.799.$$

Mean mileage between failures of the vehicles of A make:

$$L_{m,A} = \frac{1}{\sum_{j=1}^{N(0)} \left( \sum_{i=1}^{n_j} n_{ij} - 1 \right)} \sum_{j=1}^{N(0)} \sum_{i=2}^{n_j} l_{ij} = \frac{1}{\sum_{j=1}^{N(0)} (n_j - 1)} \sum_{j=1}^{N(0)} \sum_{i=2}^{n_j} l_{ij} =$$

$$L_{m,A} \cong 17,700 \text{ km.}$$

Maximum value of the mean mileage between failures had the vehicles of B make –  $L_{m,max} = 18,470$  km (Table 3).

Standardized index of the mean mileage between failures of the vehicles of A make:

$$L_{m,Anorm} = \frac{L_{m,A} - L_{m,min}}{L_{m,max} - L_{m,min}}$$

$$L_{m,Anorm} \cong 0.909.$$

The results of calculations of the individual reliability indices for the vehicles of the studied makes are summarized in Table 3 and presented in Figures 2-7.

On the basis of Table 3 and Fig. 2, it is known that on average the highest mileage to the first failure (71,835 km) is characterized by the vehicles of C make. Whereas, the lowest mean mileage to the first failure (62,566 km) is recorded for the vehicles of the A make, and the relative difference between the two makes is relatively high, amounting to approximately 13%. Therefore, the vehicles of C make compared to the other makes have the lowest failure frequency in the initial period/first year of their use.

On the basis of Fig. 3, it is known that the vehicles of A and B makes, while compared to other makes in their first year of use, are characterized by the highest failure - breakdown intensity (frequency). The failure frequency of the vehicles of both A and B makes can be considered to be three times higher than the failure frequency of the other makes - C and D. This may mean that the vehicles of both A and B makes in the first year of their use failed on average three times more frequently than the vehicles of C and D makes.

On the basis of Fig. 4, it is known that the vehicles of A and B makes during the entire period of their monitoring (over the years under the study) had on average a 30% higher mileage between subsequent failures than the vehicles of C and D makes. The vehicles of all A, B, C and D makes fulfilled the requirement for the minimum

Table 3. Results of calculations of operational reliability indices for vehicles of k (A, B, C, D) makes

Make	Operational reliability indices					
	Mean mileage to first failure $L_{u,k}$ [km]	Standardized index of the mean mileage to first failure $L_{u,knorm}$	Mean failure intensity $\lambda_{u,k}$ [failure./75000 km]	Standardized index of the mean failure intensity $\lambda_{u,knorm}$	Mean mileage between failures $L_{m,k}$ [km]	Standardized index of the mean mileage between failures $L_{m,knorm}$
A	62 566	0.576	0.141	0.799	17 700	0.909
B	64 858	0.680	0.159	0.760	18 470	1.000
C	71 835	1.000	0.059	0.980	11 269	0.150
D	68 973	0.869	0.051	1.000	13 298	0.389



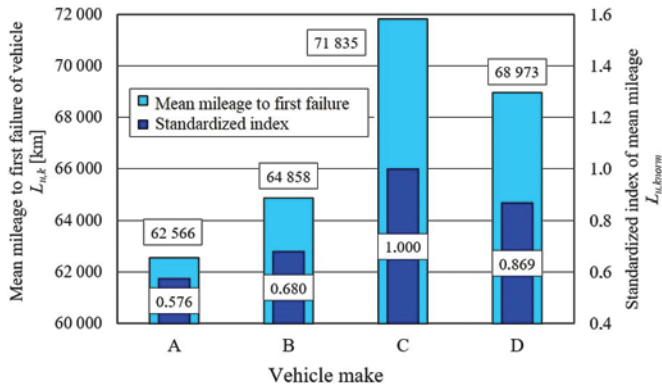


Fig. 2. Mean mileage of vehicles of a given make to first failure and the corresponding standardized mean mileage index

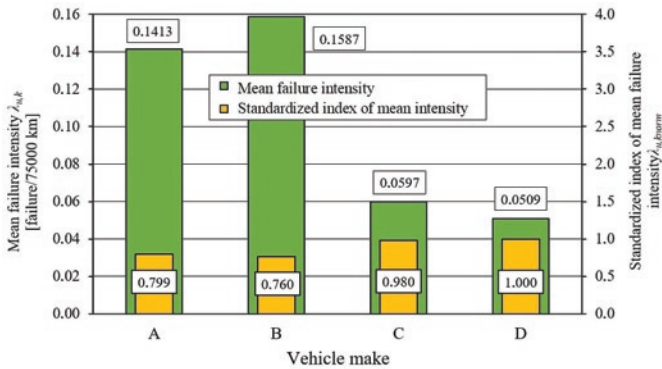


Fig. 3. Mean failure intensity of vehicles and corresponding standardized index of mean failure intensity

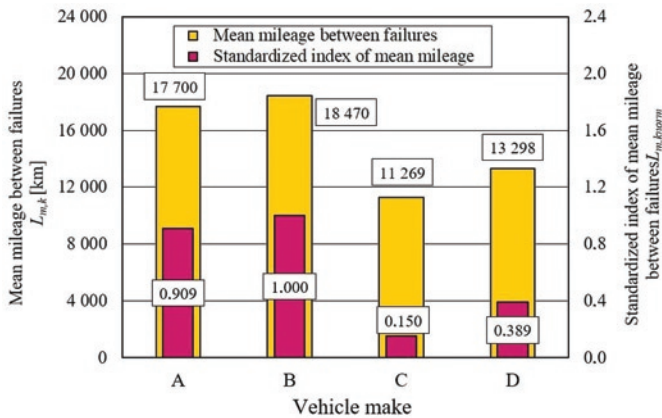


Fig. 4. Mean mileage of vehicles of given make between failures and corresponding standardized index of mean mileage between failures of vehicle

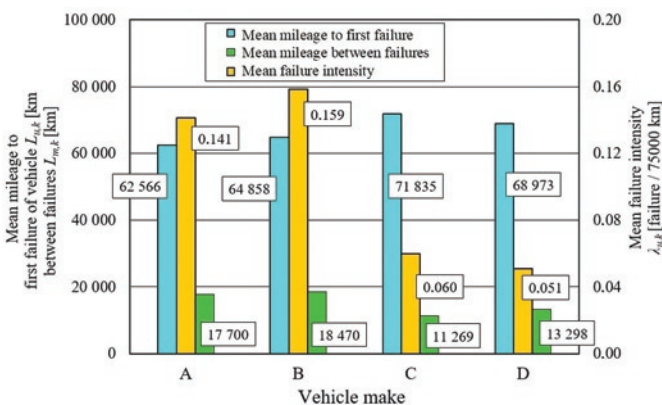


Fig. 5. Summary of reliability indices of vehicles of studied makes

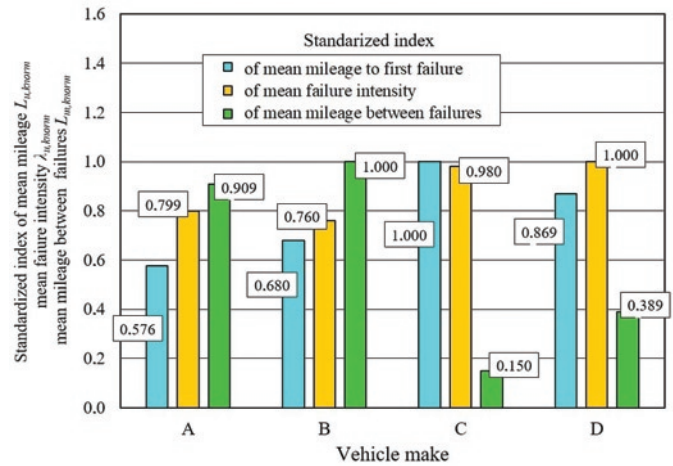


Fig. 6. Summary of standardized reliability indices for studied A, B, C and D makes of vehicles

value of  $L_{m,min} = 10,000$  km of mileage between subsequent failures/breakdowns applicable in the company.

On the basis of the results obtained above, the summary charts were drawn up, which present the results of the study of the concerned, targeted operational reliability of the vehicles of the individual makes in a form allowing for their mutual comparison (Fig. 5 and 6).

On the basis of the results of the study of the operational reliability of the vehicles of the individual makes in the company referred to herein above, it is concluded that the highest mean mileage to the first failure (71,835 km) is characteristic for the vehicles of C make. Whereas, the vehicles of D make have the lowest mean failure intensity (0.0509 failure/75,000 km), and the highest mean mileage between failures (18,470 km) is characteristic of the vehicles of B make.

Therefore, on the basis of the results obtained, it is not possible to directly answer a question as to which vehicles/makes have the highest operational reliability, targeted at their failure frequency. In order to facilitate the answer to the question posed as above, a simple in its design global index  $W_{G,k}$  of the studied operational reliability of the vehicles in the concerned "k" vehicle group/make in the following form:

$$W_{G,k}(W_{i,k}) = \frac{\sum_{i=1}^n W_{i,k}}{n}, \quad (7)$$

where:  $W_{i,k}$  – standardized value of  $i$ -th index of the operational reliability for the studied „k” vehicle make,  $n$  – quantity of the reliability indices assumed while determining the concerned operational reliability ( $n = 3$ ).

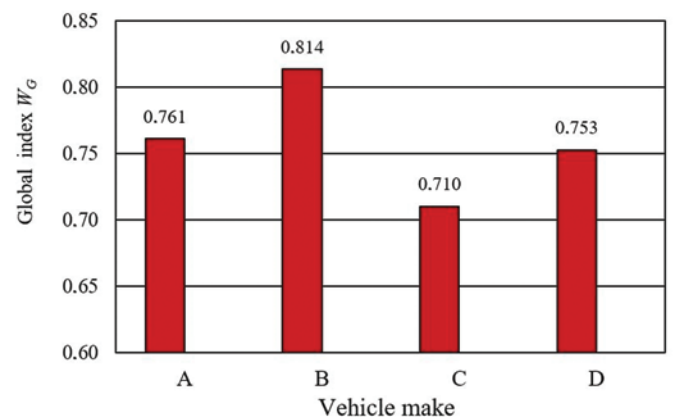


Fig. 7. Values of global index for studied A, B, C and D makes of vehicles

In the concerned study of the operational reliability, a maximum value of the target function  $W_{G,max}$  is searched:

$$W_{G,max}(W_{G,k}) = \max\{W_{G,k}\} = \max\{W_{G,A}; W_{G,B}; W_{G,C}; W_{G,D}\} \quad (8)$$

While using data from the calculations (Table 3) the value of the global index  $W_{G,A}$  for the vehicles of A make is as follows:

$$W_{G,A} = \frac{\sum_{i=1}^n W_{i,A}}{n} = \frac{L_{u,Anorm} + \lambda_{u,Anorm} + L_{m,Anorm}}{3},$$

$$W_{G,A} = 0.761.$$

The values of the global index  $W_{G,k}$  for the studied A, B, C and D makes of the vehicles calculated in the manner referred to in hereinabove are presented in Fig. 7.

The summary presented in Fig. 7 results that the highest value of the global index ( $W_{G,B} = 0.814$ ) among the analyzed A, B, C and D makes of the vehicles had the vehicles of B make, which makes it possible to conclude that they have the least failure frequency among the makes subject to the analysis within the assumed range of the utility reliability. The vehicles of B make fulfil the formal requirements of the researched company as regards the required mileage to the first failure of 64,858/50,000 km, the required mileage between failures of 18,470/10,000 km and the required failure intensity of 0.159/0.500 [failure/75,000 km] much better than the vehicles of the other three makes.

The lowest value of the global index ( $W_{G,C} = 0.710$ ) had the vehicles of C make, which was determined by their high failure frequency expressed as a small mean mileage of 11,269 km between the subsequent failures.

## 7. Summary and final conclusions

The research method presented in this article enables a road transport company to obtain, in a relatively simple manner, a targeted evaluation focused on the failure frequency of the operational reliability in respect of the commercial vehicles possessed by it, often of the varied makes of the manufacturer. Acquiring this evaluation requires prior monitoring of the use process of these vehicles and the application of a simple and short research procedure based on the use of classical/stationary and simple quantitative measures/indices. These indices are available in the basic and source literature on the theory of the vehicle operation.

The authors proposed the research method that allows for a multi-criteria evaluation of the operational reliability of the commercial vehicles studied due to their failure frequency, taking into account the following three evaluation criteria: mileage to the first failure, mileage between the subsequent failure and failure intensity. On the basis of this information, the proposed research method enables to obtain the targeted evaluation focused on the failure frequency of the operational reliability in respect of the vehicles and to identify the makes of the vehicles in the company, which had the least failure frequency in the monitored/set-point period of time. Thus, the vehicles of this make generated lower exploitation costs, which is what each owner of such a fleet of the vehicle strives for.

According to the authors, the proposed research method is effective, efficient and easy to apply in any transport company. The construction of this method is classical, typical for this category of the research issues. The indices used in this method for the evaluation of the operational reliability targeted at the failure frequency of the vehicles are simple and easy to use in each transport company.

It may turn out that, when a new generation/new model of the commercial vehicle is launched on the market, its exploitation reliability, compared to the former model, declared by the manufacturer or presented by the first users will be evaluated as overall good or more favorable than the previous model.

From the point of view of the authors, in such a case, it is up to the concerned company to take a decision about purchasing and adding to its fleet a new generation/new model of the vehicle that is currently being launched on the market. Each new model of the vehicle marketed in relation to the one already in use by the transport company will undoubtedly be more modern, which is primarily determined, as it is well known, by equipping the engine and the vehicle with more modern devices reducing, for example, a fuel consumption, extending maintenance intervals or ensuring a compliance with the next Euro standard in terms of the exhaust gas purity. This is probably what the purchasers of the new vehicle models will be guided by in case they are not aware more accurately or in more details of the operational reliability of the vehicle.

In the event of taking a purchase decision, the company has the opportunity to monitor the process of use of the newly purchased vehicle model on a periodical basis, upon which it can make a subsequent evaluation/re-evaluation in terms of the use of the currently possessed transport fleet using the algorithm presented in this paper. Depending on the result of the evaluation referred to hereinabove, the company may decide whether it is substantiated to unify the make or model of the light commercial vehicles possessed.

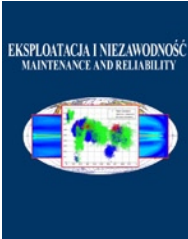
## 9. Conclusions

- 1) The extensive analyses of the exploitation process of the vehicles usually refer to a minimization of the costs linked to their use and ensuring the maximum reliability of the transport system, as well as an impact of the vehicle use on the environment or the safety aspect. However, the evaluation of the exploitation intensity of the vehicle is conducted taking into account, inter alia, a value of mileage, engine capacity, vehicle age, repair costs, revenues, technical readiness, and intensity of the use of the motor vehicle.
- 2) The commercial vehicles in the transport company constitute unquestionably the basis of its existence. They are sometimes used under the conditions which often lead to their failure. The knowledge of the failure frequency can be easily used for the purpose of evaluating the operational reliability of the vehicles in use.
- 3) It is substantiated to use the proposed in the article, the effective and efficient evaluation method to evaluate the operational reliability of the commercial vehicles featuring the varied makes used in the transport companies.
- 4) The application of the proposed evaluation method requires prior monitoring of the process of use of these vehicles and the application of the evaluation based on the use of classical quantitative measures/indices.
- 5) In the researched transport company, it is justified to unify the fleet of the light commercial vehicles of B make. The aforementioned is confirmed by the results of a targeted and multi-criteria analysis of their operational reliability obtained in the concerned study. The vehicles of B brand are characterized by the relatively low failure frequency in the first year of their use. In comparison with other makes, they are characterized by a relatively high mileage to the first failure (64,858 km on average) and the highest mileage between the subsequent failures/breakdowns (18,470 km on average) in the successive years of their use.

## References

1. Abdulqadir L B, Mohd Nor N F, Lewis R, Slatter T. Contemporary challenges of soot build-up in IC engine and their tribological implications. *Tribology - Materials, Surfaces & Interfaces* 2018; 12(3): 115-29, <https://doi.org/10.1080/17515831.2018.1464256>.
2. Abdulsattar H, Mostafizi A, Siam M R K, Wang H. Measuring the impacts of connected vehicles on travel time reliability in a work zone environment: an agent-based approach. *Journal of Intelligent Transportation Systems* 2020; 24(5): 421-436, <https://doi.org/10.1080/15472450.2019.1573351>.
3. Bai X, Yan W, Cao M, Xue D. Distributed multi-vehicle task assignment in a time-invariant drift field with obstacles. *IET Control Theory Appl.* 2019; 13(17): 2886-2893, <https://doi.org/10.1049/iet-cta.2018.6125>.
4. Baptista Ventura L M, Oliveira Pinto F, Gioda A, Almeida D'Agosto M. Inspection and maintenance programs for in-service vehicles: an important air pollution control tool. *Sustainable Cities and Society* 2019; 101956, <https://doi.org/10.1016/j.scs.2019.101956>.
5. Batsyn M V, Batsyna E K, Bychkov I S, Pardalos P M. Vehicle assignment in site-dependent vehicle routing problems with split deliveries. *Operational Research* 2019, <https://doi.org/10.1007/s12351-019-00471-7>.
6. Borucka A. Logistic regression in modeling and assessment of transport services. *Open Engineering*, 2020; 10: 1-9, <https://doi.org/10.1515/eng-2020-0029>.
7. Caban J, Drożdźiel P, Krzywonos L, Rybicka I, Šarkan B, Ján Vrábek J. Statistical Analyses of Selected Maintenance Parameters of Vehicles of Road Transport Companies. *Advances in Science and Technology Research Journal* 2019; 13(1): 1-13, <https://doi.org/10.12913/22998624/92106>.
8. Cheng R, Lan R, Liao Z, Qin Y. Research on the Flexible Operation and Maintenance Management of Urban Rail Vehicle. *Proceedings of the 4th International Conference on Electrical and Information Technologies for Rail Transportation (EITRT) 2019*: 705-718, [https://doi.org/10.1007/978-981-15-2914-6\\_67](https://doi.org/10.1007/978-981-15-2914-6_67).
9. Chłopek Z, Bebkiewicz K. Model of the structure of motor vehicles for the criterion of the technical level on account of pollutant emission. *Eksploracja i Niezawodność - Maintenance and Reliability* 2017; 19 (4): 501-507, <https://doi.org/10.17531/ein.2017.4.2>.
10. Drożdźiel P, Komsta H, Krzywonos L. An analysis of costs of vehicle repairs in a transportation company. Part I. *Transport Problems* 2012; 7(3): 67-75.
11. Drożdźiel P, Komsta H, Krzywonos L. An analysis of the relationships among selected operating and maintenance parameters of vehicles used in a transportation company. *Transport Problems* 2011; 6 (4): 93-99.
12. Franceschelli M, Rosa D, Seatzu C, Bullo F. Gossip algorithms for heterogeneous multi-vehicle routing problems. *Nonlinear Anal. Hybrid* 2013; 10: 156-174, <https://doi.org/10.1016/j.nahs.2013.03.001>.
13. Gandoman F H, Ahmadi A, Bossche P Van den, Mierlo J Van, Omar N, Nezhad A E, Mavalizadeh H, Mayet C. Status and Future Perspectives of Reliability Assessment for Electric Vehicles. *Reliability Engineering & System Safety* 2019; 183: 1-16, <https://doi.org/10.1016/j.res.2018.11.013>.
14. Golebiowski W, Wolak A, Zajac G. The influence of the presence of a diesel particulate filter (DPF) on the physical and chemical properties as well as the degree of concentration of trace elements in used engine oils. *Petroleum Science and Technology* 2019; 37(7): 746-755, <https://doi.org/10.1080/10916466.2018.1539751>.
15. Gołda P, Izdebski M, Szczepański E. The Application of Ant Algorithm in the Assignment Problem of Aircrafts to Stops Points on the Apron. *Journal of KONES* 2018; 25(4): 479-488.
16. Graboń-Chalupczak M. Reliability as an element of process in managing a fleet of railway vehicles. *Journal of KONBiN* 2019; 49(4): 265-280, <https://doi.org/10.2478/jok-2019-0086>.
17. Hebda M, Janicki D. Trwałość i niezawodność samochodów w eksploatacji (Durability and reliability of vehicles in exploitation). WKŁ, Warsaw 1977.
18. Herder P M, van Luijk J A, Bruijnooge J. Industrial application of RAM modeling development and implementation of a RAM simulation model for the Lexan plant at GE Industrial, Plastics. *Reliability Engineering & System Safety* 2008; 93: 501-508, <https://doi.org/10.1016/j.res.2006.10.019>.
19. Hoque MA, Rios-Torres J, Arvin R, Khattak A, Ahmed S. The extent of reliability for vehicle-to-vehicle communication in safety critical applications: an experimental study. *Journal of Intelligent Transportation Systems* 2020; 24(3): 264-278, <https://doi.org/10.1080/15472450.2020.1721289>.
20. Jacyna M, Izdebski M, Szczepański E, Gołda P. The task assignment of vehicles for a production company. *Symmetry* 2018; 11(10): 1-19, <https://doi.org/10.3390/sym10110551>.
21. Jacyna-Gołda I, Izdebski M, Podvieszko A. Assessment of Efficiency of Assignment of Vehicles to Tasks In Supply Chains: A Case Study of A Municipal Company. *Transport*. 2017; 2(3): 243-251, <https://doi.org/10.3846/16484142.2016.1275040>.
22. Jang J, Lee Y.-J, Kwon O, Lee M, Kim J. The Effect of Engine Oil on Particulate Matter, Emissions and Fuel Economy in Gasoline and Diesel Vehicle. *SAE Technical Paper Series* 2014, <https://doi.org/10.4271/2014-01-2837>.
23. Jbili S, Chelbi A, Radhoui M, Kessentini M. Integrated strategy of Vehicle Routing and Maintenance. *Reliability Engineering & System Safety* 2018; 170: 202-214, <https://doi.org/10.1016/j.res.2017.09.030>.
24. Kang R, Wang J, Cheng J, Chen J, Pang Y. Intelligent forecasting of automatic train protection system failure frequency in China high-speed railway. *Eksploracja i Niezawodność - Maintenance and Reliability* 2019; 21(4): 567-576, <https://doi.org/10.17531/ein.2019.4.5>.
25. Kozłowski E., Borucka A., Świdorski A. Application of the logistic regression for determining transition probability matrix of operating states in the transport systems. *Eksploracja i Niezawodność - Maintenance and Reliability* 2020; 22(2): 192-200, <https://doi.org/10.17531/ein.2020.2.2>.
26. Kristjanpoller F, Viveros P, Zio E, Pascual R, Aranda O. Equivalent availability index for the performance measurement of haul truck fleets. *Eksploracja i Niezawodność - Maintenance and Reliability* 2020; 22(4): 583-591, <https://doi.org/10.17531/ein.2020.4.1>.
27. Macián V, Tormos B, Bastidas S, Pérez T. Improved fleet operation and maintenance through the use of low viscosity engine oils: fuel economy and oil performance. *Eksploracja i Niezawodność - Maintenance and Reliability* 2020; 22(2): 201-211, <https://doi.org/10.17531/ein.2020.2.3>.
28. Nallusamy S, Balakannan K, Chakraborty PS, Majumdar G. Reliability Analysis of Passenger Transport Vehicles in Public Sector Undertaking. *International Journal of Applied Engineering Research* 2015; 10(68): 843-850.

29. Niewczas A, Móravski Ł, Dębicka E, Borucka A. Commercial vehicles incapacity risk analysis in the transport company. *Journal of Konbin* 2020; 49(4): 321-340, <https://doi.org/10.2478/jok-2019-0089>.
30. Niewczas A, Rymarz J, Dębicka E. Stages of operating vehicles with respect to operational efficiency using city buses as an example. *Eksploracja i Niezawodność - Maintenance and Reliability* 2019; 21 (1): 21-27, <https://doi.org/10.17531/ein.2019.1.3>.
31. Niewczas A., Iganciuk P., Dmowski A.: Analiza uszkodzeń i kosztów naprawy wybranych układów konstrukcyjnych samochodów dostawczych. (Analysis of failure and repair costs of selected structural systems of light commercial vehicles). *Autobusy: technika, eksploatacja, systemy transportowe* 2016; 11: 122-125.
32. Niewczas A, Rymarz J, Pieniak D. Porównawcza analiza niezawodności autobusów komunikacji miejskiej (Comparative analysis of reliability of public transport buses). *Prace Naukowe Politechniki Warszawskiej. Transport* 2014; 101: 91-101.
33. Repin S, Evtukov S, Maksimov S. A method for quantitative assessment of vehicle reliability impact on road safety. *Transportation Research Procedia* 2018; 36: 661-668, <https://doi.org/10.1016/j.trpro.2018.12.128>.
34. Sharma R K, Kumar D, Kumar P. Performance modeling in critical engineering systems using RAM analysis. *Reliability Engineering & System Safety* 2008; 93: 891-897, <https://doi.org/10.1016/j.res.2007.03.039>.
35. Siergiejczyk M, Rosiński A. Analysis of the operation process of vehicles equipped with telematic on-board systems. *Journal of Konbin* 2020; 49(4): 463-479, <https://doi.org/10.2478/jok-2019-0096>.
36. Ślęzak M, Niewczas A, Ignaciuk P, Dmowski A. Comparative Analysis of Reliability and Maintenance Costs of Commercial vehicles of Selected Makes Used Under Long-Term Rental. 11th International Scientific and Technical Conference on Automotive Safety Location: Casta Papiernicka, Slovakia 18-20, 2018, <https://doi.org/10.1109/AUTOSAFE.2018.8373348>.
37. Świdorski A, Borucka A, Jacyna-Golda I, Szczepański E. Wear of brake system components in various operating conditions of vehicle in the transport company, *Eksploracja i Niezawodność - Maintenance and Reliability*, 2019; 1(29), 1-9, <https://doi.org/10.17531/ein.2019.1.1>.
38. Szkoda M. Assessment of reliability, availability and maintainability of rail gauge change systems. *Eksploracja i Niezawodność - Maintenance and Reliability* 2014; 16 (3): 422-432.
39. Szkoda M. Wskaźniki niezawodności środków transportu szynowego (Reliability indices of means of rail transport). *Logistyka* 2012; 3: 2195-2202.
40. ten Wolde M, Ghobbar A A. Optimizing inspection intervals - Reliability and availability in terms of a cost model: A case study on railway carriers. *Reliability Engineering & System Safety* 2013; 114: 137-147, <https://doi.org/10.1016/j.res.2012.12.013>.
41. Tkaczyk S. Dobór środków transportu do realizacji procesów technologicznych (Selection of means of transport for implementation of technological processes). *Logistyka*, 2015; 4: 1045-1056.
42. Tkaczyk S. The method for selection and combining the means of transportation according to the Euro standards. *Combustion Engines*. 2015; 162(3). 958-962.
43. Vayenas N, Wu X. Maintenance and reliability analysis of a fleet of load-haul-dump vehicles in an underground hard rock mine. *International Journal of Mining, Reclamation and Environment* 2009; 23(3): 227-238, <https://doi.org/10.1080/17480930902916494>.
44. Vujanovic D. B, Momcilovic VM, Medar O. M. Influence of an integrated maintenance management on the vehicle fleet energy efficiency. *Thermal Science* 2018; 22(3): 1525-1536, <https://doi.org/10.2298/TSCI170209122V>.
45. Wang Y, Limmer S, Olhofer M, Emmerich MTM, Thomas Bäck T. Vehicle Fleet Maintenance Scheduling Optimization by Multi-objective Evolutionary Algorithms. 2019 IEEE Congress on Evolutionary Computation (CEC), Wellington, New Zealand, 10-13 June 2019; 442-449, <https://doi.org/10.1109/CEC.2019.8790142>.
46. Williams E.J, Messina A. Applications of the Multiple Failure Location Assurance Criterion, *Proceedings of the International Conference on Failure Assessment of Structures (DAMAS 99)*, Dublin, Ireland, 1999: 256-264, <https://doi.org/10.4028/www.scientific.net/KEM.167-168.256>.
47. Wojciechowski Ł, Cisowski T. Metoda optymalizacji procesu eksploatacji w ujęciu bezawaryjności pojazdu. *Wydawnictwo Politechniki Lubelskiej*, Lublin 2018.
48. Worden K, Manson G, Wardle R, Staszewski W, Allman D. *Experimental Validation of Two Structural Health Monitoring Methods*, Stanford University, Palo Alto, CA, 1999; 784-799.
49. Xu H, Zeng M, Hu W, Wang J. Authentication-Based Vehicle-to-Vehicle Secure Communication for VANETs. *Mobile Information Systems* 2019, 7016460, <https://doi.org/10.1155/2019/7016460>.
50. Yu J, Chung S J, Voulgaris P G. Target assignment in robotic networks: distance optimality guarantees and hierarchical strategies, *IEEE Trans. Autom. Control* 2015; 60(2): 327-341, <https://doi.org/10.1109/TAC.2014.2344291>.
51. Yusoff M, Ariffin J, Mohamed A. Solving Vehicle Assignment Problem Using Evolutionary Computation. *Lecture Notes in Computer Science* 2010; 6145: 523-532, [https://doi.org/10.1007/978-3-642-13495-1\\_64](https://doi.org/10.1007/978-3-642-13495-1_64).
52. Zhou J, Tian D, Member S, Wang Y, Sheng Z, Duan X, Leung VCM. Reliability-Optimal Cooperative Communication and Computing in Connected Vehicle Systems. *IEEE Transactions on Mobile Computing* 2020; 19(5): 1216-1232, <https://doi.org/10.1109/TMC.2019.2907491>.



Article citation info:

Kumar V, Kumar G, Singh RK, Soni U. Degrading systems availability analysis: analytical semi-Markov approach. *Eksploracja i Niezawodność – Maintenance and Reliability* 2021; 23 (1): 195–208, <http://dx.doi.org/10.17531/ein.2021.1.20>.

## Degrading systems availability analysis: analytical semi-Markov approach

Varun Kumar<sup>a</sup>, Girish Kumar<sup>b</sup>, Rajesh Kumar Singh<sup>c</sup>, Umang Soni<sup>a</sup>

<sup>a</sup>Netaji Subhas Institute of Technology, Delhi, India

<sup>b</sup>Delhi Technological University, Delhi, India

<sup>c</sup>Management Development Institute, Gurugram, India

Indexed by:



### Highlights

- Semi-Markov Process (SMP) is used to model the probabilistic behavior of system.
- SMP takes into account the dependencies and interactions between components of the system.
- System model is developed which is solved using SMP to evaluate system availability.
- The analysis results show that the maintenance policy: perfect repair with opportunistic maintenance is more efficient.

### Abstract

This paper deals with modeling and analysis of complex mechanical systems that deteriorate with age. As systems age, the questions on their availability and reliability start to surface. The system is believed to suffer from internal stochastic degradation mechanism that is described as a gradual and continuous process of performance deterioration. Therefore, it becomes difficult for maintenance engineer to model such system. Semi-Markov approach is proposed to analyze the degradation of complex mechanical systems. It involves constructing states corresponding to the system functionality status and constructing kernel matrix between the states. The construction of the transition matrix takes the failure rate and repair rate into account. Once the steady-state probability of the embedded Markov chain is computed, one can compute the steady-state solution and finally, the system availability. System models based on perfect repair without opportunistic and with opportunistic maintenance have been developed and the benefits of opportunistic maintenance are quantified in terms of increased system availability. The proposed methodology is demonstrated for a two-stage reciprocating air compressor with intercooler in between, system in series configuration.

### Keywords

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

availability, perfect repair, opportunistic maintenance, Embedded Markov Chain (EMC), Cumulative Density Function (CDF), Semi-Markov Process (SMP).

### Notations

$A$	Transition probability matrix of the system model	$O$	Operative
$A_s$	Steady-state system availability	$P_i$	Steady-state probability of state 'i' of SMP model
$D$	Degraded	$s_i$	Steady-state probability of state 'i' of Embedded Markov Chain
$F$	Failed	$T_i$	Mean sojourn time at state 'i' of the SMP model
$F_{ij}(t)$	Cumulative density function (CDF) of transition from state 'i' to 'j'.	$i, j$	Sequential number of states of the SMP model, where, $i, j = 1, 2, \dots, n$
$\bar{F}_{ij}(t)$	Complementary CDF of transition from state 'i' to 'j'.	$\beta_{ij}$	Shape parameter of Weibull distribution
$K(t)$	Kernel matrix of the semi-Markov model	$\theta_{ij}$	Scale parameter of Weibull distribution
$k_{ij}$	Transition probability from state 'i' to 'j'	$\mu_{ij}$	Exponential repair transition rate

(\*) Corresponding author.

E-mail addresses: V. Kumar - [varunkumar3793@yahoo.in](mailto:varunkumar3793@yahoo.in), G. Kumar - [girish.kumar154@gmail.com](mailto:girish.kumar154@gmail.com), R.K. Singh - [rajesh.singh@mdi.ac.in](mailto:rajesh.singh@mdi.ac.in), U. Soni - [umangsoni.1@gmail.com](mailto:umangsoni.1@gmail.com)

## 1. Introduction

Degradation modeling of mechanical systems has drawn special attention of plant engineers as it is a crucial aspect of the execution of an effective maintenance plan. Maintenance is the system's design feature that enables the success of various maintenance activities including inspection, repair, replacement and diagnosis [28]. Whenever a system breakdown occurs due to sub-system/component failure, it enters the repair phase. With the capability to repair or restore a system which is under breakdown, a failure-repair-failure-repair cycle is introduced. The restoration cycle can be broken down into a variety of subtasks that include supply delay, maintenance delay, diagnosis, replacement or repair, alignment and verification, etc. Therefore, repair work is a highly skilled and dynamic task in which workers come into close contact with the equipment. There are several factors which affect the maintenance procedure like part layout, ergonomic factors, maintenance skill levels, repair crew size, level of repair, and use and clarity of maintenance procedures and diagrams lay down in the manual. Although, the reliability of the components has improved dramatically, humans still remain messy and unpredictable [1]. There are several measures of maintainability but the most popular and the one discussed in this paper is mean time to repair (MTTR). The growing importance of maintenance has developed an increasing trend in designing and implementing optimal maintenance strategies to enhance system reliability, avoid system failures, and reduce system degradation maintenance cost [27].

The main objective of any maintenance regime is to maintain the system functionality to the maximum extent possible with optimum trade-offs between the down times and cost of maintenance, avoiding any catastrophic failures. Opportunistic maintenance works out to be the suitable remedy, which utilizes the opportunity of system shutdown or module dismantle to perform any maintenance required in the immediate future and saves a substantial amount of system downtime. A system of components working in a random environment is subjected to wear and damage over time and may fail unexpectedly. The components are replaced or repaired upon failure, and such unpleasant events of failure are at the same time also considered in practice as opportunities for maintenance on other components. Opportunistic maintenance basically refers to the scheme in which maintenance is carried out at opportunities, either by choice or based on the physical condition of the system. In this study, the focus is made on the situation in which the opportunities for maintenance are generated by the failure epochs of individual components. At each failure epoch, the failed components are correctively repaired and other components that are still operational are also serviced so that all the subsystems are maintained and restored to certain conditions. For example, when corrective repair on some components requires dismantling of the entire system, a corrective repair on these components combined with opportunistic repair on other or neighboring components might be worthwhile.

System performance evaluation remains a core feature in manufacturing industries as there is a need for efficient method of assessing the efficiency of modern production processes. Availability is one of the most important measures of system performance as it is directly related to the financial returns. It has a broader reach than reliability as it takes into account the measurement of maintenance times [16]. Availability gives the probability of the system being available when called to function at random. Complex mechanical systems work under high reliability and safety standards as such systems deteriorate through distinct mode resulting from different physical phenomenon or different failure attribute of individual subsystems/components with the ageing process [24] which is described by means of increasing failure rate. Certain systems may have lower failure rate, but their down time will be long, hence may interrupt the process at a higher rate as compared to the system with short down time having lower failure rate [22]. The downtime cost of such equipment that progressively degrade with the age, is high. So, it would be more effective to

consider multi-state degradation and to take appropriate maintenance actions upon failure of the system. Industries that rely on certain key performance measures have keen interest in being able to model complex mechanical systems and track the availability of such systems. System availability estimation is most frequently done through simulation. Software used for assessment of availability of complex mechanical system usually takes extra time to bring out results by virtue of simulation based technique. Therefore, analytical techniques which are quicker than simulation based techniques are utilized in this paper for system availability analysis. Due to mathematical complications, analytical techniques are hardly used for large and complex mechanical systems but still an attempt is being made to model such complex systems for system availability analysis.

Availability studies for degrading systems have been carried out by numerous researchers, but these are mainly based on Markov model using constant failure and repair rates [19], which is unrealistic in actual operating conditions. Markov model is a stochastic model which is used to model randomly changing systems over time. The basic assumption of a Markov Process is that the behavior of a system in each state is memory less which illustrates that the future evolution of the process depends only on the present state and not on the past sequence of traversed states prior to current state. Mathematically, the Markov property is stated as:

$$P(X_{t+1} = j | X_t = i, X_{t-1} = i_1, \dots, X_0 = i_0) = P(X_{t+1} = j | X_t = i) \quad (1)$$

However, it has two main limitations: with the complexity of the system, the number of state increases so rapidly that it can lead to state-explosion, restricting the solution to very complex systems [4]. For systems which are subjected to increasing or decreasing failure and repair rates, the Markov approach is inapplicable. Semi-Markov models seem to address this which considers variable failure or repair rates. Therefore, SMP is used to model systems where future state depends on current state as well as sojourn times in this state, which can obey any distribution in turn, not necessarily exponential.

The SMP model however is difficult to approach analytically [30] and no general and detailed procedure for its solution is available. In this respect, several attempts were made to transform the SMP model to a Markov model by approximating non-exponential distributions to exponential distributions in order to get the solution [30, 31]. The SMP has been applied in the areas of software reliability [11, 26, and 32] degradation-dependent reliability [10], fault detection and isolation reliability [18], optimization for condition-based maintenance [3], equipment health management [5], etc. Some integrated methods such as SMP and Bayesian networks were also presented for assessing the availability of fault tolerant systems [21]. The steady-state solution of the SMP model was implemented by a two-stage analytical approach in the field of software reliability but was not applied to mechanical systems [12, 33]. In this article, this method has been extended for availability assessment of repairable mechanical systems. The proposed approach is capable of handling non-exponential distributions i.e. Weibull distribution, as such distribution depicts actual behavior of the systems undergoing degradation. The Weibull distribution has been used for modeling other real life applications such as the deterioration of mechanical systems such as pistons, engine crankshaft, and the breakdown of insulating fluid, etc. [25]. Therefore, to capture the dynamics of real system and to model the dependencies and interactions between components of the system, semi-Markov approach have been adopted which deal with events that are non-exponentially distributed. The primary objective of this paper is to develop a mathematical model for system availability assessment and to quantify the benefits of opportunistic maintenance in terms of increased system availability. The remaining part of the paper is organized as follows: Overview of the analytical semi-Markov approach is described in Section 2. Section 3 deals with the system modelling considering

multi-state degradation and its solution is illustrated in detail in Section 4. Section 5 concludes the work while the scope for future work is described in Section 6.

## 2. Research Methodology: Analytical Semi-Markov Approach

Semi-Markov approach is an extension of Markov process, which is used to model such systems that degrade non-exponentially. Semi-Markov process is more suitable to present the deterioration process of physical system than continuous time Markov chain in terms of the mathematical generality and tractability. Semi-Markov process (Semi-MP) is of two types: discrete and continuous time chain. Availability assessment problems are usually encountered in continuous time chain for the system. However, to avoid the complexity involved in the solution of continuous time chain model, it is converted into an Embedded Markov chain (EMC) which is a discrete time chain. Hence, EMC is an ideal process for modeling and analysis of degrading systems. The solution of the Semi-Markov model involves complex integration and hence more complicated than the Markov process. Semi-Markov process is characterized as an arrangement of two dimensional random variables,  $\{(X_k, T_k): k \in 1, 2, 3, 4, 5, \dots, m\}$ , with the properties mentioned below:

- $X_k$  represents the system state after  $k$  transitions in a discrete-time Markov chain (DTMC).
- $T_k$  represents the sojourn time i.e., the amount of time the system is expected to spend at a particular state [6].

The complete framework of implementing Semi-MP to model system availability is summarized using a two stage method. In stage 1, transition probability matrix of the EMC of SMP model is determined while in stage two, availability of the system is evaluated by using sojourn time and steady-state probability of each state of the SMP model.

The steps of the semi-Markov approach expressed in the flow chart given below:

Consider a system state space with  $m$  possible states,  $m$  being a finite natural number, which is represented as  $I = \{1, 2, 3, 4, 5, \dots, m\}$ . The state space is the set of all feasible states where each state represents a different configuration of system. On the basis of this, semi-Markov model is developed which can be conveniently represented by a labeled directed graph which consists of all feasible states connected by the transition lines showing the appropriate distribution. The semi-Markov process is decomposed into two stages which are discussed below:

In stage 1, the transition probability matrix  $A$  of the SMP model's embedded Markov chain (EMC) is evaluated and this matrix helps in determining steady-state solution,  $s_i$ , of all the feasible EMC states.

In stage 2, mean sojourn time  $T_i$ , has to be evaluated for each state. The equation expressed below helps in determining the steady-state solution,  $P_i$ , of all feasible states of the SMP model [13].

$$P_i = \frac{s_i T_i}{\sum_{i=1} s_i T_i}, i \in I \quad (2)$$

where,  $s_i$  is Steady state probability of the EMC and  $T_i$  depicts Sojourn time of each system state.

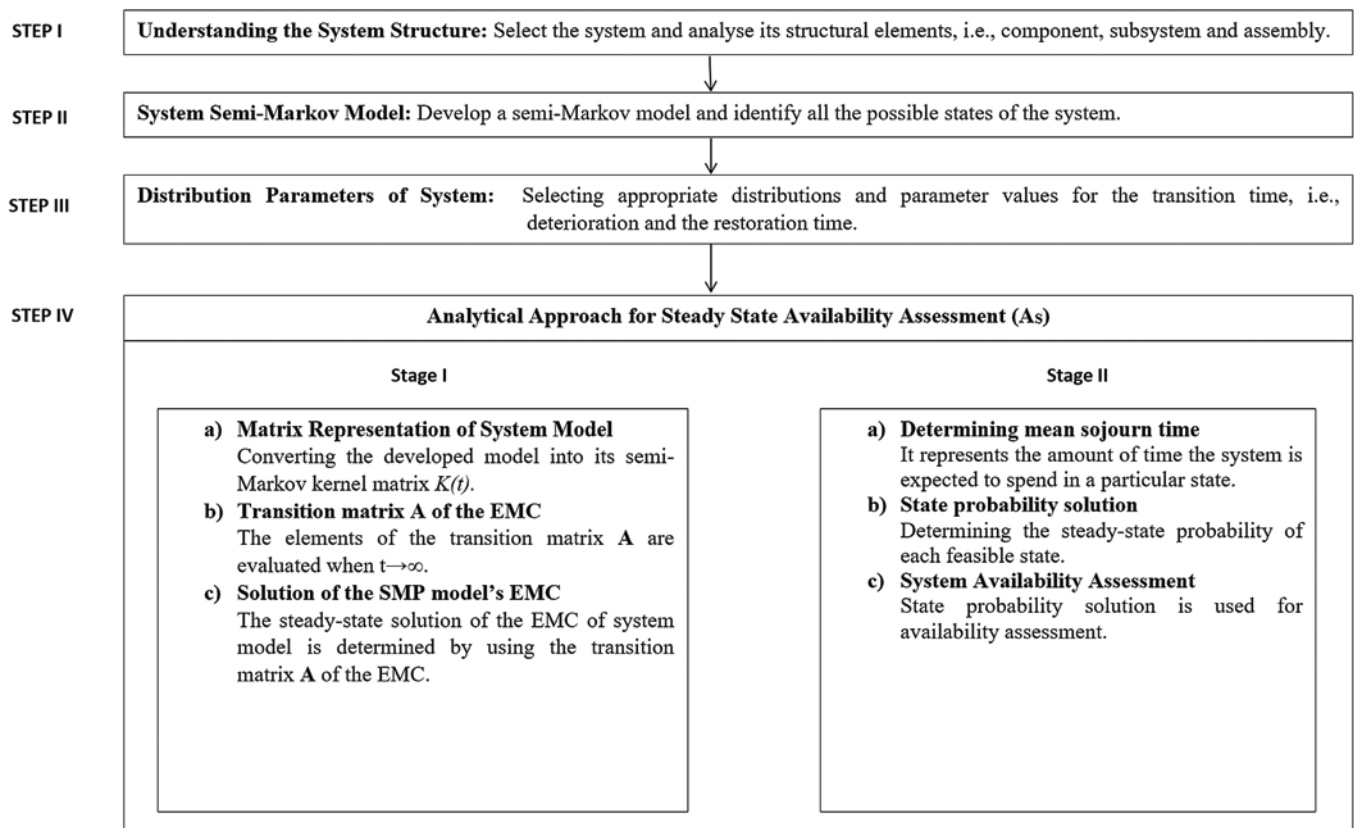
The equation expressed below is utilized for determining the System Availability ( $A_s$ ).

$$A_s = \sum_{i=1} P_i, i \in I \quad (3)$$

where,  $P_i$  is the steady-state probability of the system in working state only.

## 3. System modeling

Degradation modelling of mechanical systems for availability analysis has drawn special attention of researchers in the recent years. Whenever a system is in working state, it is always degrading with



time representing gradual degradation. Therefore, the degradation process for the system must be interpreted in such a manner that an effective model is built and implemented in operation. In this section, semi-Markov model is developed based on subsystem degradation.

### 3.1. System Description

A system is characterized as a set of subsystems/components working together towards achieving some logical end. Coupling relationships occurs between various parts of the system and also between various fault types, resulting in several routes for the propagation of faults [7]. The two stage reciprocating air compressor system consists of low pressure compressor (LPC) and high pressure compressor (HPC) with intercooler in between, system in series configuration, is selected for this paper as shown in the Fig. 1. This series system is widely used in power plants to increase the air pressure without increasing its temperature as this result in power saving in compressing the air.

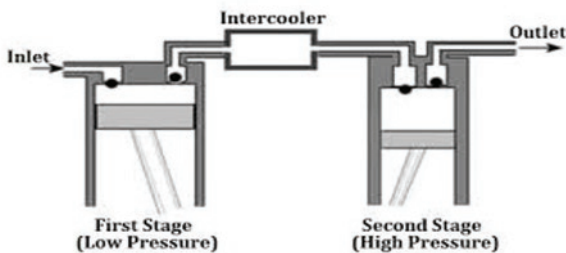


Fig. 1. Multi-stage reciprocating air compressor system

Being a series system, it should be highly available when called to function so that the power plant operations do not suffer. There is a huge loss in case of continuous operating units when unexpected shutdown occurs and their economic competitiveness implies an effective maintenance strategy to improve system availability and reduce operating cost [17]. Therefore, it becomes important to evaluate availability of such system so that overall system availability can be analyzed and appropriate maintenance policy can be suggested to improve it.

### 3.2. System Model Development

A system is characterized as a set of subsystems/components working together and its performance depends on how individual components work. Configuration of a series production system is such that failure of any machine may cause the entire system breakdown [34]. If proper attention is not given to the system in operation, it can result in a catastrophic failure that needs significant repair time and expense [35]. To scientifically research it, we also have to make a series of assumptions that typically take the form of mathematical or logical relationships that constitutes a model. Variables are required at a given time to explain the state of system. A system's failure is due to inappropriate functional interplay between its components and subsystems [29]. Many real-world systems are therefore too complicated to be evaluated analytically but still an attempt is being made to model such complex mechanical system. A system consisting of three subsystems in series configuration, with three degradation states, namely operational, degraded and failed, is considered for availability analysis. When any of its subsystems or part fails, it enters the repair process. The system is expected to perform its intended functions when an appropriate maintenance policy is adopted, however in practice it may unexpectedly fail. The corrective maintenance approach is more suitable in such conditions for failed subsystems or components as it is believed that shutdown and repair cost in case of breakdown will be less than the investment required for preventive maintenance. According to the level of repair, the corrective maintenance approach can be categorized into perfect and imperfect repair but in this paper,

only perfect repair is incorporated in the system which restores the system from failed state to operative state whenever the system enters the repair process. For mechanical systems and subsystems, the maintenance actions are varied with the age of components which is described by means of increasing failure rate therefore; it becomes evident for the study of system availability to accept multi-state degradation at the subsystem/component level, depending on their life span. The Weibull distribution, which is more appropriate for degrading systems, is considered [9].

For each subsystem/component, O, D and F are assumed in order for the system undergoing gradual degradation as shown in the Fig. 2. Each component is assumed to be in any of the three states throughout, where; O: Operative; D: Degraded; F: Failed. The states O and D are considered as the working states while F is considered as the repair state.

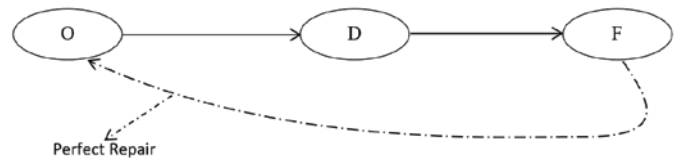


Fig. 2. Multi-state degradation of a subsystem and perfect repair

Models are developed with corrective maintenance approach for system availability assessment. Various maintenance models find the maintenance actions to be done perfectly. In fact, the effectiveness of maintenance staff to restore the failed component/subsystem usually lies between two extreme limits (“as good as new” and “as bad as old”), commonly referred as imperfect repair [23]. However, only perfect repair module is discussed in this study. The recoverability dimension quantifies how quickly and how well a system can recover after interruption to its normal state [8]. In the system model development, as mentioned earlier state O is considered as the original operating state, i.e. “as good as new” state and D as degraded state and F as the failed state at the subsystem level. Refer Fig. 2, a maintenance action, i.e. perfect repair is considered, which restores the subsystem from its failed state F to the operating state O. The states O and D of the subsystem is considered as working states, while F is the ‘repair state’ due to performance below the unacceptable level. Let a system is considered with its three subsystems in series, with each subsystem having one original operating state, one degraded state and one failed state. It is assumed that only one subsystem is changing its state at a particular time. State of a system is dependent on the state of its subsystems. All the feasible states and their corresponding transitions are identified for the system to develop SMP model as shown in the Table 1. In developing the system model simultaneous failure of two or more subsystems is not considered as it is assumed to be an infeasible state.

The criteria to be followed in decision making for corrective maintenance and/or opportunistic maintenance are discussed here. Table 1 shows system states for three subsystems ‘1’, ‘2’, and ‘3’ in series, each with three states as O, D and F. The last column of the table gives the maintenance option or possibility of ‘Corrective Maintenance and/or opportunistic maintenance’ for the subsystems ‘1’, ‘2’, and ‘3’, which is decided from the system state status (Column 2); ‘Under Repair’, i.e. 3, 6, 7, 8, 11, 14 to 20. The rationale behind the decision making ‘yes’ or ‘no’ is illustrated by considering the first system state status ‘Under Repair’ (Column 2) in the Table 1, i.e. at S. No. 3 (Column 1). In this state, states of subsystems ‘1’, ‘2’ and ‘3’ (Column 3, 4 and 5) are O, O and F respectively. For the subsystem ‘3’ in state, F i.e. the failed state, a corrective maintenance (perfect repair) is the best choice as one needs considerable time to perform the maintenance task, including use of resources needed which are also expected to be on a higher side. In view of this, it is ‘yes’ logic for corrective maintenance (perfect repair) of



subsystem '3'. With subsystem '3' under maintenance, there is need to check if opportunistic maintenance for the other subsystems, i.e. subsystem '1' and '2' can be undertaken. For this, one needs to check their state, which is O and O, i.e. 'Original Operating State', and they does not need it. In view of this, it is 'no' logic for opportunistic maintenance. In a similar way, decision 'yes' or 'no' for other system states 6, 7, 8, 11, 14 to 20, which are 'Under Repair' are carried out. In all, two models for the system have been developed as discussed in the following subsections.

### 3.2.1. System model based on corrective maintenance (perfect repair) without opportunistic maintenance

In a system with three subsystems in series configuration, out of the twenty feasible states of the system, twelve states (3, 11, 6,7, 8, 14, 15, 16, 17, 18, 19 and 20) are the 'under repair' system states, while the remaining eight states falls under the 'working states' category. Considering subsystem degradation and perfect repair, a system model is developed as shown in the Figure 3. The system model features two types of transition edges. Refer Fig. 3, a black line depicts the degradation of the subsystem from O to D and from D to F while a continuous dotted black line represents perfect repair, i.e. from F to O. Perfect repair is carried out for under repair states and such states are represented in the model by transition lines 3-1, 7-1, 8-2, 11-9, 14-12, 15-9, 16-10, 17-1, 18-2, 19-4 and 20-5 for the subsystems restoring the subsystem which is under repair from state F to O.

### 3.2.2. System Model based on corrective maintenance (perfect repair) with opportunistic maintenance

In opportunistic maintenance, whenever a system or module is grounded for corrective or preventive maintenance, that opportunity is utilized to do maintenance on other parts of the module, which are found to be damaged or have started to deteriorate. On

one hand, this improves the safety and reliability of the system, and on the other hand it reduces the downtime by avoiding unscheduled maintenance. This in turn reduces the cost of maintenance and loss of revenue due to extra groundings. In this model, opportunistic maintenance is considered along with perfect repair. Following the similar pattern, a system model is developed considering subsystem degradation and opportunistic maintenance with perfect repair, as shown in the Figure 4. In the system model, the transition lines for degradation are same as in the previous model, Fig. 3. The system being in the 'under repair' states, 3, 7 and 17, only perfect repair is carried out for these as the opportunistic maintenance is not possible because the other subsystems are still in state 'O' i.e. 'as good as new' state. In the system model, perfect repair for these states is represented by transition line 3-1, 7-1, and 17-1 for the subsystems '3', '2', and '1' respectively, restoring the system which is under repair from its state, F to O. For the remaining nine 'under repair' states, opportunistic maintenance is possible because the other subsystems are in state D, i.e. degraded state. In the system model, opportunistic maintenance with perfect repair is represented by transition lines 6→1, 8→1, 11→1, 14→1, 15→1, 16→1, 18→1, 19→1 and 20→1, restoring the system which is under repair from its state, D to O and F to O. It is assumed that the time to restore a subsystem in perfect repair from state F to O is more than the restoration time of other partially degraded states of the subsystems in opportunistic maintenance from state D to O.

## 4. Solution of the System Model

The solution of the SMP model is divided into two stages as discussed below. Stage 1 deals with matrix representation of the system model while stage 2 deals with system availability assessment using sojourn time and state probability values.

Table 1. States of a two stage reciprocating compressor system and their corresponding transitions

System State		State of subsystem 1	State of subsystem 2	State of subsystem 3	Transition to	Maintenance Possibility	
S.NO	Status					C.M*	O.M*
1	Working	O	O	O	2 4 9	-	-
2	Working	O	O	D	3 5 10	-	-
3	Under Repair	O	O	F	1	Yes(S <sub>3</sub> )	No
4	Working	O	D	O	5 7 12	-	-
5	Working	O	D	D	6 8 13	-	-
6	Under Repair	O	D	F	4	Yes(S <sub>3</sub> )	Yes(S <sub>2</sub> )
7	Under Repair	O	F	O	1	Yes(S <sub>2</sub> )	No
8	Under Repair	O	F	D	2	Yes(S <sub>2</sub> )	Yes(S <sub>3</sub> )
9	Working	D	O	O	17 12 10	-	-
10	Working	D	O	D	18 13 11	-	-
11	Under Repair	D	O	F	9	Yes(S <sub>3</sub> )	Yes(S <sub>1</sub> )
12	Working	D	D	O	19 15 13	-	-
13	Working	D	D	D	20 16 14	-	-
14	Under Repair	D	D	F	12	Yes(S <sub>3</sub> )	Yes(S <sub>1,2</sub> )
15	Under Repair	D	F	O	9	Yes(S <sub>2</sub> )	Yes(S <sub>1</sub> )
16	Under Repair	D	F	D	10	Yes(S <sub>2</sub> )	Yes(S <sub>1,3</sub> )
17	Under Repair	F	O	O	1	Yes(S <sub>1</sub> )	No
18	Under Repair	F	O	D	2	Yes(S <sub>1</sub> )	Yes(S <sub>3</sub> )
19	Under Repair	F	D	O	4	Yes(S <sub>1</sub> )	Yes(S <sub>2</sub> )
20	Under Repair	F	D	D	5	Yes(S <sub>1</sub> )	Yes(S <sub>2,3</sub> )

\*C.M- Corrective Maintenance, O.M- Opportunistic Maintenance

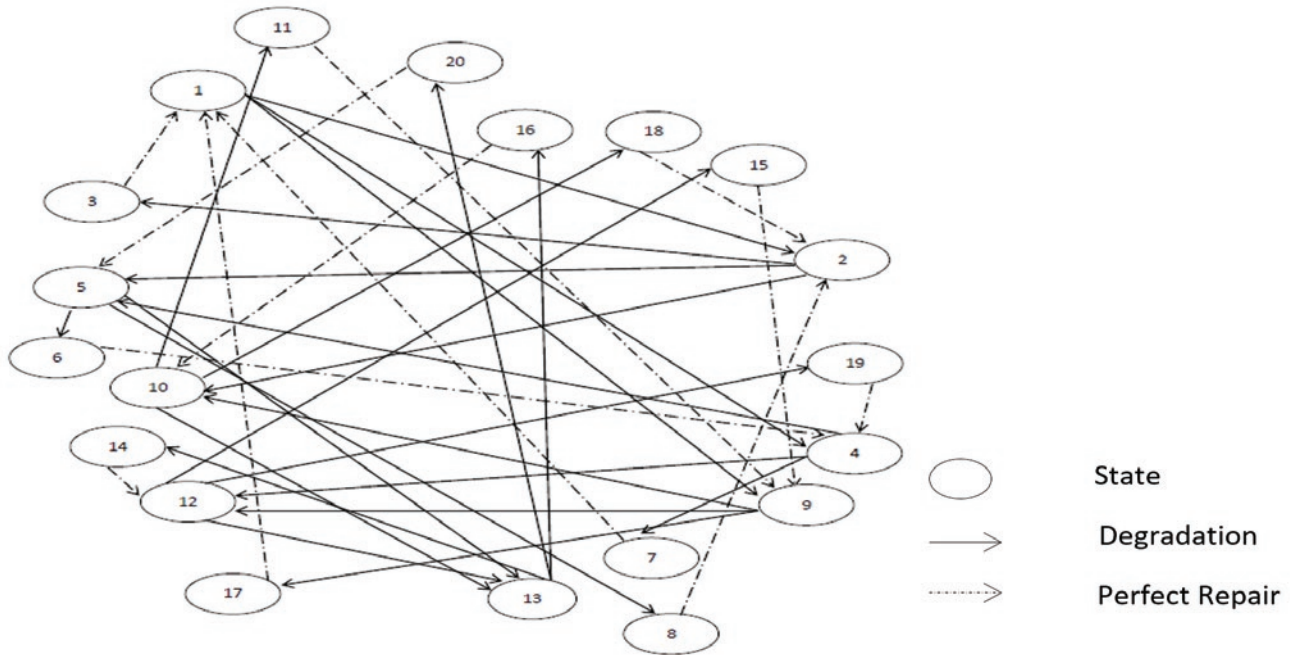


Fig. 3. Semi-Markov model for the system undergoing perfect repair without opportunistic maintenance

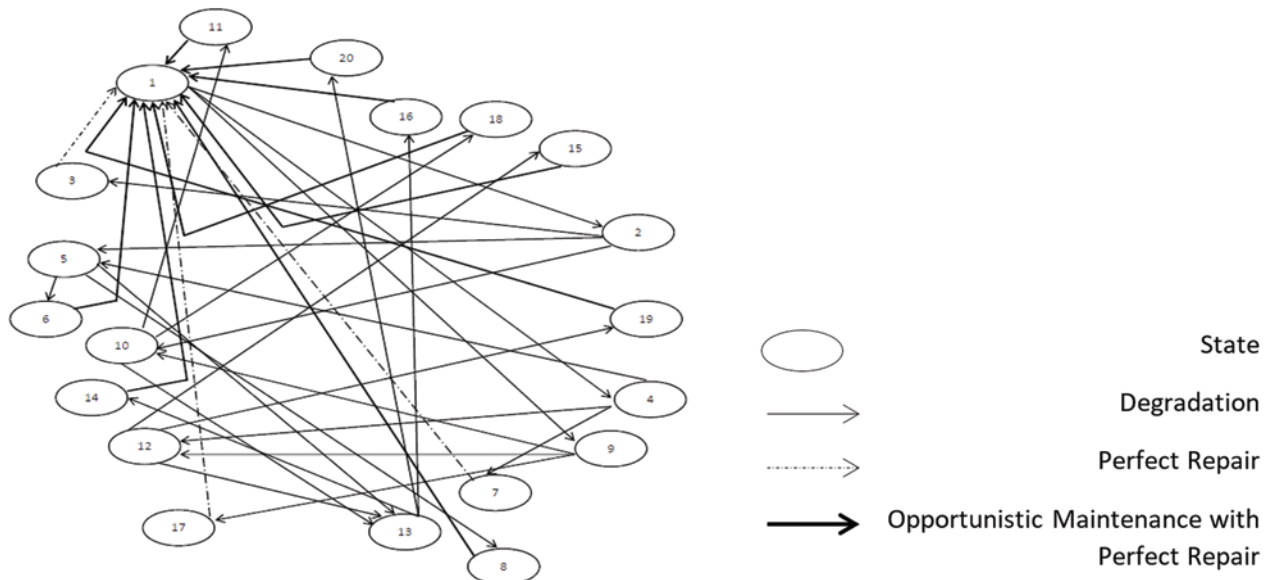


Fig. 4. Semi-Markov model for the system undergoing perfect repair coupled with opportunistic maintenance

### Stage I

a) The SMP model is described by its Kernel matrix,  $K(t)$  as shown below.

$$K(t) = \begin{bmatrix} k_{11}(t) & k_{12}(t) & \cdots & \cdots & k_{1N}(t) \\ k_{21}(t) & k_{22}(t) & \cdots & \cdots & k_{2N}(t) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ k_{N1}(t) & k_{N2}(t) & \cdots & \cdots & k_{NN}(t) \end{bmatrix}$$

The matrix  $K(t) = [k_{ij}(t)]$  is called kernel of the SMP model and the matrix elements gives the probability of jumping from one state to another. The matrix's rows and columns correspond to the number of feasible system states. The matrix elements,  $k_{ij}(t)$  are defined as [14]:

$$k_{ij}(t) = \begin{cases} 0 & \text{When there is no possible transition from state } i \text{ within time } t. \\ F_{ij}(t) & \text{When there is a single possible transition from state } i \text{ to state } j \text{ within time } t. \\ \int_0^t \bar{F}_{ik} \bar{F}_{im} dF_{ij}(t) & \text{When there are multiple possible transition from state } i \text{ to state } j, k \text{ and } m \text{ within time } t. \end{cases}$$

where,  $F_{ij}(t)$  is the CDF and  $\bar{F}_{ij}(t) = 1 - F_{ij}(t)$  represents complement of CDF associated with transition from state  $i$  to  $j$ . Therefore, the kernel matrix  $K(t)$  of the SMP model developed as shown in the Fig. 3 is obtained as:

$$K(t) = \begin{bmatrix} 0 & k_{12} & 0 & k_{14} & 0 & 0 & 0 & 0 & k_{19} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_{23} & 0 & k_{25} & 0 & 0 & 0 & 0 & k_{210} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{45} & 0 & k_{47} & 0 & 0 & 0 & 0 & k_{412} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & k_{56} & 0 & k_{58} & 0 & 0 & 0 & 0 & k_{513} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_{64} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_{71} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_{82} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{910} & 0 & k_{912} & 0 & 0 & 0 & 0 & k_{917} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{1011} & 0 & k_{1013} & 0 & 0 & 0 & 0 & k_{1018} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{119} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{1213} & 0 & k_{1215} & 0 & 0 & 0 & k_{1219} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{1314} & 0 & k_{1316} & 0 & 0 & 0 & k_{1320} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{1412} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{159} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_{1610} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_{171} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_{182} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_{194} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{205} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Table 2. Distribution parameters of compressor system [2]

CDFs	Distribution	Parameter	CDFs	Distribution	Parameter
F <sub>12</sub>	Weibull	$\beta_{12}=3.78, \theta_{12}=10000$	F <sub>917</sub>	Weibull	$\beta_{917}= 2.5, \theta_{917}=6776$
F <sub>14</sub>	Weibull	$\beta_{14}=1.2, \theta_{14}=6000$	F <sub>1011</sub>	Weibull	$\beta_{1011}= 3.78, \theta_{1011}=4892$
F <sub>19</sub>	Weibull	$\beta_{19}=2.5, \theta_{19}=16000$	F <sub>1013</sub>	Weibull	$\beta_{1013}= 1.2, \theta_{1013}=6000$
F <sub>23</sub>	Weibull	$\beta_{23}=3.78, \theta_{23}=4892$	F <sub>1018</sub>	Weibull	$\beta_{1018}= 2.5, \theta_{1018}=6776$
F <sub>25</sub>	Weibull	$\beta_{25}= 1.2, \theta_{25}=6000$	F <sub>119</sub>	Exponential	$\mu_{119}=0.001667$
F <sub>210</sub>	Weibull	$\beta_{210}=2.5, \theta_{210}=16000$	F <sub>1213</sub>	Weibull	$\beta_{1213}= 3.78, \theta_{1213}=10000$
F <sub>31</sub>	Exponential	$\mu_{31}=0.001667$	F <sub>1215</sub>	Weibull	$\beta_{1215}= 1.2, \theta_{1215}=4000$
F <sub>45</sub>	Weibull	$\beta_{45}=3.78, \theta_{45}=10000$	F <sub>1219</sub>	Weibull	$\beta_{1219}= 2.5, \theta_{1219}=6776$
F <sub>47</sub>	Weibull	$\beta_{47}= 1.2, \theta_{47}=4000$	F <sub>1314</sub>	Weibull	$\beta_{1314}= 3.78, \theta_{1314}=4892$
F <sub>412</sub>	Weibull	$\beta_{412}= 2.5, \theta_{412}=16000$	F <sub>1316</sub>	Weibull	$\beta_{1316}= 1.2, \theta_{1316}=4000$
F <sub>56</sub>	Weibull	$\beta_{56}= 3.78, \theta_{56}=4892$	F <sub>1320</sub>	Weibull	$\beta_{1320}= 2.5, \theta_{1320}=6776$
F <sub>58</sub>	Weibull	$\beta_{58}= 1.2, \theta_{58}=4000$	F <sub>1412</sub>	Exponential	$\mu_{1412}=0.001667$
F <sub>513</sub>	Weibull	$\beta_{513}= 2.5, \theta_{513}=16000$	F <sub>159</sub>	Exponential	$\mu_{159}=0.001667$
F <sub>64</sub>	Exponential	$\mu_{64}=0.001667$	F <sub>1610</sub>	Exponential	$\mu_{1610}=0.001667$
F <sub>71</sub>	Exponential	$\mu_{71}=0.001667$	F <sub>171</sub>	Exponential	$\mu_{171}=0.001667$
F <sub>82</sub>	Exponential	$\mu_{82}=0.001667$	F <sub>182</sub>	Exponential	$\mu_{182}=0.001667$
F <sub>910</sub>	Weibull	$\beta_{910}=3.78, \theta_{910}=10000$	F <sub>194</sub>	Exponential	$\mu_{194}=0.001667$
F <sub>912</sub>	Weibull	$\beta_{912}= 1.2, \theta_{912}=6000$	F <sub>205</sub>	Exponential	$\mu_{205}=0.001667$

The empirical estimators of the kernel matrix  $K(t)$ , are determined for the semi-Markov model as shown in the Fig. 3 using the distribution available in Table 2.

For example, for the non-zero element of the kernel matrix,  $k_{12}$ , see the details for its distribution available in Table 2. It is clear from

the Fig. 3 that there are three outgoing transitions from State 1 to State 2, 4 and 9 respectively. The distribution of these three transitions is Weibull, therefore the expression for the non-zero element of the kernel matrix,  $K_{12}$  in terms of Weibull parameters is expressed as follows:

$$k_{12} = \int_0^t \overline{F_{14}F_{19}} dF_{12} = \frac{\beta_{12}}{(\theta_{12})^{\beta_{12}}} \int_0^t (\beta_{12}-1) e^{-\left[ \left(\frac{t}{\theta_{12}}\right)^{\beta_{12}} + \left(\frac{t}{\theta_{14}}\right)^{\beta_{14}} + \left(\frac{t}{\theta_{19}}\right)^{\beta_{19}} \right]} dt \quad (4)$$

Similarly, for the non-zero element of the kernel matrix,  $k_{31}$ , see the details for its distribution available in Table 2. It is clear from the Fig. 3 that there is only one outgoing transition from State 3. The distribution of this transition is exponential; therefore the expression for the matrix element,  $k_{31}$  in terms of exponential parameters is expressed as follows:

$$k_{31} = F_{31}(t) = 1 - e^{-(\mu_3 t)} \quad (5)$$

Remaining elements of the kernel matrix  $K(t)$  are expressed as explained above, and all these expressions are listed in "Appendix: Table 8".

The transition matrix A of the Embedded Markov Chain (EMC), obtained is:

$k_{ij}$	Value	$k_{ij}$	Value	$k_{ij}$	Value	$k_{ij}$	Value
$k_{12}$	0.1885	$k_{45}$	0.0896	$k_{910}$	0.0753	$k_{1213}$	0.0431
$k_{14}$	0.7367	$k_{47}$	0.8678	$k_{912}$	0.5687	$k_{1215}$	0.7238
$k_{19}$	0.0748	$k_{412}$	0.0426	$k_{917}$	0.3560	$k_{1219}$	0.2331
$k_{23}$	0.4915	$k_{56}$	0.3066	$k_{1011}$	0.3743	$k_{1314}$	0.2610
$k_{25}$	0.4823	$k_{58}$	0.6741	$k_{1013}$	0.4397	$k_{1316}$	0.5993
$k_{210}$	0.0262	$k_{513}$	0.0193	$k_{1018}$	0.1860	$k_{1320}$	0.1397

0	0.1885	0	0.7367	0	0	0	0	0.0748	0	0	0	0	0	0	0	0	0	0	0
0	0	0.4915	0	0.4823	0	0	0	0	0.0262	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0.0896	0	0.8678	0	0	0	0	0.0426	0	0	0	0	0	0	0	0
0	0	0	0	0	0.3066	0	0.6741	0	0	0	0	0.0193	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0.0753	0	0.5687	0	0	0	0	0.3560	0	0	0
0	0	0	0	0	0	0	0	0	0	0.3743	0	0.4397	0	0	0	0	0.1860	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0.0431	0	0.7238	0	0	0	0.2331	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2610	0	0.5993	0	0	0.1397
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

b) Transition matrix **A** of the EMC  
 The state transition matrix **A** has to be determined such that the system spends a prescribed amount of time in each state before making transition. The system transits from the operative state to the failed state in a strongly correlated manner with the time. To imitate this behavior of system model, some coupling is required with the matrix  $K(t)$ .

The developed matrix  $K(t)$  helps in determining the matrix,  $A = K(\infty)$ , of the EMC considering  $t \rightarrow \infty$ , which is necessary condition for the steady-state analysis. As it is clear from the kernel matrix,  $K(t)$ , that there is only one element in the 3<sup>rd</sup>, 6<sup>th</sup>, 7<sup>th</sup>, 8<sup>th</sup>, 11<sup>th</sup>, 14<sup>th</sup>, 15<sup>th</sup>, 16<sup>th</sup>, 17<sup>th</sup>, 18<sup>th</sup>, 19<sup>th</sup>, and 20<sup>th</sup> row, therefore, the values of elements  $k_{31}(\infty)$ ,  $k_{64}(\infty)$ ,  $k_{71}(\infty)$ ,  $k_{82}(\infty)$ ,  $k_{119}(\infty)$ ,  $k_{1412}(\infty)$ ,  $k_{159}(\infty)$ ,  $k_{1610}(\infty)$ ,  $k_{171}(\infty)$ ,  $k_{182}(\infty)$ ,  $k_{194}(\infty)$  and  $k_{205}(\infty)$  are all equal to 1. Distribution parameter values shown in Table 2 are substituted in the expressions of kernel matrix elements and these are solved using MATLAB [20]. Table 3 shows the non-zero elements of matrix **A**. By substituting the values of non-zero elements in the kernel matrix,  $K(t)$ , the complete transition probability matrix **A** is obtained.

c) Probability of the states of EMC

The matrix **A** helps in determining the state probability of EMC. The steady-state solution of the EMC is determined using the equation expressed below which is to be solved using MATLAB and the values of  $s_i$ ,  $i \in I$  are listed in Table 4.

$$[s_1 s_2 s_3 s_4 s_5 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{12} s_{13} s_{14} s_{15} s_{16} s_{17} s_{18} s_{19} s_{20}] = [s_1 s_2 s_3 s_4 s_5 s_6 s_7 s_8 s_9 s_{10} s_{11} s_{12} s_{13} s_{14} s_{15} s_{16} s_{17} s_{18} s_{19} s_{20}] * A \quad (6)$$

Table 4. Probability of the states of EMC

$s_i$	Probability Value	$s_i$	Probability Value	$s_i$	Probability Value	$s_i$	Probability Value
$s_1$	0.2289	$s_6$	0.0191	$s_{11}$	0.0038	$s_{16}$	0.0043
$s_2$	0.0820	$s_7$	0.1705	$s_{12}$	0.0376	$s_{17}$	0.0171
$s_3$	0.0403	$s_8$	0.0370	$s_{13}$	0.0071	$s_{18}$	0.0019
$s_4$	0.1964	$s_9$	0.0481	$s_{14}$	0.0019	$s_{19}$	0.0088
$s_5$	0.0571	$s_{10}$	0.0100	$s_{15}$	0.0272	$s_{20}$	0.0009

**Stage 2**

a) Mean sojourn time evaluation of all the feasible states of the system model

The mean sojourn time,  $T_i$ , is the amount of time the system is expected to spend at state  $i$ , before leaving for another state depending upon the configuration of the system. This helps in determining the steady-state probability of all the feasible states of the system model.

See the details for its distribution available in Table 2 for sojourn time expression for state 1,  $T_1$ . It is clear from the Fig. 3 that there are three outgoing transitions from State 1 to State 2, 4 and 9 respectively. The distribution of these three transitions is Weibull, therefore the expression for the mean sojourn time,  $T_1$  in terms of Weibull parameters is expressed as follows [14]:

$$T_1 = \int_0^{\infty} \overline{F_{12} F_{14} F_{19}} dt = \int_0^{\infty} e^{-\left[ \left(\frac{t}{\theta_{12}}\right)^{\beta_{12}} + \left(\frac{t}{\theta_{14}}\right)^{\beta_{14}} + \left(\frac{t}{\theta_{19}}\right)^{\beta_{19}} \right]} dt \quad (7)$$

See the details for its distribution available in Table 2 for sojourn time expression for state 1,  $T_3$ . From Figure 3, it is evident that only one state is reachable from state 3 i.e. to state 1. The distribution is exponential for this transition; therefore the expression of the sojourn time,  $T_3$ , in terms of exponential parameters is expressed as follows [14]:

$$T_3 = \int_0^{\infty} \overline{F_{31}} dt = \int_0^{\infty} e^{-(\mu_{31}t)} dt \quad (8)$$

Likewise, the time values are represented in ‘‘Appendix: Table 9’’ for the remaining feasible states of the system model. Using the sojourn time expressions of all the states and by substituting the distribution parameter values listed down in Table 2, the sojourn time for all the states are determined. To solve sojourn time expressions, MATLAB is used and its value is shown in Table 5.

Table 5. Mean Sojourn Time of System States

$T_i$	Value(hr)	$T_i$	Value(hr)	$T_i$	Value(hr)	$T_i$	Value(hr)
$T_1$	4458.2	$T_6$	600	$T_{11}$	600	$T_{16}$	600
$T_2$	3176.8	$T_7$	600	$T_{12}$	2922.8	$T_{17}$	600
$T_3$	600	$T_8$	600	$T_{13}$	2519.4	$T_{18}$	600
$T_4$	3379	$T_9$	3611.9	$T_{14}$	600	$T_{19}$	600
$T_5$	2683.9	$T_{10}$	2941.2	$T_{15}$	600	$T_{20}$	600

b) Steady-state probability of states of the system model

The probability ‘ $P_i$ ’ of all the feasible states of the system model is determined by Eq. (1). The evaluated values of the states are listed in Table 6.

Table 6. Steady-state probability of the states of System model

$P_i$	Probability values	$P_i$	Probability values	$P_i$	Probability values	$P_i$	Probability values
$P_1$	0.3882	$P_6$	0.0044	$P_{11}$	0.00085842	$P_{16}$	0.00097674
$P_2$	0.0991	$P_7$	0.0384	$P_{12}$	0.0418	$P_{17}$	0.0039
$P_3$	0.0092	$P_8$	0.0084	$P_{13}$	0.0068	$P_{18}$	0.00042657
$P_4$	0.2525	$P_9$	0.0661	$P_{14}$	0.00042538	$P_{19}$	0.0020
$P_5$	0.0583	$P_{10}$	0.0112	$P_{15}$	0.0062	$P_{20}$	0.00022768

c) Availability Assessment

The state probability solutions are used for availability assessment. Availability is the summation of state probabilities in which the system is operational or available [15]. It is determined by using Eq. (3);  $A_s = 0.9241$ . As mentioned earlier system availability is a performance measure and is defined as a measure of the percentage of time the equipment is in operable state. The steady-state availability of a two stage recip-

Table 7. Steady-state availability analysis

S.No.	Corrective Maintenance	System Availability	
		Without Opportunistic Maintenance	With Opportunistic Maintenance
1.	Perfect Repair	0.9241	0.9342

rocating air compressor system by analytical semi-Markov approach is 0.9241. It represents that the system is available 92.41% of the time and working at 92.41% of the system’s technical limit. Availability measure is typically equal to the financial output of the system. Therefore, availability modeling and analysis is crucial for degrading system as it will give insights for its improvement. All our efforts should be focused on improving system availability in order to achieve the planned service life. The analytical approach utilized for availability assessment typically requires less time in computation but involves complex integrals. The suggested approach quantifies the impact of corrective maintenance policy in terms of system availability.

Steps explained in Section 4 are repeated for the SMP model based on corrective maintenance (perfect repair) with opportunistic maintenance. The results obtained are tabulated in Table 7.

As the results obtained provide a definite indication of the trend in the availability for different maintenance policies, these numeric results can be analyzed quantitatively to compare the relative improvement in the performance of the system in different scenarios. For the same mission time, moving from perfect repair without opportunistic maintenance to perfect repair with opportunistic maintenance, the availability shows the increasing trend. This clearly establishes that the maintenance policy: perfect repair without opportunistic maintenance is inefficient and should be seldom used unless cost of maintenance is the only dictating factor.

**5. Conclusions**

Considering the importance of maintenance, this research has attempted to establish a framework for availability modelling and analysis of degrading system. Using analytical semi-Markov methodology, a method for determining the availability based on corrective maintenance approach applied to complex mechanical system is suggested. By recognizing the configuration of the system, different subsystems/ components are identified and allocated at different stratified levels of mechanical system experiencing deterioration with ageing. The states and states transitions are identified. The state transition rates for deterioration and restoration are expressed by Weibull and exponential distribution respectively. The main contribution of this research is to develop mathematical model for system availability assessment which will give insights for its improvement. The steady-state availability of the system based on corrective maintenance (perfect repair) without opportunistic maintenance is low ( $A_s = 0.9241$ ). So, there lies a scope of improving it further. Therefore, the system needs attention on revising existing maintenance policies towards further improvement. Another model is added to investigate the gain in system availability when corrective maintenance (perfect repair) is combined with opportunistic maintenance. Hence, it is concluded that corrective maintenance (perfect repair) with opportunistic maintenance should be preferred and certain redundant strategies may also be applied in the designing of the system to enhance its availability further. The suggested approach is valuable for

maintenance personnel allowing them to establish repair and replacement policies. The designers can also use the methodology to make a decision on the implementation of the repairs and the degree of repair to achieve the required system availability values.

## 6. Practical Utility and Future Directions

In spite of the reality that the suggested technique provides numerous merits, it involves significant calculations for developing its detailed solution. The proposed approach is useful for plant engineers and maintenance personnel in designing a system with high availability by incorporating appropriate maintenance policies. The suggested methodology can be applied to variety of systems as it incorporates the

varying deterioration and restoration rates. In this paper, steady-state availability of the system is evaluated by employing the semi-Markov technique. Availability assessment under transient conditions case can also be carried out for measuring system performance. The suggested approach can also be extended for system availability assessment of mechanical systems considering various maintenance approaches like condition based maintenance, preventive maintenance, etc. The effect of human error can also be incorporated in the maintenance procedure in order to analyse its effect on system availability. After, one find a suitable way of quantifying system availability, it is possible to carry out further modifications of system design and operations thereby enhancing availability. Moreover, other performance parameters of complex mechanical systems, including resilience, maintainability and reliability will also be evaluated using this technique.

## Appendix

Table 8. Expressions of the elements of kernel matrix,  $K(t)$

$k_{ij}$	$k_{ij}(t)$	$k_{ij}(t)$ with specific distribution
$k_{12}$	$\int_0^t \overline{F_{14} F_{19}} dF_{12}$	$\frac{\beta_{12}}{(\theta_{12})^{\beta_{12}}} \int_0^t t^{(\beta_{12}-1)} e^{-\left[\left(\frac{t}{\theta_{12}}\right)^{\beta_{12}} + \left(\frac{t}{\theta_{14}}\right)^{\beta_{14}} + \left(\frac{t}{\theta_{19}}\right)^{\beta_{19}}\right]} dt$
$k_{14}$	$\int_0^t \overline{F_{12} F_{19}} dF_{14}$	$\frac{\beta_{14}}{(\theta_{14})^{\beta_{14}}} \int_0^t t^{(\beta_{14}-1)} e^{-\left[\left(\frac{t}{\theta_{12}}\right)^{\beta_{12}} + \left(\frac{t}{\theta_{14}}\right)^{\beta_{14}} + \left(\frac{t}{\theta_{19}}\right)^{\beta_{19}}\right]} dt$
$k_{19}$	$\int_0^t \overline{F_{14} F_{12}} dF_{19}$	$\frac{\beta_{19}}{(\theta_{19})^{\beta_{19}}} \int_0^t t^{(\beta_{19}-1)} e^{-\left[\left(\frac{t}{\theta_{12}}\right)^{\beta_{12}} + \left(\frac{t}{\theta_{14}}\right)^{\beta_{14}} + \left(\frac{t}{\theta_{19}}\right)^{\beta_{19}}\right]} dt$
$k_{23}$	$\int_0^t \overline{F_{25} F_{210}} dF_{23}$	$\frac{\beta_{23}}{(\theta_{23})^{\beta_{23}}} \int_0^t t^{(\beta_{23}-1)} e^{-\left[\left(\frac{t}{\theta_{23}}\right)^{\beta_{23}} + \left(\frac{t}{\theta_{25}}\right)^{\beta_{25}} + \left(\frac{t}{\theta_{210}}\right)^{\beta_{210}}\right]} dt$
$k_{25}$	$\int_0^t \overline{F_{23} F_{210}} dF_{25}$	$\frac{\beta_{25}}{(\theta_{25})^{\beta_{25}}} \int_0^t t^{(\beta_{25}-1)} e^{-\left[\left(\frac{t}{\theta_{23}}\right)^{\beta_{23}} + \left(\frac{t}{\theta_{25}}\right)^{\beta_{25}} + \left(\frac{t}{\theta_{210}}\right)^{\beta_{210}}\right]} dt$
$k_{210}$	$\int_0^t \overline{F_{25} F_{23}} dF_{210}$	$\frac{\beta_{210}}{(\theta_{210})^{\beta_{210}}} \int_0^t t^{(\beta_{210}-1)} e^{-\left[\left(\frac{t}{\theta_{23}}\right)^{\beta_{23}} + \left(\frac{t}{\theta_{25}}\right)^{\beta_{25}} + \left(\frac{t}{\theta_{210}}\right)^{\beta_{210}}\right]} dt$
$k_{31}$	$F_{31}(t)$	$1 - e^{-(\mu_{31}t)}$
$k_{45}$	$\int_0^t \overline{F_{47} F_{412}} dF_{45}$	$\frac{\beta_{45}}{(\theta_{45})^{\beta_{45}}} \int_0^t t^{(\beta_{45}-1)} e^{-\left[\left(\frac{t}{\theta_{45}}\right)^{\beta_{45}} + \left(\frac{t}{\theta_{47}}\right)^{\beta_{47}} + \left(\frac{t}{\theta_{412}}\right)^{\beta_{412}}\right]} dt$
$k_{47}$	$\int_0^t \overline{F_{45} F_{412}} dF_{47}$	$\frac{\beta_{47}}{(\theta_{47})^{\beta_{47}}} \int_0^t t^{(\beta_{47}-1)} e^{-\left[\left(\frac{t}{\theta_{45}}\right)^{\beta_{45}} + \left(\frac{t}{\theta_{47}}\right)^{\beta_{47}} + \left(\frac{t}{\theta_{412}}\right)^{\beta_{412}}\right]} dt$
$k_{412}$	$\int_0^t \overline{F_{47} F_{45}} dF_{412}$	$\frac{\beta_{412}}{(\theta_{412})^{\beta_{412}}} \int_0^t t^{(\beta_{412}-1)} e^{-\left[\left(\frac{t}{\theta_{45}}\right)^{\beta_{45}} + \left(\frac{t}{\theta_{47}}\right)^{\beta_{47}} + \left(\frac{t}{\theta_{412}}\right)^{\beta_{412}}\right]} dt$

$k_{56}$	$\int_0^t \overline{F_{58} F_{513}} dF_{56}$	$\frac{\beta_{56}}{(\theta_{56})^{\beta_{56}}} \int_0^t t^{(\beta_{56}-1)} e^{-\left[\left(\frac{t}{\theta_{56}}\right)^{\beta_{56}} + \left(\frac{t}{\theta_{58}}\right)^{\beta_{58}} + \left(\frac{t}{\theta_{513}}\right)^{\beta_{513}}\right]} dt$
$k_{56}$	$\int_0^t \overline{F_{56} F_{513}} dF_{58}$	$\frac{\beta_{58}}{(\theta_{58})^{\beta_{58}}} \int_0^t t^{(\beta_{58}-1)} e^{-\left[\left(\frac{t}{\theta_{56}}\right)^{\beta_{56}} + \left(\frac{t}{\theta_{58}}\right)^{\beta_{58}} + \left(\frac{t}{\theta_{513}}\right)^{\beta_{513}}\right]} dt$
$k_{513}$	$\int_0^t \overline{F_{58} F_{56}} dF_{513}$	$\frac{\beta_{513}}{(\theta_{513})^{\beta_{513}}} \int_0^t t^{(\beta_{513}-1)} e^{-\left[\left(\frac{t}{\theta_{56}}\right)^{\beta_{56}} + \left(\frac{t}{\theta_{58}}\right)^{\beta_{58}} + \left(\frac{t}{\theta_{513}}\right)^{\beta_{513}}\right]} dt$
$k_{64}$	$F_{64}(t)$	$1 - e^{-(\mu_{64}t)}$
$k_{71}$	$F_{71}(t)$	$1 - e^{-(\mu_{71}t)}$
$k_{82}$	$F_{82}(t)$	$1 - e^{-(\mu_{82}t)}$
$k_{910}$	$\int_0^t \overline{F_{912} F_{917}} dF_{910}$	$\frac{\beta_{910}}{(\theta_{910})^{\beta_{910}}} \int_0^t t^{(\beta_{910}-1)} e^{-\left[\left(\frac{t}{\theta_{910}}\right)^{\beta_{910}} + \left(\frac{t}{\theta_{912}}\right)^{\beta_{912}} + \left(\frac{t}{\theta_{917}}\right)^{\beta_{917}}\right]} dt$
$k_{912}$	$\int_0^t \overline{F_{910} F_{917}} dF_{912}$	$\frac{\beta_{912}}{(\theta_{912})^{\beta_{912}}} \int_0^t t^{(\beta_{912}-1)} e^{-\left[\left(\frac{t}{\theta_{910}}\right)^{\beta_{910}} + \left(\frac{t}{\theta_{912}}\right)^{\beta_{912}} + \left(\frac{t}{\theta_{917}}\right)^{\beta_{917}}\right]} dt$
$k_{917}$	$\int_0^t \overline{F_{912} F_{910}} dF_{917}$	$\frac{\beta_{917}}{(\theta_{917})^{\beta_{917}}} \int_0^t t^{(\beta_{917}-1)} e^{-\left[\left(\frac{t}{\theta_{910}}\right)^{\beta_{910}} + \left(\frac{t}{\theta_{912}}\right)^{\beta_{912}} + \left(\frac{t}{\theta_{917}}\right)^{\beta_{917}}\right]} dt$
$k_{1011}$	$\int_0^t \overline{F_{1013} F_{1018}} dF_{1011}$	$\frac{\beta_{1011}}{(\theta_{1011})^{\beta_{1011}}} \int_0^t t^{(\beta_{1011}-1)} e^{-\left[\left(\frac{t}{\theta_{1011}}\right)^{\beta_{1011}} + \left(\frac{t}{\theta_{1013}}\right)^{\beta_{1013}} + \left(\frac{t}{\theta_{1018}}\right)^{\beta_{1018}}\right]} dt$
$k_{1013}$	$\int_0^t \overline{F_{1011} F_{1018}} dF_{1013}$	$\frac{\beta_{1013}}{(\theta_{1013})^{\beta_{1013}}} \int_0^t t^{(\beta_{1013}-1)} e^{-\left[\left(\frac{t}{\theta_{1011}}\right)^{\beta_{1011}} + \left(\frac{t}{\theta_{1013}}\right)^{\beta_{1013}} + \left(\frac{t}{\theta_{1018}}\right)^{\beta_{1018}}\right]} dt$
$k_{1018}$	$\int_0^t \overline{F_{1013} F_{1011}} dF_{1018}$	$\frac{\beta_{1018}}{(\theta_{1018})^{\beta_{1018}}} \int_0^t t^{(\beta_{1018}-1)} e^{-\left[\left(\frac{t}{\theta_{1011}}\right)^{\beta_{1011}} + \left(\frac{t}{\theta_{1013}}\right)^{\beta_{1013}} + \left(\frac{t}{\theta_{1018}}\right)^{\beta_{1018}}\right]} dt$
$k_{119}$	$F_{119}(t)$	$1 - e^{-(\mu_{119}t)}$
$k_{1213}$	$\int_0^t \overline{F_{1215} F_{1219}} dF_{1213}$	$\frac{\beta_{1213}}{(\theta_{1213})^{\beta_{1213}}} \int_0^t t^{(\beta_{1213}-1)} e^{-\left[\left(\frac{t}{\theta_{1213}}\right)^{\beta_{1213}} + \left(\frac{t}{\theta_{1215}}\right)^{\beta_{1215}} + \left(\frac{t}{\theta_{1219}}\right)^{\beta_{1219}}\right]} dt$
$k_{1215}$	$\int_0^t \overline{F_{1213} F_{1219}} dF_{1215}$	$\frac{\beta_{1215}}{(\theta_{1215})^{\beta_{1215}}} \int_0^t t^{(\beta_{1215}-1)} e^{-\left[\left(\frac{t}{\theta_{1213}}\right)^{\beta_{1213}} + \left(\frac{t}{\theta_{1215}}\right)^{\beta_{1215}} + \left(\frac{t}{\theta_{1219}}\right)^{\beta_{1219}}\right]} dt$
$k_{1219}$	$\int_0^t \overline{F_{1215} F_{1213}} dF_{1219}$	$\frac{\beta_{1219}}{(\theta_{1219})^{\beta_{1219}}} \int_0^t t^{(\beta_{1219}-1)} e^{-\left[\left(\frac{t}{\theta_{1213}}\right)^{\beta_{1213}} + \left(\frac{t}{\theta_{1215}}\right)^{\beta_{1215}} + \left(\frac{t}{\theta_{1219}}\right)^{\beta_{1219}}\right]} dt$

$k_{1314}$	$\int_0^t \overline{F_{1316} F_{1320}} dF_{1314}$	$\frac{\beta_{1314}}{(\theta_{1314})^{\beta_{1314}}} \int_0^t t^{(\beta_{1314}-1)} e^{-\left[\left(\frac{t}{\theta_{1314}}\right)^{\beta_{1314}} + \left(\frac{t}{\theta_{1316}}\right)^{\beta_{1316}} + \left(\frac{t}{\theta_{1320}}\right)^{\beta_{1320}}\right]} dt$
$k_{1316}$	$\int_0^t \overline{F_{1314} F_{1320}} dF_{1316}$	$\frac{\beta_{1316}}{(\theta_{1316})^{\beta_{1316}}} \int_0^t t^{(\beta_{1316}-1)} e^{-\left[\left(\frac{t}{\theta_{1314}}\right)^{\beta_{1314}} + \left(\frac{t}{\theta_{1316}}\right)^{\beta_{1316}} + \left(\frac{t}{\theta_{1320}}\right)^{\beta_{1320}}\right]} dt$
$k_{1320}$	$\int_0^t \overline{F_{1316} F_{1314}} dF_{1320}$	$\frac{\beta_{1320}}{(\theta_{1320})^{\beta_{1320}}} \int_0^t t^{(\beta_{1320}-1)} e^{-\left[\left(\frac{t}{\theta_{1314}}\right)^{\beta_{1314}} + \left(\frac{t}{\theta_{1316}}\right)^{\beta_{1316}} + \left(\frac{t}{\theta_{1320}}\right)^{\beta_{1320}}\right]} dt$
$k_{1412}$	$F_{1412}(t)$	$1 - e^{-(\mu_{1412}t)}$
$k_{159}$	$F_{159}(t)$	$1 - e^{-(\mu_{159}t)}$
$k_{1610}$	$F_{1610}(t)$	$1 - e^{-(\mu_{1619}t)}$
$k_{171}$	$F_{171}(t)$	$1 - e^{-(\mu_{171}t)}$
$k_{182}$	$F_{182}(t)$	$1 - e^{-(\mu_{182}t)}$
$k_{194}$	$F_{194}(t)$	$1 - e^{-(\mu_{194}t)}$
$k_{205}$	$F_{205}(t)$	$1 - e^{-(\mu_{205}t)}$

Table 9. Sojourn time expressions for each system state

$T_i$	Sojourn time expressions	Sojourn time expressions with parameter values
$T_1$	$\int_0^{\infty} \overline{F_{12} F_{14} F_{19}} dt$	$\int_0^{\infty} e^{-\left[\left(\frac{t}{\theta_{12}}\right)^{\beta_{12}} + \left(\frac{t}{\theta_{14}}\right)^{\beta_{14}} + \left(\frac{t}{\theta_{19}}\right)^{\beta_{19}}\right]} dt$
$T_2$	$\int_0^{\infty} \overline{F_{23} F_{25} F_{210}} dt$	$\int_0^{\infty} e^{-\left[\left(\frac{t}{\theta_{23}}\right)^{\beta_{23}} + \left(\frac{t}{\theta_{25}}\right)^{\beta_{25}} + \left(\frac{t}{\theta_{210}}\right)^{\beta_{210}}\right]} dt$
$T_3$	$\int_0^{\infty} \overline{F_{31}} dt$	$\int_0^{\infty} e^{-(\mu_{31}t)} dt$
$T_4$	$\int_0^{\infty} \overline{F_{45} F_{47} F_{412}} dt$	$\int_0^{\infty} e^{-\left[\left(\frac{t}{\theta_{45}}\right)^{\beta_{45}} + \left(\frac{t}{\theta_{47}}\right)^{\beta_{47}} + \left(\frac{t}{\theta_{412}}\right)^{\beta_{412}}\right]} dt$
$T_5$	$\int_0^{\infty} \overline{F_{56} F_{58} F_{513}} dt$	$\int_0^{\infty} e^{-\left[\left(\frac{t}{\theta_{56}}\right)^{\beta_{56}} + \left(\frac{t}{\theta_{58}}\right)^{\beta_{58}} + \left(\frac{t}{\theta_{513}}\right)^{\beta_{513}}\right]} dt$
$T_6$	$\int_0^{\infty} \overline{F_{64}} dt$	$\int_0^{\infty} e^{-(\mu_{64}t)} dt$
$T_7$	$\int_0^{\infty} \overline{F_{71}} dt$	$\int_0^{\infty} e^{-(\mu_{71}t)} dt$



$T_8$	$\int_0^{\infty} \overline{F_{82}} dt$	$\int_0^{\infty} e^{-(\mu_{82}t)} dt$
$T_9$	$\int_0^{\infty} \overline{F_{910} F_{912} F_{917}} dt$	$\int_0^{\infty} e^{-\left[\left(\frac{t}{\theta_{910}}\right)^{\beta_{910}} + \left(\frac{t}{\theta_{912}}\right)^{\beta_{912}} + \left(\frac{t}{\theta_{917}}\right)^{\beta_{917}}\right]} dt$
$T_{10}$	$\int_0^{\infty} \overline{F_{1011} F_{1013} F_{1018}} dt$	$\int_0^{\infty} e^{-\left[\left(\frac{t}{\theta_{1011}}\right)^{\beta_{1011}} + \left(\frac{t}{\theta_{1013}}\right)^{\beta_{1013}} + \left(\frac{t}{\theta_{1018}}\right)^{\beta_{1018}}\right]} dt$
$T_{11}$	$\int_0^{\infty} \overline{F_{119}} dt$	$\int_0^{\infty} e^{-(\mu_{119}t)} dt$
$T_{12}$	$\int_0^{\infty} \overline{F_{1213} F_{1215} F_{1219}} dt$	$\int_0^{\infty} e^{-\left[\left(\frac{t}{\theta_{1213}}\right)^{\beta_{1213}} + \left(\frac{t}{\theta_{1215}}\right)^{\beta_{1215}} + \left(\frac{t}{\theta_{1219}}\right)^{\beta_{1219}}\right]} dt$
$T_{13}$	$\int_0^{\infty} \overline{F_{1314} F_{1316} F_{1320}} dt$	$\int_0^{\infty} e^{-\left[\left(\frac{t}{\theta_{1314}}\right)^{\beta_{1314}} + \left(\frac{t}{\theta_{1316}}\right)^{\beta_{1316}} + \left(\frac{t}{\theta_{1320}}\right)^{\beta_{1320}}\right]} dt$
$T_{14}$	$\int_0^{\infty} \overline{F_{1412}} dt$	$\int_0^{\infty} e^{-(\mu_{1412}t)} dt$
$T_{15}$	$\int_0^{\infty} \overline{F_{159}} dt$	$\int_0^{\infty} e^{-(\mu_{159}t)} dt$
$T_{16}$	$\int_0^{\infty} \overline{F_{1610}} dt$	$\int_0^{\infty} e^{-(\mu_{1619}t)} dt$
$T_{17}$	$\int_0^{\infty} \overline{F_{171}} dt$	$\int_0^{\infty} e^{-(\mu_{171}t)} dt$
$T_{18}$	$\int_0^{\infty} \overline{F_{182}} dt$	$\int_0^{\infty} e^{-(\mu_{182}t)} dt$
$T_{19}$	$\int_0^{\infty} \overline{F_{194}} dt$	$\int_0^{\infty} e^{-(\mu_{194}t)} dt$
$T_{20}$	$\int_0^{\infty} \overline{F_{205}} dt$	$\int_0^{\infty} e^{-(\mu_{205}t)} dt$

## References

1. Aju Kumar VN, Gandhi OP. Quantification of human error in maintenance using graph theory and matrix approach. *Quality and Reliability Engineering International* 2011; 27(8): 1145-72, <https://doi.org/10.1002/qre.1202>.
2. Barringer (2018) Weibull database. <http://www.barringer1.com/wdbase>.
3. Chen D, Trivedi KS. Optimization for condition-based maintenance with semi-Markov decision process. *Reliability Engineering & System Safety* 2005; 90(1): 25-29, <https://doi.org/10.1016/j.res.2004.11.001>.
4. Cheng Z, Remenyte-Priscott R. Two probabilistic life-cycle maintenance models for the deteriorating pavement. *Eksplotacja i Niezawodność - Maintenance and Reliability* 2018; 20 (3): 394-404, <https://doi.org/10.17531/ein.2018.3.7>.
5. Dong M. A novel approach to equipment health management based on auto-regressive hidden semi-Markov model (AR-HSMM). *Science in China Series F: Information Sciences* 2008; 51(9): 1291-304, <https://doi.org/10.1007/s11432-008-0111-4>.
6. Foley AJ. Regime switching in currency markets and portfolio flows. *Journal of Asset Management* 2002; 2(4): 353-367, <https://doi.org/10.1057/palgrave.jam.2240058>.
7. Hu J, Zhang L. Risk based opportunistic maintenance model for complex mechanical systems. *Expert Systems with Applications* 2014;

- 41(6): 3105-3115, <https://doi.org/10.1016/j.eswa.2013.10.041>.
8. Hu Z, Mahadevan S. Resilience assessment based on time-dependent system reliability analysis. *Journal of Mechanical Design* 2016; 138(11), <https://doi.org/10.1115/1.4034109>.
  9. Jager P, Bertsche B. A new approach to gathering failure behavior information about mechanical components based on expert knowledge. In *Annual Symposium Reliability and Maintainability, 2004-RAMS 2004* Jan 26 (pp. 90-95). IEEE.
  10. Kharoufeh JP, Solo CJ, Ulukus MY. Semi-Markov models for degradation-based reliability. *IEEE Transactions* 2010; 42(8): 599-612, <https://doi.org/10.1080/07408170903394371>.
  11. Koutras VP, Platis AN. Semi-Markov availability modeling of a redundant system with partial and full rejuvenation actions. In *2008 Third International Conference on Dependability of Computer Systems DepCoS-RELCOMEX 2008* Jun 26 (pp. 127-134), <https://doi.org/10.1109/DepCoS-RELCOMEX.2008.13>.
  12. Koutras VP, Platis AN. Semi-Markov performance modeling of a redundant system with partial, full and failed rejuvenation. *International Journal of Critical Computer-Based Systems* 2010; 1(1-3): 59-85, <https://doi.org/10.1504/IJCCBS.2010.031909>.
  13. Kulkarni VG. *Modeling and analysis of stochastic systems*. Crc Press; 2016 Nov 18.
  14. Kumar G, Jain V, Gandhi OP. Availability analysis of repairable mechanical systems using analytical semi-Markov approach. *Quality Engineering* 2013; 25(2): 97-107, <https://doi.org/10.1080/08982112.2012.751606>.
  15. Kumar G, Jain V, Gandhi OP. Reliability and availability analysis of mechanical systems using stochastic petri net modelling based on decomposition approach. *International Journal of Reliability, Quality and Safety Engineering* 2012; 19(01): 1250005, <https://doi.org/10.1142/S0218539312500052>.
  16. Kumar S, Tewari PC, Kumar S, Gupta M. Availability optimization of CO-shift conversion system of a fertilizer plant using genetic algorithm technique. *Bangladesh Journal of Scientific and Industrial Research* 2010; 45(2): 133-140, <https://doi.org/10.3329/bjsir.v45i2.5711>.
  17. Laggoune R, Chateaneuf A, Aissani D. Opportunistic policy for optimal preventive maintenance of a multi-component system in continuous operating units. *Computers & Chemical Engineering* 2009; 33(9): 1499-1510, <https://doi.org/10.1016/j.compchemeng.2009.03.003>.
  18. Li H, Zhao Q. Reliability evaluation of fault tolerant control with a semi-Markov fault detection and isolation model. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering* 2006; 220(5): 329-338, <https://doi.org/10.1243/09596518JSCE225>.
  19. Lotovskiy E, Teixeira AP, Guedes Soares C. Availability analysis of an offshore oil and gas production system subjected to age-based preventive maintenance by Petri Nets. *Eksplotacja i Niezawodność - Maintenance and Reliability* 2020; 22 (4): 627-637, <https://doi.org/10.17531/ein.2020.4.6>.
  20. MATLAB, R2018a. (2018). Natick, Massachusetts, USA: Mathworks.
  21. Moura MD, Droguett EL. A continuous-time semi-markov bayesian belief network model for availability measure estimation of fault tolerant systems. *Pesquisa Operacional* 2008; 28(2): 355-375, <https://doi.org/10.1590/S0101-74382008000200011>.
  22. Rajeevan AK, Shouri PV, Nair U. Markov modeling and reliability allocation in wind turbine for availability enhancement. *Life Cycle Reliability and Safety Engineering* 2018; 7(3): 147-157, <https://doi.org/10.1007/s41872-018-0054-8>.
  23. Samet S, Chelbi A, Hmida FB. Optimal availability of failure-prone systems under imperfect maintenance actions. *Journal of Quality in Maintenance Engineering* 2010, <https://doi.org/10.1108/13552511011084544>.
  24. Samhouri MS. An intelligent opportunistic maintenance (OM) system: a genetic algorithm approach. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH) 2009* Sep 26 (pp. 60-65), <https://doi.org/10.1109/TIC-STH.2009.5444428>.
  25. Shinde V, Biniwale D, Bharadwaj SK. Availability analysis for estimation of repair time of performance based logistics under operating condition. *Journal of Reliability and Statistical Studies* 2019; 12(1).
  26. Vaidyanathan K, Trivedi KS. A comprehensive model for software rejuvenation. *IEEE Transactions on Dependable and Secure Computing* 2005; 2(2): 124-137, <https://doi.org/10.1109/TDSC.2005.15>.
  27. Wang H. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research* 2002; 139(3): 469-489, [https://doi.org/10.1016/S0377-2217\(01\)00197-7](https://doi.org/10.1016/S0377-2217(01)00197-7).
  28. Wani MF, Gandhi OP. Development of maintainability index for mechanical systems. *Reliability Engineering & System Safety* 1999; 65(3): 259-270, [https://doi.org/10.1016/S0951-8320\(99\)00004-6](https://doi.org/10.1016/S0951-8320(99)00004-6).
  29. Wani MF, Gandhi OP. Failure analysis of mechanical systems based on function-cum-structure approach. *International Journal of Performability Engineering* 2008; 4(2): 141-152.
  30. Welte TM. A rule-based approach for establishing states in a Markov process applied to maintenance modelling. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 2009; 223(1): 1-2, <https://doi.org/10.1243/1748006XJRR194>.
  31. Xie M, Kong H, Goh TN. Exponential approximation for maintained Weibull distributed component. *Journal of Quality in Maintenance Engineering* 2000, <https://doi.org/10.1108/13552510010346761>.
  32. Xie W, Hong Y, Trivedi K. Analysis of a two-level software rejuvenation policy. *Reliability Engineering & System Safety* 2005; 87(1): 13-22, <https://doi.org/10.1016/j.res.2004.02.011>.
  33. Zhou B, Yu J, Shao J, Trentesaux D. Bottleneck-based opportunistic maintenance model for series production systems. *Journal of Quality in Maintenance Engineering* 2015, <https://doi.org/10.1108/JQME-09-2013-0059>.
  34. Zhu Y, Elsayed EA, Liao H, Chan LY. Availability optimization of systems subject to competing risk. *European Journal of Operational Research* 2010; 202(3): 781-788, <https://doi.org/10.1016/j.ejor.2009.06.008>.

**Prof. Andrzej Niewczas**

*Chair of Editorial Board*

**Prof. Holm Altenbach**

*Otto-von-Guericke-Universität, Magdeburg, Germany*

**Prof. John Andrews**

*University of Nottingham, Nottingham, UK*

**Prof. Karol Andrzejczak**

*Poznań University of Technology, Poznań, Poland*

**Prof. Christophe Bérenguer**

*Institut Polytechnique de Grenoble, Grenoble, France*

**Prof. Gintautas Bureika**

*Vilnius Gediminas Technical University, Vilnius, Lithuania*

**Dr Alireza Daneshkhan**

*Warwick Centre for Predictive Modelling*

*University of Warwick, UK*

**Prof. Luis Andrade Ferreira**

*University of Porto, Porto, Portugal*

**Prof. Mitra Fouladirad**

*Troyes University of Technology, France*

**Dr Ilia Frenkel**

*Shamoon College of Engineering, Beer Sheva, Israel*

**Prof. Olgierd Hryniewicz**

*Systems Research Institute of the Polish Academy of Science,*

*Warsaw, Poland*

**Prof. Hong-Zhong Huang**

*University of Electronic Science and Technology of China,*

*Chengdu, Sichuan, China*

**Prof. Marianna Jacyna**

*Warsaw University of Technology, Warsaw, Poland*

**Prof. Vaclav Legat**

*Czech University of Agriculture, Prague, Czech Republic*

**Prof. Jerzy Merkisz**

*Poznań University of Technology, Poznań, Poland*

**Prof. Gilbert De Mey**

*University of Ghent, Belgium*

**Prof. Maria Francesca Milazzo**

*University of Messina, Italy*

**Prof. Tomasz Nowakowski**

*Wrocław University of Technology, Wrocław, Poland*

**Prof. Marek Orkisz**

*Rzeszów University of Technology, Rzeszów, Poland*

**Prof. François Pérès**

*Toulouse University, Toulouse, France*

**Prof. Yi Ren**

*Beihang University, Beijing, China*

**Prof. Jan Szybka**

*AGH University of Science and Technology,*

*Cracow, Poland*

**Prof. Marcin Ślęzak**

*Motor Transport Institute, Warsaw, Poland*

**Prof. Katsumi Tanaka**

*Kyoto University, Kyoto, Japan*

**Prof. David Vališ**

*University of Defence, Brno, Czech Republic*

**Prof. Lesley Walls**

*University of Strathclyde, Glasgow, Scotland*

**Prof. Min Xie**

*City University of Hong Kong, Hong Kong*

**Prof. Irina Yatskiv**

*Riga Transport and Telecommunication Institute, Latvia*

**Indexed by:**



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.



**Task** „Implementation of procedures ensuring the originality of scientific papers published in the quarterly „Eksploracja i Niezawodność – Maintenance and Reliability” financed under contract 532/P-DUN/2018 from the funds of the Minister of Science and Higher Education for science dissemination activities.

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

**Our 2018 Impact Factor is 1.525**

**Editorial staff:**

Dariusz Mazurkiewicz, PhD, DSc (Eng), Associate Professor (Editor-in-Chief, Secretary of the Editorial Board)

Tomasz Klepka, PhD, DSc (Eng), Associate Professor (Deputy Editor-in-Chief)

Teresa Błażnio-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:**

“Eksploracja i Niezawodność” – Editorial Office

ul. Nadbystrzycka 36, 20-618 Lublin, Poland

e-mail: office@ein.org.pl

http://www.ein.org.pl/

**Circulation:**

550 copies

# INFORMATION FOR AUTHORS

## Terms and Conditions of Publication

1. The quarterly „Eksploracja i niezawodność - Maintenance and Reliability” publishes original papers written in English. Authors are strongly recommended to have their manuscripts checked by a native English speaker or a professional language editing service before submission, in order to ensure that the language is acceptable. Submissions may be rejected due to low quality of the translation.
2. The Quarterly welcomes all original submissions of articles that comply with its Aims and Scope. If a paper contains any material that has already been published (e.g. figures), then this must be declared by referencing the source in the manuscript, and authors are responsible for obtaining relevant permissions to republish. Evidence of permission granted must be available upon request.
3. All submissions should be made online via the peer-review system. Authors can use the „Articles” button to make a new submission or check the status of their manuscript. New users will need to create an account.
4. Submissions are made on the understanding that the manuscript has not been published elsewhere and is not currently under consideration by another journal published by any other publisher.
5. The submitting author is responsible for ensuring that the submission has been approved for publication by all co-authors. Submitting authors' responsibilities include especially the following issues: plagiarism, authorship, conflict of interest and misleading or misreporting of findings. In all cases, the Editorial Office reserves the rights to contact the authors' institutions if serious misconduct has been identified, and to retract any articles in which serious misconduct has been identified.
6. Each submission is checked for suitability when received by the Editorial Office, and may be rejected without review if it is outside the scope of the journal, is obviously of insufficient quality, or is missing important sections. This ends the assessment procedure. Re-submission of a rejected manuscript is not possible.
7. Reviewers are selected in accordance with the principle of avoiding conflict of interest (examples of conflicts of interest include a personal relationship between reviewer and author, occupational subordination, direct scientific cooperation over the last two years prior to the review). The journal operates a single-blind peer review system.
8. The review is provided in written form and concludes with an explicit recommendation to accept, revise or reject the paper for publication. If a negative review is given, the submission is rejected, which ends the assessment procedure. Re-submission of a rejected manuscript is not possible.
9. The Editor-in-Chief makes the publication decision after receiving at least two external reviewer reports with recommendations. On receiving the required number of reviews, the Editors pass them immediately to authors. The Editors require that authors prepare answers to reviewers' comments, which are then sent to the referees. The Editor-in-Chief will make a decision to accept, accept with minor revisions, accept with major revisions, or reject the paper for publication. The reasons for the decision will be communicated to the authors. When the decision of minor/major revisions is made, and the authors do not revise their articles satisfactorily after receiving reviewer reports, then the Editor-in-Chief reserves the right to reject the article. When revised articles are received, they will either be sent out for further review or the Editor-in-Chief will make a decision depending on the level of revision requested.
10. The time required for reviewing a manuscript and making a publication decision may vary a lot from one submission to the next since it is sometimes difficult to find suitable reviewers, and there may be delays in receiving reviewer reports. The Editor-in-Chief and the Editorial Office try their best to minimize the time from submission to first decision. The journal aims to make the first decision within 7 days after the Editorial Office check and the second decision within 30 days after the review, but no guarantees can be made in this regard.
11. Submissions with positive reviews are subjected to an originality check. Only those articles that have successfully passed the anti-plagiarism test are finally accepted for publication.
12. All accepted articles are published as open-access and distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. Full texts of all printed articles are available on the website of the Quarterly.

## Fees

The publication fee for one text is 410 EUR (1700 PLN for submissions from Polish institutions).

## Technical requirements

1. All papers must be submitted in a Microsoft Word document format. Drawings and photos should be submitted as separate graphic files in \*.tif, \*.jpg or \*.cdr format.
2. A manuscript should include:

**Names of authors, title, abstract, and keywords** that should complement the title and abstract. **The corresponding author** should be indicated. Abstract should not exceed 1100 characters (including spaces) and **must be divided into the following sections**: Objectives, Methods, Results, and Conclusions.

**Highlights** (a short collection of bullet points that convey the core findings and provide researchers with a quick overview of the article in text form) - three to five bullet points should describe the essence of the study (e.g. results or conclusions) and highlight what is distinctive about it. Highlights should not exceed 90 characters (including spaces) in each bullet point.

**The text 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 English-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 enclosed reference format). Authors using the Mendeley and the Zotero citation plugins and styles support may download the required format file from the repository.

**Author data** should include first names and surnames along with scientific titles, affiliation, detailed postal address and official e-mail address.